Supplementing the studbook. Using genetic analysis to complement a captive breeding programme of an endangered primate, *Hylobates moloch.*

Katherine Stanbury

A thesis submitted for the degree of Doctorate of Philosophy

Writtle College

University of Essex

Date of submission January 2015

Abstract

Genetic analysis of animals involved in captive breeding and reintroduction programmes can provide valuable information to aid in maintaining wild type genotypes and genetic variability. *Hylobates moloch,* also referred to as the silvery gibbon, is an Endangered primate species endemic to the Indonesian island of Java. As part of an overall conservation programme, a captive breeding and reintroduction programme is being organised. In order to aid both the management decisions within the breeding programme and success rates of re-introductions analyses at three genetic regions were carried out, with DNA extracted from non-invasively collated faecal samples. In order to assay if captively bred individuals were representative of their wild conspecifics, the population was split into two groups representative of wild born and captive born individuals. Genetic analyses at mitochondrial DNA hypervariable region-I (mtDNA HV-I), 15 microsatellite loci and the second exon of the major histocompatibility complex (MHC) *DRB* region, were carried out to ascertain genetic variability levels, levels of inbreeding, signs of selection and confirm the pedigree. Captive born individuals had markedly lower levels of variability at mtDNA HV-I, which was significant versus the wild group. The second neutral marker of microsatellites revealed no differentiation between wild and captive-born; moreover measures of standardised heterozygosity demonstrated a fairly high level of genomic variability overall. Pedigree analysis using the microsatellites produced information that differed from studbook entries. This was further supported by haplotypic data compiled from the MHC *DRB* exon 2 analysis. The MHC study revealed a total of 14 *DRB* alleles, 10 of which are from unknown lineages when compared to human and chimpanzees. As with microsatellites, no group differentiation between wild and captive has occurred but there are more rare

alleles present within wild individuals. In conclusion, whilst genetic variation is both high and shows no deviation from wild-born to captive-born at neutral microsatellite loci, care should be taken to maintain rare mtDNA haplotypes and MHC DRB alleles in future generations.

Acknowledgements

I would like to express very special thanks to Dr. Carlos de Luna, who introduced me to this project and served as my PhD supervisor. You recognised my passion for animal conservation whilst I was an animal science undergraduate and then encouraged me to pursue it further. In all the ups and downs on my PhD journey that then ensued you always believed in me.

Thank you to Matthew Ford and Simon Jeffreys from Howletts and Port Lympne who collected the samples for this research. Thank you also to Dr. Joe Smith from Fort Wayne Childrens Zoo who also collected samples but for reasons beyond my control were not included within the study.

I would also like to express special appreciation to Dr.Gaby Doxiadis from the Biomedical Primate Research Centre and her amazing team, Annemiek de Vos Rouweler and Nanine de Groot. Your help was invaluable for work carried out for the final experimental chapter. Thank you for allowing me to work in your laboratory and for sharing your expert knowledge in the subject.

Finally, I have to thank my family and my friends. This project became my life and they stood by me and encouraged me every step of the way. To my parents, Val and Ken and my brother Stephen, your faith in me kept me motivated. To my husband, words can not express my thanks to you for all of your help, from making endless cups of coffee to proof reading. Your unwavering support was instrumental to my success. Lastly, to my son Noah. I dedicate this to you.

Table of Contents

List of Tables

List of Figures

1 Chapter one – Introduction

1.1 The importance of conservation

Extinction is a natural process representative of the end point of a species evolution. It has, however, become distorted by anthropogenic actions such as habitat destruction, excessive hunting and as a consequence of human induced climate change (Andrabi & Maxwell 2007). These, often irreversible, activities have resulted in an acceleration of extinction rates of up to 1000 times their natural occurrence (Brooks et al. 2006). The recent "Living Planet Report" (2014) published by the World Wide Fund for Nature International stated that vertebrate species have declined by over 50 percent within a forty year period (from 1970 to 2010). Extinctions are not solely a severely negative event for the species in question, but also have knock on effects on the ecosystem from which it is extirpated that can be extremely detrimental. Alterations to species' richness within habitats have been theoretically and empirically shown to change biotic interactions, such as those evinced within trophic webs, and also alter energy production levels within ecosystems (Chapin III et al. 2000; Nichols et al. 1998; Paine 1980; Worm et al. 2006). The consequences of alterations to an ecosystem and changes of faunal community structure have also resulted in losses of species richness within the localised habitat. Paine (1971) observed that in an intertidal habitat the removal of a single carnivorous starfish (*Stichaster australis*) for a period of just 9 months resulted in the loss of 6 different species (from 20 to 14 species) owing to a change in predation and thus in the resources available.

Arguably, the most gainful approach to conserving wildlife species is to preserve their habitats; preferably on the largest scale possible. This conservation strategy not only offers the potential to maintain biodiversity, but also achieves economies of scale as multiple organisms are safeguarded as opposed to focusing on a single species (Simberloff 1998; Pukazhenthi et al. 2006). However, in practice this option is not always possible and the maintenance of a growing number of wildlife populations requires more taxon specific targeted management strategies. One such strategy is that of captively breeding endangered species whose objectives involve future reintroductions of individuals into their wild environment. If an Endangered species can be successfully bred in captivity, the resultant group has the potential to act as a metapopulation to their wild conspecifics (Britt, Welch & Katz 2003). As a consequence of the acceleration of extinction rates of differing organisms, the practice of captive breeding is postulated to undergo a large increase, with estimates of 2000 to 3000 vertebrate species alone destined for captive breeding if they are to subsist for future generations (Frankham 2008). This, of course, would only be able to transpire if space within zoological institutions and funds were to be made available. In a bid to secure the future of the silvery gibbon, Howletts and Port Lympne, operating under the remit of the charitable organisation, the Aspinall Foundation, has selected the species for an overall conservation programme that includes both a captive breeding and re-introduction programme. Collectively, these zoo parks have already had great success in breeding the *H. moloch* gibbon thus increasing *ex-situ* population numbers (Aspinall Foundation 2010).

1.2 Captive breeding and reintroduction programmes

Captive breeding programmes, in particular those that include aims of reintroductions, have over the years recognised that naturalistic behaviours of species can be altered owing to the captive environment. This phenomenon has been shown to negatively impact on the survival rates of captive-born released animals. For example, after a long-term conservation project for golden lion tamarins (*Leontopithecus rosalia rosalia*) the survival rate of reintroduced captive-born individuals was very low, owing to adaptation to their captive environment that impacted foraging and locomotor skills once released into the wild (Stoinski & Beck 2004). Other vital behaviours such as predator avoidance and hunting ability as evinced for example in carnivores, are not apparent after a number of generations in captivity (Jule, Leaver & Lea 2008; McPhee 2003). All of these examples pose difficulties for conservation managers. They are issues addressed in breeding and reintroduction programmes so as to bolster success rates. However, another aspect which is equally valuable but is rarely addressed within such programmes, is that changes may be occurring at the genetic level as well as at the behavioural one.

The IUCN recommends that conservation be carried out at three levels, the ecosystem, the species and at the genetic level (Frankham 2010). The creation and maintenance of a captive breeding programme addresses the IUCN recommended approaches to conservation at the species level. Many breeding programmes in zoological institutions then use a proxy measure known as Mean Kinship (Ballou & Lacy 1995) to address the issue of genetic conservation of their captive species (Rudnick & Lacy 2008).

The Mean Kinship method functions by using accurate studbook records that record the pedigree of individuals within a population. It quantifies the relationship of an individual to all other members of the population by a kinship coefficient (f_{ij}) that is defined by the likelihood of random alleles drawn from two individuals (i and j) and assumes that alleles are identical by descent (Falconer 1981). The mean kinship of each individual is then ascertained by defining the average of the kinship coefficients of the individual under assay to all other living members within the population. The calculation to ascertain mean kinship is as follows, where *N* is the number of individuals in the population (Ballou & Lacy 1995). :

$$\mathrm{mk_i} = \frac{\sum_{j=1}^{N} \mathrm{f_{ij}}}{\mathrm{N}}$$

The result is used in the captive breeding process by selecting individuals who have a low mean kinship score. Individuals with a high mean kinship value are deemed to have a high level of genetic representation amongst the breeding group. Therefore, they will most likely be excluded from future pairings. An assumption of the method is that founder individuals are unrelated, a deduction that may not be correct. However, it has been stated that including an extra step to test founder relatedness is of little benefit as the additional information, after computer simulation, was found to have a minimal impact on overall results (Ivy et al. 2009; Rudnick & Lacy 2008). Overall, the method is deemed to be an effective approach to maximise diversity and minimize inbreeding in a captive breeding programme (Montgomery et al. 1997).

Whilst the mean kinship method derived from accurate studbooks appears to be a valuable measure of the levels of relatedness within a captive breeding population, it is only a proxy measure, inferring levels of inbreeding and genetic variability. Furthermore, it is reliant on studbook entries to be accurate, and thus if an entry were to be incorrect it would confound all data pertaining to that individual. Moreover, the true level of genetic diversity within the population, whether from the genetic 'starting point' of the group from its founders or the subsequent generations of captively born individuals is unknown. Whilst the aim of zoological institutions may be to maintain a genetic structure representative of wild conspecifics (Ivy & Lacy 2010) within subsequent generations of captively born individuals, the method does not yield information pertaining to rare and potentially valuable alleles within a population. An individual may harbour a rare allele but also be deemed to have a high mean kinship value and thus may be excluded from future pairings. However, it is not known whether the rare allele has been transmitted to progeny and maintained within the population, or whether the allelic richness of the population is lowered by the exclusion of the individual.

1.2.1 The problems of inbreeding, outbreeding and adaptation

A pertinent issue for Endangered species within captivity is the risk of inbreeding, as a consequence of populations very often existing in small numbers (Frankham 2005). With a limited number of individuals in the available gene pool a trend towards homozygosity may be expected. Genetic drift may become influential in small populations, and culminate in an overall diminished level of genetic diversity (Frankham, Ballou & Briscoe 2010). Inbreeding depression is described as a universal event both in captivity and in the wild, owing to the fact that the population numbers of randomly mating individuals within any species is finite (Reed et al.

2003). The effects in a small population, however, can result in severely negative results acting on attributes required for a successful captive breeding programme. With a decline in genetic heterozygosity, traits such as sperm production in males and fecundity levels within females have been reported to have been negatively affected, and juvenile survival rates have declined (Frankham 2005). A further characteristic of inbreeding depression, that has been documented to impact species' fitness levels is the fixation of deleterious alleles within a population. The Californian condor, a 'flagship' species of North America suffered a drastic decline in population numbers that by the year 1987, just 27 individuals remained (Adams & Villablanca 2007). Although numbers have since been bolstered owing to the remaining individuals being taken into captivity for breeding, a recessive allele known to cause chondrodystrophy, a mortal form of dwarfism, has been fixed within the population and is present at a relatively high frequency (Ralls et al. 2000).

Genetic variability is described by Amos and Harwood (1998) as the "clay of evolution". This reflects the concept that organisms that possess high levels of genetic diversity are better equipped to respond to intrinsic and extrinsic threats to their survival (Reed & Frankham 2003).. This ability to adapt is an extremely pertinent issue for individuals in a breeding program where they are destined for reintroduction to their wild environment. There are reports of adaptations occurring owing to captive environments (e.g. Čížková et al. 2012; Christie, Marine & Blouin 2012; Montgomery et al. 2010; Ping-ping et al. 2005), all be it largely within the context of laboratory conditions but the phenomena is considered by some such as Frankham (2008) to warrant a greater focus from the scientific community. Under

laboratory conditions the eye size of fruit flies (*Drosphila melanogaster*) decreased over a number of generations possibly as a consequence of reduced light within their laboratory cage environment (Pelletier et al. 2009). The impact that the number of generations that are born and remain in the captive environment, on genetic and phenotypic changes has been highlighted as a significant issue (Frankham 2008; Williams & Hoffman 2009). For example, the Mallorcan Midwife toad (*Alytes multensis*) develops a tail during the tadpole lifestage which is presumed to function as a predator response mechanism to escape predation (Kraaijeveld-Smit et al. 2006). After 9 to 12 generations in a captive breeding programme this predator defence mechanism not only matured on a slower scale than evinced in both wild counterparts and those bred within fewer generations, but the physical nature of the tail also changed. This would render them potentially vulnerable to predators known to inhabit their native habitats. In addition, a decrease in neutral genetic variation quantified by microsatellites was also observed (after an equal number of generations.) After a prolonged period in captivity of over 100 generations, the fecundity rate of large white butterflies (*Pieris brassicae*) increased and general morphological attributes such as body size also increased (Lewis & Thomas 2001). For the butterflies in captivity these physical alterations may not pose a problem. However, in the wild an increased body mass would mean that an elevated amount of trophic resources may be required and increased fecundity result in a greater number of butterflies, but both of these factors may result in a subsequent decline in numbers if the carrying capacity of the habitat is exceeded. Christie *et al.* (2012) observed genetic adaptation to the captive environment in steelhead fish (*Oncorhynchus mykiss*) after just one generation. The

fitness of captive-born steelhead was drastically reduced when compared to their wild counterparts when quantifying success of first generation hatchery fish.

Adaptation to a localised environment has also been cited as one of the factors that can contribute to a reduction in fitness of progeny via outbreeding of two genetically dissimilar parents. The concerns of outbreeding are that localised selection pressures may translate to phenotypic adaptations in a population, which may then be disrupted at the genetic level if an individual not bearing these adaptations were to mate with a localised individual (Sagvik, Uller & Olsson 2005). Outbreeding has been reported in rainbow trout (*Oncorhynchus mykiss*) to affect fitness levels of progeny after only three generations (Tymchuk, Sundström & Devlin 2007). A similar reduction in survival rates of offspring was observed in a small population of ornate dragon lizards (*Ctenophorus ornatus*) (LeBas 2002). It was hypothesised that outbreeding was more detrimental to individuals born from distantly related parents than to those born to more inbred parents. As the population had been physically and thus genetically isolated for a great deal of time it appeared that the introduction of new genes did not function well in their new environment (LeBas 2002). However, a more complex result from outbreeding events was studied between two populations of common frog (*Rana temporaria*) (Sagvik, Uller & Olsson 2005). In two populations, one large and a smaller one located 130km away found that when males from the small population mated with a female from the large population, tadpoles were smaller and more malformed. However, this was not the case when males from the large population mated with females from the small population. Therefore, it was advised that translocations or introductions between

populations be carried out with caution so as not to negatively impact on fitness levels (Sagvik, Uller & Olsson 2005). There are those however, that believe that inbreeding is a more detrimental process than outbreeding and with careful introductions gene flow may improve fitness levels of isolated or small populations over time (Beauclerc, Johnson & White 2010; Frankham et al. 2010b; Hogg et al. 2006). The recommendations from Frankham *et al.* (2010b) are that to avoid of inbreeding depression, mixing of intra-species populations occur if genetic isolations has taken place for a period of less than 500 years and that habitats be similar.

1.3 Hylobates moloch – the silvery gibbon

Hylobates moloch, also referred to as the silvery gibbon or Javan gibbon, belongs to one of the four recognized genera (*Hylobates, Hoolock, Nomascus* and *Symphalangus*) of gibbons (Kim et al. 2011). Each genus shows several unique characteristics including morphology, and in the case of the *moloch* species song bouts (Geissman & Nijman 2006), as well as karyotypes which range from *2n*=38 to 52 (Chan et al. 2010; Kim et al. 2011). The *H. moloch* gibbon is a small arboreal ape endemic to the Indonesian island of Java. The *Hylobates* genus is thought to be a monogamous taxon with the *moloch* group usually comprised of an adult male and female and with 1 to 3 immature offspring (Dallman & Geissman, 2009; Oka & Takenaka, 2001).

According to the IUCN red list, all but one (*Hoolock leuconedys,* listed as Vulnerable) of the gibbon species is either Endangered or Critically Endangered (IUCN 2008). A population survey of *H. moloch* carried out between 1994 and 2002

found that 4000 to 4500 individuals were present in habitat fragments in the west of Java, but also in central Java (Nijman 2004). This estimate exceeded previous studies carried out on the species and thus the IUCN downgraded its status from Critically Endangered to Endangered in 2008 (IUCN 2008). However, despite what appears to be positive news with this revised estimate of population numbers the trend noted by the IUCN is that the species is in decline. This is of little surprise when considering that one of the major driving forces effecting losses of *moloch* individuals in the wild is anthropogenic actions, between 96-98% of their forest habitat has been subjected to deforestation (Geissman & Nijman 2006; Nijman 2004). Consequently, the silvery gibbon is viewed as one of the most urgent conservation priorities of all Asian primates (Geissman & Nijman 2006).

The global captive population of *H. moloch* totals 119 individuals, 71 of which are located in zoological institutions in Indonesia where attempts to breed the species have been largely unsuccessful (IUCN 2008). Outside of Indonesia, however there are 48 individuals spread across several zoological institutions, and half of this number reside at Port Lympne and Howletts parks within the United Kingdom.

1.4 The aims of this study

It is evident that inbreeding, loss of genetic diversity and adaptation to the captive environment are factors that may negatively impact fitness levels and thus success levels of captive breeding programmes. Whilst the Mean Kinship approach has been shown to be a valuable measure of some of these factors, it remains a proxy and thus not an actual measure of an individual's genetics. The aim of this study is

to analyse three genetic regions in a captive population of Endangered *Hylobates moloch* to provide vital and supplemental information to aid in pairing individuals and selecting individuals for captive breeding and reintroduction to Java. The information will pertain to the actual levels of inbreeding within the population, how genetically variable the individuals are and whether any genetic processes are acting on the group that would affect how individuals are managed. In addition, the genetic analyses will verify pedigree data in the studbook.

1.5 The objectives of this study

The objectives of the study were to analyse three differing genetic regions within *H. moloch* individuals. The individuals included in the study were all residents of either Howletts or Port Lympne zoological institutions who are members of the European Association of Zoos and Aquaria (EAZA). Collectively at the time the study commenced, they housed 24 *moloch* gibbon individuals, which is half of the global captive population (excluding Indonesia). As one of the integral aims of a captive breeding programme is to maintain wild type genetic diversity within individuals who are captive-born, the study population was split into two groups to discern whether individuals born within the captive environment remained genetically similar to their wild conspecifics or whether they had been subjected to inbreeding, or losses in genetic diversity. The first group was representative of wild type genetic levels derived from individuals previously extracted from their native habitat and were thus named the wild-born group. The second group were all individuals born within the captive environment and were thus named the captive-born group.

There are very few studies published on the genetics of *Hylobates moloch*. Studies of the *moloch* gibbon and indeed other members of the *Hylobatidae* family have largely focused on areas of mitochondrial DNA (mtDNA) (e.g Chan et al. 2010; Monda et al. 2007; Takacs et al. 2005; Thinh et al. 2010; Roos & Geissman 2001). This genetic region has received a great deal of focus as mtDNA is often employed in phylogenetic studies owing to its lack of recombination, and the phylogeny of the *Hylobatidae* is unresolved. Andayani *et al* (2001), for example, proposed that as a consequence of a molecular study at the mtDNA *H. moloch* should be split into two sub-species. This has been contested. In 2014, Carbone *et al.* published findings procured from a genome assembly study of a northern white-cheeked gibbon (*Nomascus leucogenys*) which was the first study investigating a gibbon species on a large scale. The lack of genetic studies that focus on the gibbon genera is surprising as they represent a unique node within primate phylogeny and have been described as having experienced a "near instantaneous" radiation approximately 5 million years ago (Carbone et al. 2014). Thus within the context of primate evolution, the *Hylobatidae* are an intriguing taxon as the observed chromosomal reshuffling has occurred on a relatively short time scale (Carbone et al. 2014; Kim et al. 2011; Müller, Hollatz & Wienberg 2003).

Within this study two genetic markers that are presumed to evolve on a neutral basis were chosen for analysis as it was deemed that any deviations from neutrality would highlight if selection is acting on the population. The first was derived from the control region of mtDNA known as hypervariable region-I. This genetic region is known to evolve 5 to 10 times faster than nuclear DNA and provides information on

a differing timescale than derived from nuclear DNA. This was deemed to be of value for comparing Wild and Captive groups as captive-born individuals represent a limited number of generations. The second neutral set of markers analysed were non-coding microsatellites derived from nuclear DNA, providing a more varied picture of genetic variability. These repeat motifs of DNA are located throughout the genome in eukaryotes (Kelkar et al. 2011) and can be highly polymorphic (Morin et al. 2004) and thus are of benefit to confirm the pedigree of the population and also provide a measure of overall genomic diversity. The third genetic region assayed differs in that it is of adaptive importance (Frankham 2010). It was derived from the major histocompatibility complex (MHC), a complex gene family that has vital links with fitness owing to its involvement in immune function (Smith, Belov & Hughes 2010) and is described as the most polymorphic within vertebrates (Ejsmond & Radwan 2011; Piertney & Oliver 2006).

2 Chapter two - Mitochondrial DNA analysis

2.1 Mitochondrial DNA – Form and Function

The mitochondrion is a unique eukaryotic cellular organelle that provides vital functionality predominantly, but not exclusively, owing to its role within cellular energetics. It is unique in comparison to other cytoplasmic cellular components as it contains its own genetic material in the form of mitochondrial DNA (mtDNA), a significant factor that lead to the formulation of the theory that this organelle evolved via endosymbiosis (Sato & Sato 2013). The overall structural formation is shown in Figure 1 which shows that the mammalian mitochondrial genome is organised in a closed double stranded circular configuration. As opposed to its nucleus bound counterpart, mtDNA has a fairly conserved architecture within the Kingdom *Animalia* (Freeland 2005). Mammalian mtDNA encodes for 13 polypeptides that form the blocks to create the bioenergetic pathway of the electron transfer chain, in addition to two rRNAs and 22 tRNAs (St. John 2014). These 13 proteins work in conjunction with approximately 70 nuclear proteins (Chen & Butow 2005) to perform biogenesis of the cellular fuel known as adenosine triphosphate (ATP) via the oxidative phosphorylation process. As functionality of this molecule is highly demarcated to a number of specific roles, the formation is purported to have evolved as a highly organised and efficient molecule with little overlap, if any between differing genes (Pereira et al. 2008).

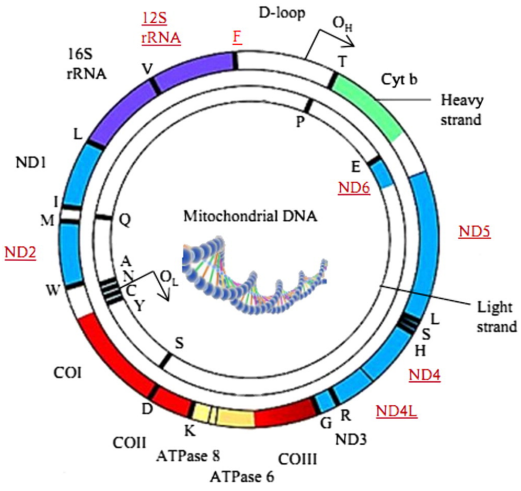


Figure 1. Mitochondrial DNA molecule (taken from Lacobazzi et al., 2013)

Despite this apparent rigidity within its formation and function, there are two non-coding regions embedded within the circular structure. The first is an extremely small region comprised of just 30 base pairs located between the genes NADH dehydrogenase subunit 2 (ND2) and cytochrome oxidase I (COI). The second is illustrated above in Figure 1 as the D-loop and located between ribosomal RNA (12s rRNA) and cytochrome b (Cytb) genes. The D-loop, which is a diminutive term for displacement loop is also referred to as the control region of mtDNA. Although the terminology D-Loop is widely used in the literature, Pereira *et al*. (2008) clarified that its use as an equivalent term to that of the control region is in fact a misnomer as D-loop is actually descriptive of a loop formation created by early termination of the heavy strand synthesis at the 5' of the region, and thus representative of a specific point in the control region rather than its entirety. The exact functionality of this particular folding element within the control region is unknown. However, theories postulated include that the D-loop modulates mutation rates within the molecule owing to the assumption that such secondary structures are binding sites for a number of transcription factors (Pereira et al. 2008).

The control region is the non-coding segment of mtDNA charged with the task of regulating transcription and replication of this molecule and in humans it extends approximately 1.1 kb (Tzen, Hsu & Wang 2008). It is then defined into two further segments, the classifications of which elucidate the high variability of this section: hypervariable region–I (HV-I) and hypervariable region-II (HV-II). Within primates, the HV-I region is purported to mutate at a rate 5 to 10 times faster than that of the nuclear genome (Andayani et al. 2001). This section mutates at a rate significantly greater than any other mtDNA segment (Roos & Geissman, 2001; Whittaker et al., 2007). The central area of the control region that concatenates the two hypervariable regions is conserved in humans (Tamura & Nei 1993).

A further attribute of the mtDNA molecule that sets it apart from nuclear DNA is that the molecule is transmitted as a haploid unit, with uniparental inheritance solely from the mother in most animal species (Frankham, Ballou & Briscoe 2010). The exact mechanisms of how this occurs are not fully understood (Sato & Sato 2013), and the processes may differ from species to species (Birky, Jr 1995). It is known in mammals that paternal mitochondria penetrate the oocyte cytoplasm post fertilization but the genetic information contained within is not then transmitted to the offspring (Sato & Sato 2013). However, there have been reports of paternal 'leakage' in some analyses (Lunt & Hyman 1997; Zhang & Hewitt 1996). It has long been thought that recombination within this molecule does not take place in animals (Rokas, Ladoukakis & Zouros 2003). However, Tsaousis *et al* (2005) note that recombination within the mtDNA molecule may be occurring within an increasing number of animal species than previously thought, and this could potentially impact on genetic studies of the molecule. However, this phenomenon is unlikely to affect

the outcome of this study, as although it is possible that paternal mtDNA may be present within the *H. moloch* samples, prior studies have shown that haplotypes have not been sufficiently altered to produce a new haplotypic sequence (Rokas et al., 2003; Tsaousis et al., 2005).

2.1.1 Mitochondrial DNA as a molecular marker

Mitochondrial DNA has been used as a molecular marker since the end of the 1970s (Zhang & Hewitt 1996). The nature of mtDNA such as its conserved structure, lack of recombination in most animal species and small size make it a more amenable molecule to assay than nuclear DNA. With a conserved architecture, it is possible to utilise previously described generic primers that can be applied across a variety of species, thus conserving both finances and time (Frankham 2010). In addition mtDNA is present at high numbers within the cell, for example reaching 250,000 copies in the mature metaphase II oocyte (St. John 2014). The number then fluctuates later in life as a result of differing metabolic demands, however this high copy number render it a more compliant molecule when utilised in the polymerase chain reaction (PCR) procedure.

The applications of mtDNA within genetic analyses studies are numerous. Studies include those focusing on phylogenetics (e.g. Belay & Mori, 2006; Whittaker et al., 2007; Zenger et al., 2005), population structure (e.g. Ahlering et al., 2011; Caballero et al., 2013; Chow et al., 2009), disease (e.g. Kenney et al., 2014; Stewart et al., 2008) and more recently a focus on the effects on health as a result of the post-translational process of DNA methylation (e.g., Lacobazzi et al., 2013).

Owing to its haploid character, when analysing the mtDNA molecule the population under review is effectively a quarter of the actual population size. Whilst this factor makes it a more amenable molecule to assay in many cases such as in studies of maternal lineage, it must be considered when calculating genetic variability. A species' mating system can also influence population genetic dynamics. For example, low genetic diversity at this molecular region may not be as a result of a bottleneck as a result of a decline in population size, but may be attributable to the social dynamics of the target population in question. For example, a study to ascertain genetic variability between and within populations of bull shark (*Carcharhinus leucas*), found that individuals exhibited low diversity at both the nucleotide and haplotype levels (Karl et al. 2011). The results were not as a consequence of declining populations, this is not a species that is considered Endangered and enjoys a widespread global distribution, rather that female bull sharks showed a high degree of philopatry at nursing sites (Karl et al. 2011).

The importance of understanding the underlying diversity of the mtDNA region is highlighted in a study of Black Rhinoceros (*Diceros bicornis*) (O'Ryan, Flamand & Harley 1994). When selecting animals that have originated from different regions for breeding and reintroduction purposes the diversity at this region was considered with regards to outbreeding. Diversity between the different subjects was low and therefore it was decided that to bolster numbers and mix gene flow, the different populations could be interbred. A similar approach to conserve Eastern barred bandicoot (*Perameles gunnii*) was used where mtDNA variability was ascertained prior to mixing two different populations originating from different areas (Robinson 1995). Results obtained illustrated that there was a marked sequence divergence

from the two populations and therefore it was decided that mixing individuals may impact negatively on fitness owing to outbreeding.

The neutral theory of molecular evolution was first described by Kimura in 1968, a hypothesis that was then further defined over the following decade and has become a core concept within population genetics (Kimura 1968). The theory postulates that the process of selection is less of a factor in driving molecular variation, and rather that genotypes that persist in a population have equal neutrality with respect to each other when considering their connections with fitness traits (Hedrick 2005). The mtDNA molecule appears to split opinion as to whether it follows a neutral pattern of evolution or whether it is under selection. Many studies select mtDNA on the basis that it evolves in a neutral manner (Kanthaswamy et al., 2006; Nachman et al., 1994). This assumption may be based in the fact that the molecule is of such vital importance as the 'power house' of the cell that alterations within the structure may impair vital biological function. To function at the most efficient level possible, it may follow that mtDNA is in fact a non-neutral marker and is susceptible to selection thus allowing change over future generations and to fixate the most desired alleles for optimal function. One selective force postulated to influence mtDNA genes is that of alterations in climate (Mishmar et al., 2003; Tomasco & Lessa, 2014). Alterations were observed in a number of mtDNA genes responsible for ATP production in accordance with different climates in the study by Mishmar and colleagues (2003) using samples derived from human populations from differing global regions. A further study that also raises the notion that external environmental conditions influence mtDNA genotypes is that of Tomasco and Lessa (2014). They postulated that in rodents, alterations in two mtDNA genes occurred as a result of modifications

in their ecological surroundings. Changes within COX2 and CytB genes were discovered within two codons and were hypothesised to be under episodic selection in accordance with their external environmental conditions. The species in the study is a subterranean rodent that dwells in low oxygen habitats but requires high energy levels for burrowing. A further hypothesis surmises that there is strong purifying selection within mtDNA as a means to limit deleterious accumulations to resist disease (Stewart et al. 2008). Modes of evolution within mtDNA may follow both selective and neutral theories of mutation and influential forces may come from both biotic and abiotic factors.

2.2 Materials and Methods

2.2.1 Sample population

The sample population for the study was comprised of 21 *Hylobates moloch* individuals residing at Howletts and Port Lympne zoological institutions both of whom are EAZA (European Association of Zoos and Aquaria) members. The studbook for the *moloch* gibbon is managed by Howletts for individuals within the United Kingdom. There should have been a total of 24 *H. moloch* for the study, however three infants from the population would not take marked hand fed food items thus identification of faecal deposits was not possible.

The population was then split into groups representative of wild-born individuals (Wild) and those born within the captive environment (Captive). Wild-born individuals (*n*=8) are denoted with the reference H.mol1 to H.mol8 and Captive-born individuals (*n*=13) with H.mol9 to H.mol21. Further information for each animal is in Table 1.

All Mean Kinship values were provided by Howletts. Missing values represent individuals that when MK values were created were no longer in the group.

A further group of 31 wild *H. moloch* individuals were included in statistical analyses. The sequences derived from GenBank were the only other published data available on this genomic segment within this species. This group is designated Wild_pop2 and all sequences were deposited in GenBank with the following accession numbers by Andayani *et al.* (2001):

AF338908.1, AF338906.1, AF338904.1, AF338902.1, AF338900.1, AF338898.1, AF338896.1, AF338894.1, AF338892.1, AF338890.1, AF338888.1, AF338886.1, AF338884.1, AF338882.1, AF338880.1, AF338878.1, AF338874.1, AF338907.1, AF338901.1, AF338899.1, AF338897.1, AF338895.1, AF338893.1, AF338891.1, AF338889.1, AF338887.1, AF338885.1, AF338881.1, AF338879.1, AF338875.1, AF338873.1.

Table 1 - *Hylobates moloch* sample population information. Primates in bold are wild-born.

Primate Ref	Sex	Year of Birth	Generation
H.mol1	**F**	**Unknown**	**1**
H.mol2	**M**	**Unknown**	**1**
H.mol3	**F**	**Unknown**	**1**
H.mol4	**M**	**Unknown**	**1**
H.mol5	**M**	**Unknown**	**1**
H.mol6	**M**	**Unknown**	**1**
H.mol7	**M**	**Unknown**	**1**
H.mol8	**F**	**Unknown**	**1**
H.mol9	M	2001	2
H.mol10	F	1999	2
H.mol11	M	2006	3
H.mol12	F	2004	3
H.mol13	M	2009	4
H.mol14	F	2009	3
H.mol15	F	1996	3
H.mol16	M	1990	2
H.mol17	M	1993	2
H.mol18	F	1994	2
H.mol19	M	2007	3
H.mol20	F	2005	3
H.mol21	F	1998	3

2.2.2 DNA extraction

DNA was extracted from faecal samples collected from 21 of the 24 *H. moloch* individuals residing at Howletts and Port Lympne zoological institutions. Faecal samples were allocated to specific individuals either by direct observation and collected immediately or by hand feeding foodstuffs containing maize that could then be identified on subsequent days. Samples were immediately frozen upon collection. Two faecal samples per individual were collected. It has been widely reported that both DNA yield and quality derived from faecal samples is lower than that extracted from blood or tissue (e.g. Chaves et al., 2006; Marrero et al., 2009; Wasser et al., 1997). However, faecal samples are a non-invasive medium from which to obtain genetic material, which is desirable from both a practical standpoint as it is not necessary to obtain licences but much more importantly it causes no stress to the animal under observation.

DNA was extracted from the frozen faecal samples using the QIAamp DNA stool Mini Kit (Qiagen) following the Stool Larger Volumes protocol. This commercially available kit was chosen as it included a step that involved binding secondary compounds found in plant matter that is present in herbivorous diets, which is applicable to the *H. moloch* species. Plant secondary compounds negatively impact the PCR process by interfering with the taq polymerase enzymatic reaction (Marrero et al. 2009). For each DNA extraction an amount of 400mg of frozen stool was used ensuring that both internal and external surfaces of the faeces were present as a precautionary measure as sloughed epithelial cells may not be homogenously distributed throughout the sample (Piggott & Taylor 2003).

2.2.3 Hypervariable Region I (HV-I) PCR amplification

Once two DNA samples had been prepared for each individual, extracted DNA was used to amplify the desired target region of mtDNA, the HV-I. The HV-I region of the mitochondrial DNA control region was amplified using the gibbon specific primers GIBDLF3 (5'-CTT CAC CCT CAG CAC CCA AAG C 3') and GIBDLR4 (5'-GGG TGA TAG GCC TGT GAT C-3') as published in Andayani et al., 2001. Total volume of PCR mixture per reaction was 50µL consisting of: 1x Q Solution (a reagent that is part of the HotStarTaq® kit, modifies the melting behaviour of DNA and useful for difficult template DNA, Qiagen), 1x PCR buffer (contains Tris-CL, KCL, $(NH_4)SO_4$, 15mM $MgCL_2$) (Qiagen), 1.75mM $MgCl_2$, 0.2µM dNTPs, 0.1µM Primer GIBDLF3, 0.1µM Primer GIBDLR4, 2.5 Units HotStarTaq® DNA Polymerase (Qiagen), 20-50ng template DNA. A negative control was included to monitor for contamination. No positive control was used here as primers utilised were specific to the gibbon genera. PCR reactions were run using a S1000 Thermal Cycler (Bio-Rad) with the following conditions: 95°C for 15 minutes to activate HotStarTaq, then 40 cycles of denaturing phase at 94°C for 1 minute, annealing at 60°C for 1 minute, extension at 72°C for 3 minutes and a concluding extension stage at 72°C for 6 minutes. The PCR product was then purified using SpinPrep™ PCR Clean-up kit (Merck, KGaA). PCR products were visualised on a 0.8% agarose gel to ensure amplification success. Successfully amplified products were then sequenced using Sanger ABI 3730xl. For quality control the process was repeated a second time.

2.2.4 HV-I data analysis

The first step of data analysis was to reconcile the two HV-I sequences from each individual, to ensure that each pair of sequences pertaining to a specific individual were identical. For control purposes, nucleotide quality at each site was first checked by viewing sequence chromatograms within 4Peaks version 1.7.2 (Griekspoor & Groothuis 2006). Then, duplicate sequences for each individual were aligned using ClustalX version 2.0 (Larkin et al. 2007). Once duplicate sequences had been verified to be identical, the sequences taken from GenBank from the study of Andayani *et al.* (2001) known as Wild_pop2 were added to ClustalX. All sequences were then aligned in accordance with the 'multiple alignment mode' as opposed to the 'profile alignment mode'. The multiple alignment mode functions by first analysing sequences in a pairwise manner and identifies regions of similarity between sequences. The distances between each pair of sequences within the dataset were calculated in accordance with the 'slow-accurate' alignment parameters (as opposed to the fast but approximate). Alignments were carried out with the default transition weighting of 0.5 as sequences were expected to be similar as they were all derived from the same species. Although mtDNA control region data analysed is non-coding by default the programme assigns a protein weight matrix, thus BLOSUM was selected which is the most applicable for studies utilising similar evolutionary distances as evinced here.

Once all sequences from the three groups (captive, wild and wild_pop2) were aligned they were edited within MacClade OSX version 4.08 (Maddison & Madison 2005). The wild_pop2 sequences were of the entire control region and thus they

were trimmed to equal the HV-1 segment. The resultant segment length of all group sequences was 518bp.

A consideration of aligned sequences from each individual is that there is a possibility that the target mtDNA segment may co-amplify with a similar sequence from nuclear DNA. This is as a consequence of a phenomenon where mtDNA sequence copies are inserted into nuclear DNA and are referred to as nuclear mitochondrial sequences (NUMTS). This occurrence has been reported in a number of different taxa (Hazkani-Covo, Zeller & Martin 2010) and with a differing number of mtDNA regions, for example an entire D-loop discovered in a domestic cat (Zhang & Hewitt 1996). Investigations in both humans and rhesus monkeys found that sections within the control region NUMTs were rare which lead the researchers to deduce that this region is under-represented within the primate taxa (Tsuji et al. 2012).

2.2.5 Statistical analysis

Statistical tests of neutrality on the HV-I mtDNA segment carried out were Tajima *D* and Fu's F_s test. Both tests are applicable for intraspecies data sets and use nucleotide information to perform calculations. Tajima *D* (Tajima 1989) compares the number of segregating nucleotide sites within a sequence with the mean pairwise difference between two random sequences. It is defined as follows:

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_\omega}{S_{\hat{\theta}_\pi} - \hat{\theta}_\omega}$$

Where $\hat{\theta}_\pi$ is an estimator of $\hat{\theta}$ based on the average number of pairwise differences, $\hat{\theta}_\omega$ is an estimator of θ based on the number of segregating sites and $S_{\hat{\theta}_\pi}$ is an estimate of the standard error of the difference of these two estimations. The significance of the *D* value was tested by 1000 simulations (as recommended within the programme) using Arlequin version 3.5.1.3 (Excoffier & Lischer 2010).

Fu's F_s test (Fu 1997) also analyses the nucleotide sequence information and assumes that no recombination has occurred but this computation is carried out at the haplotype level of data. This statistic is given by the following equation:

$$F_s = 1n\left(\frac{\acute{S}}{1-\acute{S}}\right)$$

Where $\acute{S}$ is an estimation of the probability of observing a random sample with the same number of alleles that are equal to or smaller than the observed value within a given haplotype dataset. The significance of F_s was also tested by 1000 simulations (as recommended within the programme) using Arlequin version 3.5.1.3 (Excoffier & Lischer 2010).

Statistical measures of genetic variability estimated were number of haplotypes, the number of polymorphic sites, gene diversity (*H*) and nucleotide diversity (π) and mean number of pairwise differences. To ascertain measures of genetic differentiation between the three *H. moloch* groups a measure of F_{ST} was calculated.

Gene diversity (also referred to as haplotype diversity) is representative of the probability that two random sequences are different and was calculated as follows (Nei 1987):

$H = \frac{n}{n-1}\left(1 - \sum_{i=1}^{h} p_i^2\right)$

Where *n* is the number of sequences, *h* is the number of haplotypes and p_i is the relative frequency of haplotype *i.* This calculation was carried out using DnaSP version 5.0 (Librado & Rozas 2009).

The measure of nucleotide diversity is representative of the average number of nucleotide differences per site and was carried out according to the following formula (Nei 1987). :

$\pi = k/m$

Where *m* is the total number of nucleotide positions inclusive of both variable and monomorphic sites, but excluding alignment gaps and *k* is the mean number of nucleotide differences calculated as follows:

$$k = \frac{2}{n(n-1)} \sum_{i<j} d_{ij}$$

Where *n* is the number of nucleotide sequences and d_{ij} is the number of nucleotide differences between sequences *i* and *j.* This calculation was also carried out in the software DnaSP version 5.0 (Librado & Rozas 2009).

The mean number of pairwise differences between all pairs of haplotypes was calculated in Arlequin version 3.5.1.3 (Excoffier & Lischer 2010) as follows (Tajima 1983) :

Mean No. Pairwise Diffs. $= \frac{n}{n-1} \sum_{i=1}^{k} \sum_{j=1}^{k} p_i\, p_j \hat{d}_{ij}$

Where $\hat{d}_{ij}$is an estimate of the number of mutations that have occurred since the divergence of haplotypes *i* and *j, k* is the number of haplotypes, p_i is the frequency of haplotype *i* and *n* is the sample size.

To ascertain if there is any differentiation between the three groups a test based on F_{ST} was calculated. Genetic distances were computed as pairwise F_{ST} measures between populations and in accordance with Reynold *et al.* (1983) version for a short divergence time which was applicable to the groups under assay here:

$$F_{ST} = 1 - (1 - \frac{1}{N})^t \approx 1 - e^{-t/N}$$

In order to estimate evolutionary distance and to visualise how haplotypes derived from the three *H. moloch* groups clustered, a phylogenetic species tree was created using MEGA version 5.2.2 (Tamura et al. 2013). The algorithm chosen was the maximum likelihood method. A constraint of the maximum likelihood model is that whilst the algorithm searches for the most applicable phylogenetic tree from the data presented, it is unable to search within every possible scenario of evolutionary relationships. As the number of taxa input increases, the number of possible outcomes increases greatly and renders it extremely difficult to compute every possible topology (Vandamme 2009). Preliminary tests however, showed that the percentage of branches that clustered together utilising the maximum likelihood was greater than by application of other algorithms (neighbour joining and UPGMA). UPGMA does not allow for the inclusion of an evolutionary model, which is also the case for maximum parsimony methods. The maximum likelihood algorithm does allow this and thus it was a further factor contributing to the decision to use this approach. This was pertinent for the mtDNA HV-I region sequences, as they are

known to evolve at a faster rate then in nuclear DNA and other genes within the mtDNA molecule. A further important consideration to take into account within the analysis methodology was to acknowledge the inequality of the frequency of transitions and transversions that occur in mtDNA. There is a large disparity between the occurrences of transitions (purine to purine or pyrimidine to pyrimidine), which are far more common, to the occurrence of transversions (purine to pyrimidine and vice versa) in primates (Hasegawa et al., 1985; Moritz et al., 1987). In a study of both human and chimpanzee control region mtDNA sequences, 93.7% of mutations were represented by transitions (Seligmann, Krishnan & Rao 2006). Therefore, a test of 24 differing nucleotide substitution models was performed within MEGA which found that the Hasegawa-Kishino-Yano model (HKY) (Hasegawa, Kishino & Yano 1985) was the best model for the mtDNA data presented and scored the lowest Bayesian Information Criterion (BIC) score. The HKY model recognises that transversions are less frequent that transitions, and also that base pairs do not occur at an equal frequency. This HKY model is similar to the Kimura-2 parameter (Kimura 1980) model, but the latter assumes all nucleotide base frequencies are equal.

To allow for the non-uniformity of the rate of evolution of the control region of mtDNA a discrete gamma distribution with a rate of 5 (the recommended gamma values range between 4 and 8) was also applied. A heuristic methodology was also applied to aid in tree construction included within the software which will finalise computations once the superior log likelihood of the tree is computed. In this case the nearest-neighbour-interchange heuristic algorithm was applied which functions by examining space within the tree and then rearranging the tree topologies, which

is representative of the pairwise distances of underlying nucleotide data. If the resultant tree is deemed to be a better fit than the previous arrangement this tree is kept. This process continues until it reaches the aforementioned optimal criterion.

To ascertain the reliability of the tree produced by the maximum likelihood methodology the statistical non-parametric bootstrapping technique was also applied with 1000 replicates chosen which is as per programme recommendations. This technique is the most widely employed in phylogenetic analyses (Schmidt & Haeseler 2009) to test branch support within the tree. The technique approximates the underlying distribution of the data by the creation of replica data sets that are identical in size to the original data and then randomly resamples each nucleotide position column and creates a new tree. This is then repeated in accordance with the number of times input by the user. The column resampled may or may not then be used again in the pseudo sample sets that the technique creates. Once all bootstrap replicates have been carried out the reliability that each branch was found within each bootstrap replication is shown as a percentage next to each branch within the tree. There was no root utilised in the phylogram as all sequences analysed were derived from the same species.

2.3 Results

Amplified product bands in both runs did not produce multiple or unexpected bands after gel electrophoresis, which suggests that NUMTs were not present within sequences. Both experimental rounds produced identical bands for each subject animal. Furthermore, nucleotide sequences were also identical for each duplicate

PCR product for each individual. In addition to this, Wild-born and Captive-born HV-I sequences were akin to those aligned with the Wild_pop2 group data procured from GenBank.

The HV-I sequence length derived from the *H. moloch* individuals (and then subsequently applied to the Wild_pop2 group) totalled 518 bp inclusive of gaps. This size is similar to that found in five *Nomascus* gibbon species (*concolor, leucogenys, nasutus, gabriellae* and *hainanus*) by Monda *et al.* (2007) of 477bp and to a phylogenetic study of the mtDNA control region of the *Hylobates* genus where sizes ranged between 487-520 bp dependent on the species (Whittaker, Morales & Melnick 2007). One sample (sample h.mol 17 from the Captive group) was excluded from further analyses as the resultant amplified fragment length was considerably shorter after both PCR reactions. It is thought that sequencing of the HV-I segment was unsuccessful in that particular individual.

2.3.1 Genetic diversity results

The measures of genetic diversity for the HV-I segment of mtDNA reveal a marked difference between captive born individuals and both wild born groups. Despite a larger number of individuals assayed within the captive-born in comparison to the wild-born groups, the number of haplotypes observed within the latter group was three times greater than the captive-born. The wild_pop2 exhibited a very high level of diversity with regards to the number of haplotypes and uniqueness to each individual. With just 2 haplotypes in the Captive group it is perhaps not surprising that *H* exhibited a much lower value than within both wild groups. At the nucleotide

level, the disparity between wild born and captive born is also evident. A total of just 4 polymorphic sites were observed within the Captive group and this low level of nucleotide variability is further illustrated in the value of π. The greatest value of π was evinced in the 31 individuals in the wild_pop2 group and followed by the 8 in the wild-born group. The greatest number of polymorphic sites was observed in the wild_pop2 group which was not surprising considering the high number of haplotypes harboured within this group. Unique to this group was also the presence of both transitions and transversions, an occurrence not observed within the other two groups. A factor, further illustrative of uniqueness within the wild_pop2 group is the number of private substitution sites which totalled 78, however only 1 private site was observed in the wild-born group and no private sites were evinced in the captive-born. Finally, the greatest value of pairwise differences between haplotypes was detected in wild_pop2. The group with the least value of pairwise differences was illustrated in the captive-born group, an unsurprising result considering that this group harbours just two haplotypes. The wild-born group yielded a high value when considering the total number of haplotypes observed within this group was 6.

There is no differentiation between the wild-born group and wild_pop2 (Table 2). There is significant differentiation, however, between the captive-born group with both wild groups.

Table 2 - HV-1 measures of mtDNA genetic diversity. *n* is the number of individuals, *H* is gene diversity and π is a measure of nucleotide diversity.

	Captive-born	Wild-born	Wild_pop2
n	12	8	31
No. of haplotypes	2	6	29
H	0.409 ± s.d (0.133)	0.929 ± s.d (0.086)	0.998 ± s.d (0.009)
π	0.003 ± s.d (0.001)	0.023 ± s.d (0.005)	0.036 ± s.d (0.005)
No. of polymorphic sites	4	30	114
Transitions	4	30	83
Transversions	0	0	31
Mean pairwise differences	1.636 ± s.d (1.036)	12.071 ± s.d (6.116)	20.652 ± s.d (9.373)

Table 3 - Measures of population differentiation using F_{ST}. (*) indicates statistical significance *p<0.05*

	Wild_pop2	Wild-born	Captive-born
Wild-born	0.008		
Captive-born	0.187 (*)	0.225 (*)	

There were a total of 34 haplotypes observed over all three *H. moloch* groups. There is only one haplotype common to all three, haplotype 26 which was observed

in 9 individuals from the captive-born group, 2 from the wild-born group and 1 within wild_pop2. The only other haplotype shared with wild_pop2 is number 18 which is present within 2 individuals from the wild-born group and 1 from wild_pop2. Haplotype 21 is shared with three *H. moloch* from captive-born and 1 individual from the wild-born groups.

The MK values for wild and captive-born are shown with the mtDNA found within each individual. The latter part of the table from H.mol 20 downward illustrate the most common haplotype within the two groups and individuals have the greatest MK values in this part of the table also. Rare alleles highlighted in yellow however appear in individuals with an intermediate value such as in H.mol 2 (Table 5).

Tests of neutrality, *Tajima D* and Fu's F_s tests are positive for both the captive-born and wild-born groups (Table 6). The wild_pop2 results, however, are negative, with a statistical significant value in the Fu's F_s test.

Table 4 - HV-I mtDNA haplotype frequencies occurring in all three groups.

Haplotype Number	Captive-born	Wild-born	Wild_pop2
Hap_1	0	0	0.032
Hap_2	0	0.125	0
Hap_3	0	0	0.032
Hap_4	0	0	0.032
Hap_5	0	0	0.032
Hap_6	0	0	0.065
Hap_7	0	0	0.032
Hap_8	0	0	0.032
Hap_9	0	0	0.032
Hap_10	0	0	0.032
Hap_11	0	0	0.032
Hap_12	0	0.125	0
Hap_13	0	0.125	0
Hap_14	0	0	0.032
Hap_15	0	0	0.032
Hap_16	0	0	0.032
Hap_17	0	0	0.032
Hap_18	0	0.250	0.032
Hap_19	0	0	0.032
Hap_20	0	0	0.032
Hap_21	0.250	0.125	0
Hap_22	0	0	0.032
Hap_23	0	0	0.032
Hap_24	0	0	0.032
Hap_25	0	0	0.032
Hap_26	0.750	0.25	0.032
Hap_27	0	0	0.032
Hap_28	0	0	0.032
Hap_29	0	0	0.032
Hap_30	0	0	0.032
Hap_31	0	0	0.032
Hap_32	0	0	0.032
Hap_33	0	0	0.032
Hap_34	0	0	0.032

Table 5 - mtDNA per individual from captive and wild-born groups and their mean kinship (MK) values. Table sorted according to lowest MK value to greatest. Highlighted cells represent unique mtDNA haplotypes

Primate Ref	MK	mtDNA Haplotype
H.mol8	-	Hap_21
H.mol21	-	Hap_26
H.mol17	-	Hap_26
H.mol6	**0.0000**	Hap_18
H.mol4	**0.0000**	Hap_26
H.mol3	**0.0000**	Hap_18
H.mol5	**0.0083**	Hap_2
H.mol1	**0.0083**	Hap_12
H.mol7	**0.0208**	Hap_26
H.mol15	0.0604	Hap_21
H.mol16	0.0688	Hap_21
H.mol2	**0.0875**	Hap_13
H.mol13	0.1083	Hap_21
H.mol20	0.1167	Hap_26
H.mol19	0.1167	Hap_26
H.mol14	0.1271	Hap_26
H.mol12	0.1271	Hap_26
H.mol11	0.1271	Hap_26
H.mol9	0.1354	Hap_26
H.mol10	0.1479	Hap_26
H.mol18	0.1521	Hap_26

Table 6 - Results of Tajima D and Fu's F_s neutrality tests. (**) indicates statistical signficance *P<0.001*

	Captive-born	Wild-born	Wild_pop2
Tajima *D* test	0.828	0.115	-1.139
Fu's F_s test	3.698	1.31	-12.732 (**)

2.3.2 The phylogeny of mtDNA HV-I within Captive and Wild haplotypes

Basal sequences within the topology are derived from two wild_pop2 sequences with a 99% probability after 1000 bootstrap replications. Sequences from wild_pop2 are largely clustered at the base of the tree, however there is one sequence from a wild-born individual that forms a clade with another wild_pop2 member. There was an 85% probability that all three groups clustered within one clade which is presumably representative of Haplotype number 26, where 9 captive-born individuals share the same haplotype with 2 wild-born and one individual from wild_pop2. Some nodes in the tree have yielded confidence levels below 50%. However, the overall topology appears to support results procured within the genetic diversity measures (section 2.3.1).

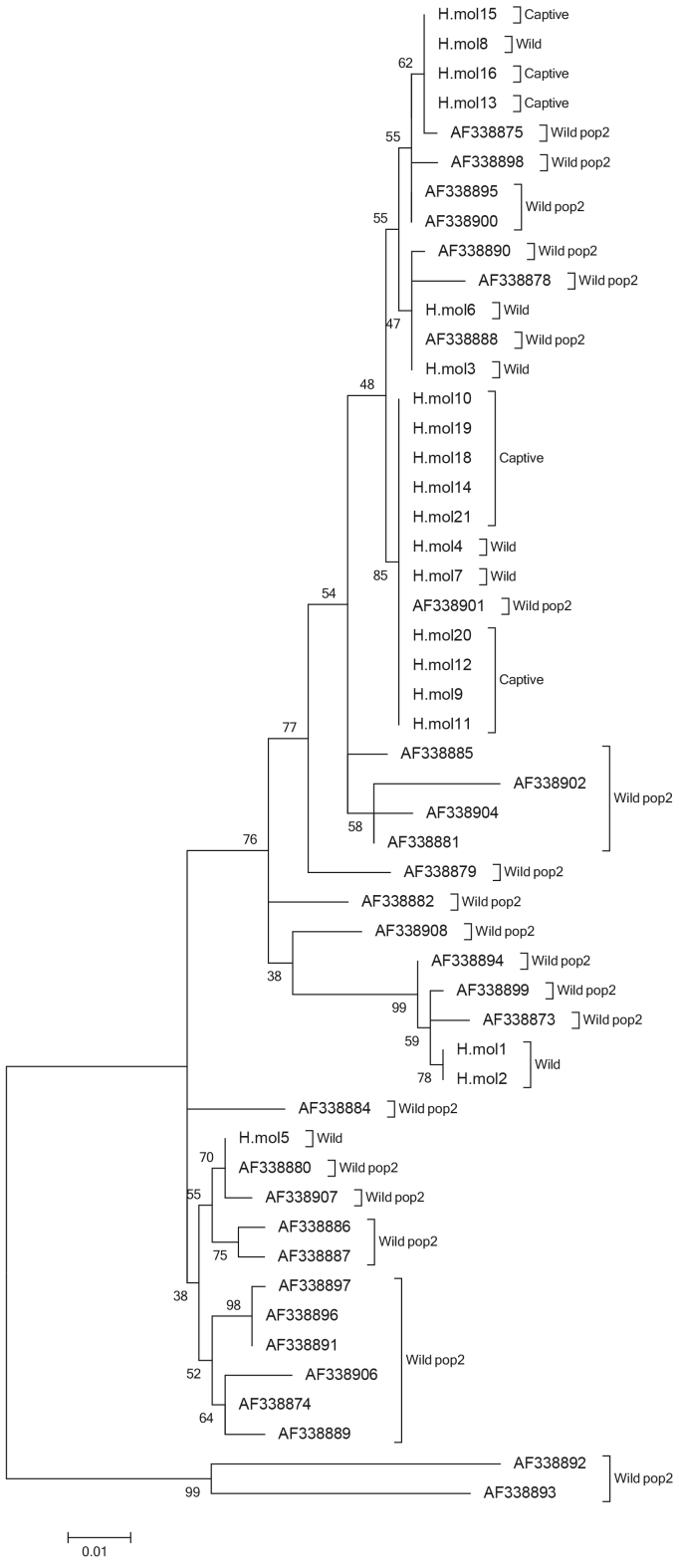


Figure 2 - Maximum likelihood phylogenetic tree illustrating evolutionary distances of mtDNA HV-I sequences. Nucleotide substitution model used was Hasegawa-Kishino-Yano. Percentage of replicate trees is shown next to each branch (1000 bootstrap replications).

2.4 Discussion

The amplification of the mtDNA HV-I region within the *H. moloch* population consisting of the two groups, wild born (Wild) and captive born (Captive) was successful in all but one (H.mol17) of the 21 individuals. The results obtained, show a disparity between wild born and captive born groups with regards to levels of mtDNA diversity within the HV-I region.

A clear lack of variability in the number of haplotype is illustrated in the Captive group in comparison to both the wild-born and wild_pop2 groups, with 2 haplotypes detected in 12 captive, 6 in 8 wild and 29 in 31 wild_pop2 individuals. Low mtDNA haplotype counts have been observed in a number of species and the occurrence is attributed in some cases to female philopatry. For example, the bull shark (*Carcharhinus leucas*), a species that enjoys a wide distribution across the globe was found within two populations (*n*=17 and 23) analysed to yield a total of 4 mtDNA control region haplotypes (Karl et al. 2011). The apparent lack in maternal gene flow was in large attributed to a preference in females to remain in particular nursing areas, both mother and female progeny. Similarly, the Brazilian stingless bee (*Plebeia remota*) is known to exhibit low female dispersal pattern within females and the outcome of this behaviour was shown to impact on mtDNA haplotype counts with only 1 haplotype occurring within two populations (*n*=12 and 13) (Francisco, Santiago & Arias 2013). The American alligator (*Alligator mississippiensis*) was found to harbour just 3 haplotypes within the control region of mtDNA (*n*=25) and one of the contributing factors was deemed to be female philopatry to particular areas as with the bull shark (Glenn et al. 2002). Fidelity to a particular area within a habitat is one observation that has been purported to maintain mtDNA haplotypes at

a low count: however, ecological barriers also serve as a limiting factor with regards to variability at this genetic region. One such example is provided by a study into an endangered freshwater Saimma ringed seal (*Phoca hispida saimensis*) (Valtonen et al. 2012). This seal species is an endemic organism found only in a landlocked Lake in Finland. The ringed seal population consisting of fewer than 300 individuals is believed to have been genetically isolated since the last ice age. A study of mtDNA variability found 8 haplotypes in the control region of 215 individuals. This number is stated within the study to be low, however it is evidently greater than the Captive and Wild populations of *H. moloch* assayed here. Although haplotype counts may increase with an increase in sample sizes across all groups, the ratio of haplotype count to number of individuals is further confirmation that captive-born group lack diversity at this mtDNA segment (captive-born = 0.16, wild-born, 0.75 and wild_pop2 0.93).

The low count exhibited within the Captive group of *H. moloch* here however, can not be attributed to such an occurrence. This is postulated to have occurred as a consequence of the limitations of captive breeding, which often involves a limited number of founders contributing wild type alleles when dealing with endangered species, such as here with *H. moloch* (Hedrick et al. 1997). With a limited number of females available, and if those females do not fall within the prerequisites of desired mean kinship values, this number may be even smaller and thus mtDNA gene flow is likely to decrease. This occurrence has also been reported in studies that have compared captive bred individuals to their wild conspecifics with captive groups exhibiting a lower genetic diversity. A study of a group of captively bred and wild born Elliot pheasants (*Syrmaticus ellioti*) exhibited similar results to this study and

results were perhaps more akin to the Wild_pop2 and Captive *H. moloch* haplotype count results (Ping-ping et al. 2005). Captive pheasants harboured 3 haplotypes which was observed within 36 individuals, however their wild counterparts, a group that totalled 17 individuals, exhibited 16 haplotypes that were unique to each individual. Further observations of disparity between wild and captive groups of the same species have been made in the wild Matschie tree kangaroo from Papua New Guinea (*Dendrolagus matschiei*) (McGreevy Jr et al. 2009), African giraffe (*Giraffa camelopardalis*) (Hassanin et al. 2007) and mallard (*Anas platyrhynchos*) (Čížková et al. 2012).

The Captive group exhibited the least number of haplotypes and of the 12 individuals assayed, 3 possess Hap_21 and the remaining 9 individuals maintain Hap_26. Interestingly, this common haplotype within the Captive group is also present not only in the Wild group which was to be expected as haplotype number 26 is also present in a dam derived from the Wild group, but it is also present in Wild_pop2. The origins of individuals from the Wild_pop2 group are unknown, however the persistence of Hap_26 across all three groups suggests a level of matrilineal connectivity between the three groups. This phenomenon has been observed in other species and subspecies. In 16 populations, the koala (*Phascolarctos cinereus*) was observed to share particular haplotypes across differing geographical regions, although the species was thought to have had limited gene flow between the populations (Houlden et al. 1999). Observations of a persistent mtDNA haplotype across species and sub-species taxonomic levels has also been observed within the Mexican wolf (*Canis lupus baileyi*) (Hedrick et al. 1997), tiger (*Panthera tigris*) (Luo et al. 2004) and mahseer fish (*Tor tambroides* and

Tor douronensis) (Nguyen et al. 2006). With an identical haplotype occurring across sub-species and within same species within differing geographical areas, it may be surmised that this occurrence is derived from a more ancient lineage and from previous distributions of the species in question (Hedrick et al. 1997).

Uniqueness is clearly the greatest in the Wild_pop2 group as only one haplotype (Hap_6) is shared and only between two individuals. All other 28 haplotypes within this group are unique to a particular individual. This pattern is similar to the Wild group, where 2 haplotypes (Hap_18 and Hap_26) are shared with 2 intra-group individuals, the remainder are unique to an individual as within Wild_pop2. This occurrence may be attributed to the social behaviour of the gibbon, whereby in the wild, offspring within a group will disperse and locate a new territory where they will seek to pair with a mate and reproduce themselves once sexual maturity is reached (Brockelman et al. 1998). However, in captivity the potential partners available for mature offspring are limited with only 48 individuals thus the gene pool is already small. In addition, *H. moloch* is a Cites Appendix I species, therefore procuring individuals from the wild to supplement mtDNA variability for captively bred individuals is prohibited (IUCN 2008).

When viewing the mean kinship values with the haplotypes per individual there is a risk that a rare haplotype be lost if MK values were the only factor employed when selecting individuals for breeding. A wild individual harbouring a unique haplotype has an MK value of 0.0875 and yet another wild individual harbours the most common haplotype yet has an MK value of 0.000.

The group with the least number of variable sites and subsequently the lowest value of π, was the Captive group with 4 polymorphic sites and 0.003 (±0.001) for nucleotide polymorphism. Even when considering that the number of haplotypes within this group totalled just 2, the level of diversity at the nucleotide level with only 4 variable sites is clearly extremely low. The two groups representative of wild alleles differed at both nucleotide values, with the Wild group exhibiting 30 variable sites, all consisting of transitions but the Wild_pop2 group had 114 variable sites and 31 of those were transversions. This number of transversions observed is representative of approximately 27% of variable sites, a value that is greater than reported in chimpanzee and humans (Seligmann, Krishnan & Rao 2006; Tamura & Nei 1993). The control region is not a coding segment of mtDNA, therefore the transversions will not produce altered amino acid sequences and therefore proteins. However there may be an evolutionary function as the region is responsible for regulating transcription and replication (Sbisà et al. 1997).

The values for nucleotide diversity (π) were 0.023 (±0.005) and 0.036 (±0.005) for the Wild and Wild_pop2 groups respectively, thus a lower value was illustrated in the *H. moloch* group assayed here in comparison to the larger population within Wild_pop2. There may be an element of bias with regards to the sample number which is low in the Wild group and if looking at the large differential between haplotype sequence numbers it is not surprising that the Wild_pop2 group yielded larger values of diversity. However, the differential between the two π values is not particularly large, and when viewed in conjunction with gene diversity (0.929 for wild-born and 0.998 for wild_pop2 individuals), the values are significantly different and are further corroborated by analysis of F_{ST} measures. If taking just the Wild_pop2

nucleotide diversity measure of 0.036, it is an extremely similar value to that observed in a group of *Hylobates klossii* that is also an endemic gibbon species, but to the Mentawi islands in Indonesia (Whittaker 2005). A measure of π within the entire control region of *H. klossii* gibbons from North Siberut found that nucleotide diversity equalled 0.0314 (±0.024) from 3 haplotypes (*n*=3) (Whittaker 2005). As a larger segment was analysed within the *H. klossii* group this may have allowed for greater nucleotide variability.

In comparison to the Saimaa ringed seal (*Phoca hispida saimensis*), a species depauperate of mtDNA genetic variability, *H. moloch* variability (Valtonen et al. 2012). Similarly, the cheetah was found to harbour 7 haplotypes the nucleotide diversity of which was 0.0018 (Menotti-Raymond et al. 1999). These values are more akin to the captive-born *H. moloch* assayed here (0.003), which certainly appears to illustrate a much lower variability and a contrast to their Wild and Wild_pop2 conspecifics. This differentiation at the nucleotide level was also observed within captive born and wild born mallards (*Anas platyrhynchos*) where π was found to be 0.007 within the wild individuals but it was much lower with 0.002 within the captive born group (Čížková et al. 2012).

The wild_pop2 group deviated from the expected neutral equilibrium in the *Fu's* F_s test. This test incorporates haplotypic frequencies in its computations and is stated to be the most powerful when analysing regions of non-recombining DNA (which has been assumed to be the case in the HV-I region assayed here) in detecting alterations in the demographics of a species (Ramírez-Soriano et al. 2008). A

negative result from the *Fu's* F_s test can be interpreted in a number of ways, it may be indicative of a selective sweep that has occurred throughout the population, it may also be symptomatic of an expansion in population numbers (de Jong et al. 2011; Houlden et al. 1999). With the loss of habitat that *H. moloch* has experienced, population numbers have declined, and are purported to continue to decline (IUCN 2008). Genetic variability has the potential to be influenced by such environmental pressures, as populations that previously enjoyed a continuous distribution are forced into fragmented units (Moreira et al. 2010). Therefore, this is an unlikely interpretation of neutrality results, and if further considered in conjunction with the number of transversions for wild_pop2 then a selective sweep may be the more applicable hypothesis. If a selective sweep is indeed exerting positive selection on the mtDNA molecule then it would be detected within this genetic region alone, however if there has been a rapid expansion of population numbers within the *H. moloch* habitat then it would be detected within other loci. The final test illustrative of mtDNA genetic disparity between captive born and both Wild groups is shown in population differentiation statistic (F_{ST}).

2.5 Conclusions

Analysis of mtDNA HV-I variability and comparisons of population differentiation between the three groups, show that the captive-born group of *H. moloch* differs from its wild conspecifics and harbours a low variability at this genetic region. Not only did the haplotype count within the 12 individuals analysed yield a low number of 2, but the nucleotide sequences themselves show a great deal of homogeneity. The disparity shown within the genetic diversity results between the Captive group in comparison to both Wild groups is then confirmed by the population differentiation statistic of $F_{ST.}$

From the limited studies available that compare wild and captive born species at the mtDNA molecule, it appears that this phenomenon is not exclusive to the *H. moloch* species analysed here (e.g. Čížková et al., 2012; Luo et al., 2008; Miller et al., 2009; Ping-ping et al., 2005; Ray et al., 2013). In most examples listed, there is a marked difference between wild and captive individuals with the latter exhibiting a lower diversity than the former. However, efforts have been made in some breeding programmes to reverse the trend of a decline in mtDNA diversity by specifically addressing the issue within their captive breeding objects. One such success story is exemplified in the tiger (*Panthera tigris*) whereby a worldwide collaboration of zoological institutions allowed for the introduction of between 1 and 10 novel haplotypes within the differing subspecies assayed (Luo et al. 2008). This focused approach of breeding has resulted in a higher number of captive mtDNA haplotypes in comparison to wild tigers studied (Luo et al. 2008). Thus, theoretically the captive tiger population has the potential to bolster diversity as a metapopulation to populations within the wild.

Although the example of the tiger breeding programme illustrates how the incorporation of molecular management can improve genetic variability, it is not advocated here that *H. moloch* be bred solely to increase mtDNA variability, and regard should also be given to markers derived from the nuclear DNA to complement the mtDNA data. Although results showed that mtDNA analysis provided a greater scope of information as opposed to using MK values it is still advisable to look at nuclear markers. This would allow for a wider understanding of the genome wide view of this species. Thus, although it is evident that an increase of mtDNA variability is required for future generations of captively born *H. moloch* if an effort to mirror the wild type data is desired, it should not be effected at the detriment of nuclear DNA variability.

3 Chapter three - Analysis at the Nuclear DNA Level - Using Microsatellites to Estimate Genetic Diversity

The next stratagem to further analyse the genetic variability of the *Hylobates moloch* population was to look beyond mtDNA and see if observations made within the maternally inherited molecule were also occurring within both neutral and adaptive regions within nuclear DNA. This chapter evaluates diversity from a neutral nuclear marker.

Microsatellites, also referred to as short tandem repeats (STRs) or simple sequence repeats (SSRs), are tandem repetitions of DNA motifs consisting of one nucleotide base pair up to six (Buschiazzo & Gemmell 2010). Unlike mtDNA, microsatellites are inherited on a Mendelian basis and therefore provide genetic information from both the sire and dam in a subject. They have been identified within both eukaryotes and prokaryotes (Buschiazzo & Gemmell 2010) and in eukaryotes, exhibit a vast and scattered distribution throughout the genome as a whole (Crouau-Roy & Clisson 2000; Kelkar et al. 2010). It has been postulated that a microsatellite loci is present every 6Kb (Zhang et al. 2001) however there are reports that this number could be even greater at every 2Kb (Guichoux et al. 2011) and in total they constitute approximately 3% of the entire human genome (Kelkar et al. 2011). In addition to their abundance, there is further appeal in their use as a molecular marker within the field of conservation genetics as a consequence of the high level of mutation that can be found within these co-dominant markers. Mutation rates have been reported to occur at 10^{-3} per locus, but further still that this can occur at every generation (Leclercq, Rivals & Jarne 2010). Mutation of microsatellite loci

takes the form of either insertions or deletions of one or more tandem repeats. The underlying mechanisms that effect these genetic alterations are not known exactly, but it is thought that mutation occurs as a consequence of a combination of DNA replication slippage and perhaps to a lesser extent, recombination events (Leclercq, Rivals & Jarne 2010). DNA slippage is the process whereby repeats are added if newly synthesised DNA dissociates from the template DNA or lost if the template DNA chain dissociates (Chambers & MacAvoy 2000). Viguera *et al.* (2001) stated that this occurrence takes place as a consequence of DNA polymerase pausing mid-replication of a microsatellite region, which effects dissociation from the DNA. The newly synthesised DNA then realigns itself with another repeat and the DNA polymerase resumes replication.

Microsatellites follow a process that is akin to a life cycle, that is that they are born, undergo a period of maturation and then die (Kelkar et al. 2011). The birth is representative of the appearance of tandem repeats that reach a liminal number that constitutes the definition of a microsatellite, the threshold of which remains under discussion. Some authors designate the minimum definition of a microsatellite as 8 nucleotides (e.g. Chambers and MacAvoy, 2000), consisting of dinucleotide motifs. Others class a microsatellite as one that is between 7 and 9 nucleotides in a mononucleotide locus and between 4 to 8 for di- to tetra-nucleotide loci (Kelkar et al. 2011). Once the minimum threshold has been met, the maturation phase is exhibited by elevated rates of DNA slippage (Kelkar et al. 2011). The death of a microsatellite occurs when deletion events culminate in the loci contravening the minimum threshold prerequisite and thus slippage rates are no longer sustained.

A long-standing assumption with regards to the mutability of microsatellites is that they evolve on a neutral basis and are not actively involved in actual gene function (Buschiazzo & Gemmell 2010). However, the role of this form of DNA sequence remains unresolved. For example, dinucleotide microsatellites in human DNA appeared to show microsatellite mutability in a more systematic manner than would be expected under a neutral basis, with a greater tendency of mutations occurring at the 3' end of the loci (Varela et al., 2008; Viguera et al., 2001). This observation suggests that alterations are not random but have a specified direction within the loci. Withal, reports of microsatellites mutating under possible selective pressure rather than as a consequence of random genetic drift have been suggested. For example, a study of levels of infection with *Plasmodium falciparum* in children residing in Gabon, researched the role of reactive oxygen intermediates, which are known to be key players in the cellular nonspecific innate immune response (Uhlemann et al. 2004). It was found that infection levels of *Plasmodium falciparum* which were designated within the study as severe and mild were dependent on a microsatellite positioned upstream of the promoter region for the NADPH oxidase *gp91^{phox}* subunit. It was deduced that the length of the microsatellite in question was responsible for regulating the NADPH oxidase activity, with a shorter TA_{11} offering greater protection against malaria versus TA_{16} repeats. Further correlations, in humans, have been made between tri-nucleotide repeat loci and disease (Kruglyak et al., 1998; Rubinsztein et al., 1999; Varela et al., 2008) and examples of selection acting on microsatellites are highlighted by Galindo *et al.* (2009). Within the cases reviewed links are made between microsatellite variation and animal morphology to polymorphisms influencing both human and non-human primate behaviour (Galindo et al. 2009).

Differences in allele lengths within orthologous microsatellites loci have been observed in cross species comparisons among human and non-human primates (Galindo et al., 2009; Lathuilliere & Crouau-Roy, 2000; Sainudiin et al., 2004). Microsatellite lengths within humans are comparatively longer than detected in non-human primates and a number of explanations to account for this phenomenon have been suggested. One theory postulates that within humans the DNA slippage rate is higher owing to a poorer functionality of DNA polymerase resulting in a longer microsatellite length. As a microsatellite enters the maturation phase and experiences a higher rate of DNA slippage, the rate at which this occurs within monomorphic microsatellites has been reported to be lower than observed in polymorphic loci (Chambers & MacAvoy, 2000; Rubinsztein et al., 1999). Whilst a locus within one species may be polymorphic, it is not certain that this will be the case within another species. Thus it would be expected that a polymorphic loci be longer than its equivalent in another species that exists as a monomorphic locus.

A number of mutational models have been devised as follows, all of which are based on the assumption that microsatellite are neutral markers and encompass the varying forms of mutation.

The Infinite Allele Model (IAM) - (Kimura & Crow 1964)

This theoretical model is the most simple. It functions under the assumption that every novel mutation results in an allele that has not existed in the population previously. Thus for each mutation a new allele is created (which has an infinite number of possible states), but it is counterbalanced by the effects of genetic drift,

thus creating an equilibrium. Under the IAM microsatellites mutations can culminate in an increase or reduction of one or more motif repeats.

***K*-Allele Model (KAM) -** (Crow & Kimura 1970)

This model poses a more restrictive approach. Here, the number of possible allelic states are limited to '*k'*, each with an equal probability that it can mutate to any other of the existing allelic states.

Stepwise Mutation Model (SMM) - (Kimura & Ohta 1978)

The stepwise mutation model assumes that mutation occurs only within adjacent states and can manifest as an addition or a deletion of one repeat. The assumption is that addition or deletion occur at equal frequencies. Unlike the IAM, SMM mutations are not all presumed to create a novel allele within the population but may manifest as one already present.

Two Phase Model - (Di Rienzo et al. 1994)

This model functions as a two-step approach. The two phases allow for the possibility that a microsatellite loci may mutate via one repeat length, or with multiple repeats. The first phase of the model has a probability (*p)* that the descendent allele has mutated via one repeat, whether it is an addition or a deletion from its ancestral state. The probability in the second phase has a probability of *p-1* that the increase or decrease in allele length is longer than one repeat, the number of which is deduced by a prior specified distribution of alleles within the population.

The development of models to better understand the evolution and nature of microsatellites are invaluable for downstream analysis and the derivation of

biologically meaningful results. The theories are not an exhaustive list of microsatellite models, since their inception there are been alterations and updates to try and encompass the variable nature of these nuclear markers (Chambers & MacAvoy 2000). However, it is this very complexity that renders modelling microsatellites a very difficult task. The first difficulty lies in the fact that mutations do not occur exclusively as a one-step increase or decrease (Haasl & Payseur 2011) therefore the stepwise mutation model may not be sufficient to analyse all loci. Although it is thought that the majority of observed alterations in length do occur as a consequence of one deletion or addition (Kruglyak et al. 1998). Further complications are that that slippage rates and thus microsatellite lengths vary between species, and polymorphic markers mutate at a higher rate than monomorphic loci (Kruglyak et al., 1998; Sainudiin et al., 2004). Both the stepwise mutation model and the two-phase model function on the basis that microsatellite loci mutate at a constant rate and moreover that they can increase or decrease without any upper or lower limitations of size. It is evident that microsatellites do not expand exponentially as otherwise there would be a plethora of extremely long loci, but at the same time adding an upper band limit within a model is a difficult task as the value may not be a true representative of the loci in question and constrain results if in fact it can mutate longer than a stipulated limit (Sainudiin et al. 2004).

3.1 The advantages and disadvantages of microsatellites in analysing genetic variation and ascertaining pedigree

In addition to analysing genetic variation the nuclear DNA marker was required to be able to evaluate the pedigree of the population. Genetic testing of individuals may not only reinforce information already gained of a population but also, provide additional data that can serve to alter studbook entries (*e.g.* Jones et al., 2002). *H. moloch* is a monogamous species. This is also true of *Hylobates muelleri.* However, a study of the parentage of a group of wild *muelleri* gibbons found that in two of five groups studied, one subadult in each was not a descendent from the male or the female in their respective groups (Oka & Takenaka 2001). Thus, in this case it appears that a group will allow a subadult gibbon that is completely unrelated to live with them. There have been reports of extra-pair matings in parbonded species, including birds (Griffith, Owens & Thuman 2002), or in polygynous mating systems such as those evinced in *Gorilla gorilla* (Vigilant & Bradley 2004). Furthermore, captive individuals within a population may have originated from the wild and are therefore of unknown origins with regards to their parents. This is the case for the *H. moloch* population. Within the captive breeding programme remit, founder pairs are deemed to be unrelated, however this is not genetically verified in most cases.

SNPs are present within both introns and exons of DNA and are a difference in a single nucleotide within a sequence and thus can be a change of up to four differing base pairs. They have been heralded as a successor marker from microsatellites, albeit within large-scale studies and in model organisms (Varela & Amos 2010). One of the reasons for this apparent surge of appreciation and interest in this marker

is their ubiquitous presence throughout the mammalian genome and is representative of approximately 90% of human DNA sequence polymorphism (Varela & Amos 2010). It is estimated that a SNP occurs at spacing of between 200 to 500 base pairs in many species (Morin et al. 2004). A further attraction of using SNPs is that as opposed to the numerous models that have been created to accommodate microsatellite mutations, SNPs are purported to evolve in accordance with a simple model such as the infinite sites model (Morin et al. 2004).

However, several factors make microsatellites more applicable to the questions involved in this project and in addition, to the population under assessment. In the first instance, *Hylobates moloch* is not a model organism. It is an Endangered species existing in limited numbers both within captivity and in the wild. As is the case with both the *moloch* species and other species within the *Hylobates* genus, there are few genetic data. Although SNPs may be more widely distributed throughout the mammalian genome than evinced by microsatellites, a great deal of this information is derived from humans. SNPs located within the human genome may not necessarily be present at the same loci in other species (Morin et al., 2004; Varela & Amos, 2010). This renders replication from human data to other species such as the endangered *Hylobates moloch* more difficult. In order to locate a sufficient number of SNP locations (the number of which depends on the information that is desired to be procured from the data) a screening process would need to be undertaken of a number of differing genome segments in the target population, the number of which Morin and colleagues (2004) state to be between seventy-five to a hundred. From this, approximately 50 SNPs may be located. Although microsatellites exist in smaller numbers in comparison to SNPS throughout the

mammalian genome, the ability to assess homologous loci across species is much greater (Buschiazzo & Gemmell 2010; Jarne & Lagoda 1996; Sharma, Grover & Kahl 2007). This potential capacity to utilise a model organism, such as humans, as a basis to locate microsatellites within a differing species is beneficial from a laboratory and a cost viewpoint. In addition primer sequences designed for human microsatellite analysis may potentially be applied in the desired target species. Therefore, the creation of species specific primers and of course subsequent testing within individuals may not be required, thus keeping costs at a minimum. The success rate of utilising primers across species is deemed to decrease in accordance with an increasing evolutionary distance between target species (Jarne & Lagoda 1996). With regards to primate taxa, cross species amplifications are frequently carried out utilising human loci and their respective published primer sequences (Arandjelovic et al. 2010; Bradley, Boesch & Vigilant 2000; Goosens et al. 2000; Zhang et al. 2001). However, success rates either decrease or are completely absent when utilising the same loci applied to New World Monkeys (Coote & Bruford 1996). This phenomenon is not exclusive to applications within primate studies, microsatellite primers designed for a particular species have been reported to be efficacious in cross species studies where the evolutionary distance between the subjects is not too disparate. Examples have been made in a variety of differing species such as, snakes (Bushar, Maliga & Reinart 2001), nematodes (Temperley et al. 2009), fish (Scribner, Gust & Fields 1996) and cetaceans (Coughlan et al. 2006).

Quantifiable variation within microsatellites is potentially much greater than illustrated in SNPs. SNPs are low information markers that exist in a diallelic form (Haasl & Payseur 2011). In contrast, microsatellites may contain a high number of alleles within a population, often in excess of double figures (Guichoux et al. 2011). This provides a more effective measure of genetic variation, allowing a greater depth of genetic variation analysis. Phenomena such as allelic richness or uniqueness within a population are more readily apparent with microsatellites (Guichoux et al. 2011). Furthermore, as a consequence of the potential high number of alleles the statistical power of microsatellites for pedigree analysis is greater than can be derived from SNPs. To ascertain paternity of a population and reach an equilibrium of statistical power in differentiating it, a set of between three (with expected heterozygosity >0.8 and a minimum of 7 alleles per locus) (Hübner et al. 2013; Rodriguez-Barreto et al. 2013) and fourteen microsatellites (with an expected heterozygosity of between 0.6-0.8) are stated to be sufficient (Morin et al. 2004). However, to obtain the same information utilising SNPs, a panel of between forty and one hundred (with an expected heterozygosity between 0.2-0.4) would be required. A further factor that would be of benefit for a conservation programme that incorporates a re-introduction of individuals from a captive environment, is that microsatellites have the ability to provide information as to the individual identity of a subject. The utilisation of SNPs for this purpose is 2 to 4 times less powerful than employing microsatellites (Haasl & Payseur 2011; Morin et al. 2004).

The previously stated rate of mutation also played a role in the selection of microsatellites as the nuclear marker to ascertain genetic variation within the captive *H. moloch* group. The order of magnitude applicable to the mutation rate of SNPs in

comparison to microsatellites is much lower, with a typical rate of 10^{-9} compared to the previously quoted rate of approximately 10^{-3} (Guichoux et al. 2011; Leclercq, Rivals & Jarne 2010). The greater rate of mutation is desirable for the captive *moloch* group as the number of generations within captivity is limited. If any genetic divergence has occurred within the captive environment the microsatellite marker is more applicable to perceive this (Haasl & Payseur 2011) and furthermore genetic variability has the potential to be more easily quantified if the underlying data are more diverse.

With all the beneficial traits that microsatellites offer as mentioned above, there are also factors that render them problematic in analyses. The potential high level of polymorphism within microsatellites can convey and impart valuable information with regards to nuclear genetic variability within a population, and prove useful for the assignation of parentage. However, a phenomenon known as size homoplasy may confound the biological significance of results. Size homoplasy is particular to microsatellites, and derived from the evolutionary concept of homoplasy, which refers to a common character present within differing species but is not inherited from the same ancestral character. Similarly, size homoplasy is an occurrence within microsatellites whereby different copies of a locus are identical when observing allele length, but that the driving force behind it is not via identity by descent but rather by a mutational event (Estoup, Jarne & Cornuet 2002). Chambers and MacAvoy (2000) describe this issue as potentially serious when attempting to understand microsatellite data. The consequence of size homoplasy with regards to assessing genetic variation is that the overall allele count within a population may be lowered, and result in an overestimation of homozygous

individuals (Estoup, Jarne & Cornuet 2002; Jarne & Lagoda 1996). This may then affect estimations of genetic differentiation between populations. The most impact it may be argued, however, is on phylogenetic studies which heavily rely on data to be identical by descent rather than identical in state to allocate correct lineage within tree topologies. The only one not to incorporate size homplasy is the Infinite Allele Model because this theorem dictates that a mutation is always novel.

A further issue that is pertinent when reviewing results post-genotyping is the occurrence of stutter peaks as illustrated in Figure 3 below.

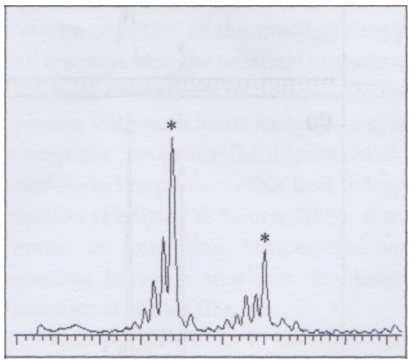


Figure 3. Image of a microsatellite profile illustrating two alleles both with stutter prior to the correct allele peak (Guichoux et al. 2011)

The presence of stutter peaks when viewing microsatellite profiles can lead to an incorrect identification of the allele by one or perhaps more repeats. This is a common problem, however, incorrect designation of an allele could result in an over estimation of alleles within a population.

Finally, further difficulties of using microsatellites to analyse pedigree and genetic variation of a population are the phenomena known as allelic dropout and false

allele identification. There is the potential to amplify PCR artefacts not related to the desired target PCR product, an occurrence which is of greater concern when utilising non-invasive samples such as faeces (Bradley, Boesch & Vigilant 2000; Gang et al. 2011; Goosens et al. 2000). The issue of both low quantity and quality of genomic DNA extracted from faecal samples have been widely reported (Nsubuga et al. 2004). Thus, it is not solely the low quantities of DNA that may lead to erroneous results, or render amplification difficult, but also the possibility that DNA from this medium may be degraded. Allelic dropout, also referred to as null alleles, is the process where only one allele is amplified during the PCR procedure in a heterozygous individual. This is exacerbated when using non-invasive mediums from which the genomic DNA is extracted. The occurrence of allelic dropout will result in false homozygotes, thus skewing data.

3.2 Materials and Methods

3.2.1 Microsatellite analysis subjects

To date, there are no available published microsatellite sequence data of *H. moloch,* therefore it was not possible within the microsatellite sample preparation to utilise a comparative outgroup of the same species. It was not deemed biologically meaningful to compare the study population of *moloch* gibbons with another species. Therefore, the sample population for the nuclear marker analysis of microsatellites was carried out solely from the *H. moloch* populations residing at Howletts and Port Lympne zoological institutions and split as in the previous chapter in to wild-born and captive-born.

3.2.2 DNA extraction

DNA was extracted as per the protocol described in chapter 2.2.2 with two faecal samples per individual.

3.2.3 Choice of microsatellite loci

Microsatellites were chosen based on the following: the chromosome on which they had been located in previous studies (to maximise overall genetic variation a number of chromosomes were wanted), the phylogenetic distance from *H. moloch* (species closest on an evolutionary scale were preferred), and the polymorphism found within previous studies (although a polymorphic locus may be monomorphic in another species) (Galindo et al. 2009). All loci tested including the primers and references are shown in Table 7. References given are for the publications in which the loci were described, the primers however were not all included within the publications as shown and thus primers were based on human microsatellite loci and retrieved from GenBank.

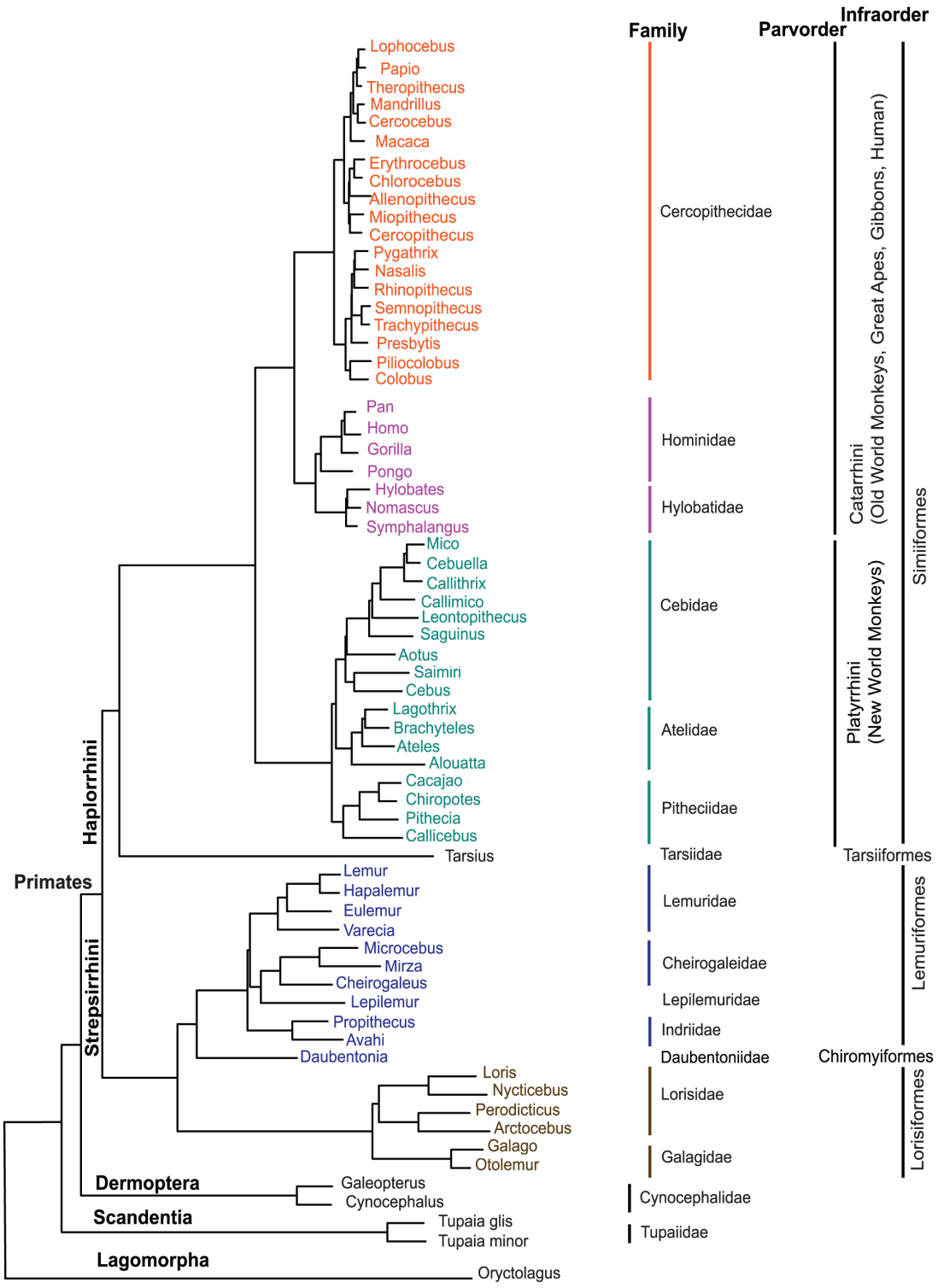


Figure 4 – The molecular phylogeny of 61 primate genera taken from Perelman *et al.* 2011.

Table 7 - Microsatellite loci analysed with primer sequences. References refer to loci tested in another primate species unless indicated with *

Locus	Primer Sequence	Reference
D1S548	*5' GAA CTC ATT GGC AAA AGG AA* *5' GCC TCT TTG TTG CAG TGA TT **	(Chambers et al. 2004)
D2S1329	*5' TTG TGG AAC CGT TCT CAA AT* *5' GAA ACT TCC ACC TGG GTT CT **	(Chambers et al. 2004)
D3S1766	*5' ACC ACA TGA GCC AAT TCT GT* *5' ACC CA ATTA TGG TGT TGT TAC C*	(Chambers et al. 2004)
D3S2459	*5' CTG GTT TGG GTC TGT TAT GG* *5' AGG GAC TTA GAA AGA TAG CAG G*	(Chambers et al. 2004)
D5S1457	*5' TAG GTT CTG GGC ATG TCT GT* *5' TGC TTG GCA CAC TTC AGG*	(Chambers et al. 2004)
D10S1432	*5' CAG TGG ACA CTA AAC ACA ATC C* *5' TAG ATT ATC TAA ATG GTG GAT TTC C*	(Chambers et al. 2004)
D13S321	*5' TAC CAA CAT GTT CAT TGT AGA TAG A* *5' CAT ACA CCT GTG GAC CCA TC*	(Chambers et al. 2004)
D20S206	*5' TCC ATT ATT CCC CTC AAA CA* *5' GGT TTG CCA TTC AGT TGA GA*	(Chambers et al. 2004)
D11S1366	*5' GCT ACA ATG ATA GGG AAA TAA TAG A* *5' GGT GGG ATC CTT TGC TAT TT*	(Whittaker 2005)
D12S391	*5' AAC AGG ATC AAT GGA TGC AT* *5' TGG CTT TTA GAC CTG GAC TG*	(Whittaker 2005)
D17S1290	*5' GCC AAC AGA GCA AGA CTG TC* *5' GGA AAC AGT TAA ATG GCC AA*	(Whittaker 2005)
D19S714	*5' ATG CCC TCT TCT GTC TCT CC* *5' GCA GAG AAT CTG GAC ATG CT*	(Whittaker 2005)
D14S306	*5' AAA GCT ACA TCC AA ATTA GGT AGG* *5' TGA CAA AGA AAC TAA AAT GTC CC*	(Whittaker 2005)
D6S2854	*5' TCA TGA GCG TGC CAC TGC AC* *5' CCG TAT ATT GCA ACC AGG AG*	(Otting et al. 2012)

D6S2859	*5' ACC CTG TCA TTC CAT GAA AC* *5' CCA CTG TTC CAG AAG CCT TG*	(Otting et al. 2012)
D7S513	*5' CAG GAG TGT TTT GAA GGT TGT AGG* *5' GCA GGA AAG ATA GAC AGA TAG ATA G*	(de Groot et al. 2008)
D8S1106	*5' TTG TTT ACC CCT GCA TCA CT* *5' TTC TCA GAA TTG CTC ATA GTG C*	(de Groot et al. 2008)
D3S1768	*5' GGT TGC TGC CAA AGA TTA GA* *5' CAC TGT GAT TTG CTG TTG GA*	(de Groot et al. 2008)
D6S2833	*5' GTA AAG TGG TGC GAT CAC AG* *5' AGT GGC TCA TGC CTT CAA TG*	(de Groot et al. 2008)
D6S2792	*5' ATC CAA TCA CCT CTG CTC AC* *5' AGA TTT CAT CCA GCC ACA GG*	(de Groot et al. 2008)
D6S2810	*5' CTA CCA TGA CCC CCT TCC CC* *5' CCA CAG TCT CTA TCA GTC CA*	(de Groot et al. 2008)
D6S2811	*5' TGG GCA ATG AGT CCT ATG AC* *5' TGC CAT TTG GCC CTA AAT GC*	(de Groot et al. 2008)
D6S2972	*5' GAA ATG TGA GAA TAA AGG AGA* *5' GAT AAA GGG GAA CTA CTA CA*	(de Groot et al. 2008)
D6S276	*5' TCA ATC AAA TCA TCC CCA GAA G* *5' GGG TGC AAC TTG TTC CTC CT*	(de Groot et al. 2008)
D6S1691	*5' AGG ACA GAA TTT TGC CTC* *5' GCT GCT CCT GTA TAA GTA ATA AAC*	(de Groot et al. 2008)
D11S295	*5' GCT CCT CCA GTA ATT CTG TC* *TTA GAC CAT TAT GGG GGC AA*	***
D13S765	*5' TGT AAC TTA CTT CAA ATG GCT CA* *5' TTG AAA CTT ACA GAC AGC TTG C*	(Chambers et al. 2004)
D18S72	*5' GCT AGA TGA CCC AGT TCC C* *5' CTG CAG AAA GGT TAC ATA TTC CA*	***
MFGT21	*5' AACTTCAGTAAGATAAGGACC* *5' CCTGAGGTCTGGACTTTAT*	***

D5S820	*5' ATT GAC TGG CAA CTC TTC TC* *5' GTT CTT CAG GGA AAC AGA ACC*	***
DXS2506	*5' GGA GAA ATG GGG AGT AAC TG* *5' ACA CAT GGC TGG CTA GCT T*	(Nagaraja et al. 1997)
MFGT18	*5' GCC CCA AAT GCC AGC AGA AC* *5' TCT GAG AGC TGT GAT GGG AC*	(Oka & Takenaka 2001)

Loci with asterisks *** do not have a reference. These loci and the respective primers were provided by the Biomedical Primate Research Centre, Netherlands.

3.2.4 Microsatellite PCR optimisation

Reactions commenced with the following components per 20*µ*l reaction: 1x CoralLoad PCR buffer (Contains Tris·Cl, KCl, (NH4)2SO4,15 mM MgCl2, gel loading reagent, orange dye, red dye, Qiagen), 1x Q Solution (a reagent part of the HotStart Taq Plus kit which modifies the melting behaviour of DNA which is stated as useful for difficult template DNA, Qiagen), 0.2*µ*M dNTP mix, 0.2 *µ*M primer (*F*), 0.2*µ*M primer (*R*) (primers as per Table 7), 1.75mM $MgCl_2$, 2.5 units HotStarTaq® *plus* DNA polymerase (Qiagen), 8 to 198 ng/*µ*l template DNA (quantified by Nanodrop ND 1000 spectrophotometer, Thermo Scientific). Both a negative control and a positive control with 100ng/*µ*l human DNA were added to each PCR cycle. Reaction components were altered if PCR amplification was not successful. These included increasing $MgCl_2$ concentration and increasing template DNA concentration if bands were present but weak. The addition of Bovine Serum Abumin (Sigma Aldrich) was utilised at 10% of the total reaction volume as a final resort for loci that continually failed to amplify after PCR. There are conflicting reports as to the efficacy of adding BSA, however it is purported to act as a binding agent to PCR inhibitors (Chaves et

al. 2006; Ernest et al. 2000; Morin et al. 2001). The addition of BSA did not improve results and loci that did not amplify sufficiently were discarded.

PCR reactions were run using a S1000 Thermal Cycler (Bio-Rad) with the following conditions: 95ºC for 5 minutes to active HotStarTaq plus, then 45 cycles of denaturing phase at 95ºC for 30 seconds, annealing at (various – see Table 8) for 30 seconds, extension at 72ºC for 30 seconds with a final extension at 72ºC for 30 minutes.

Annealing temperatures were initially calculated using the Ta Calculator function of the Thermal Cycler (Bio-Rad). Then to optimise annealing temperatures, each loci was then tested using the resultant T*a* calculator temperature as the median with a 10ºC differential either side of this figure.

All resultant products from the optimisation were visualised on a 1.2% agarose gel using the Gel Doc™ EZ System (Bio-Rad) and its affiliated software Image Lab version 3.0 (Bio-Rad). Successfully amplified loci were then approximated in size using the 100bp ladder (New England Biolabs) that had been loaded with PCR products.

3.2.4.1 Final PCR conditions

After the optimisation period a number of loci were discarded owing to repeated failures after attempted amplification. The following conditions applied to the

remaining loci: 20μl reaction: 1x CoralLoad PCR buffer (Qiagen), 1x Q Solution (Qiagen), 0.2μM dNTP mix, fluoresced modified primer (*F*) 0.02 μM (as per Table 8), primer (F) 0.18 μM, primer (*R*) 0.2μM (primer sequences as per Table 7), $MgCl_2$ 1.75mM, HotStarTaq *plus* DNA polymerase 2.5 units, template DNA between 8 to 198 ng/μl.

Each locus had the addition of a fluoresced forward primer labelled at the 5' of the primer sequence. This was added so that when subsequent genotyping was carried out within the sequencer, amplified alleles would be detected. Cycling conditions were carried out as listed above, with the addition of the following annealing temperatures as shown in Table 8.

Table 8 - List of selected loci after optimisation with annealing temperatures and fluorescent labels

Locus	Annealing Temp.	Primer Label
MFGT21	58ºC	6-FAM
MFGT18	58ºC	6-FAM
DXS2506	58ºC	6-FAM
D6S2972	49.2ºC	6-FAM
D6S2859	50.9ºC	HEX
D6S2854	62.3ºC	NED
D6S2833	50ºC	NED
D6S2792	52.3ºC	HEX
D6S1691	45.9ºC	HEX
D5S820	58ºC	HEX
D5S1457	58ºC	HEX
D3S1768	58ºC	6-FAM
D19S714	59ºC	HEX

D18S72	58ºC	NED
D17S1290	59ºC	6-FAM
D14S306	59ºC	6-FAM
D13S765	58ºC	NED
D12S391	55ºC	6-FAM
D11S295	58ºC	HEX

For quality assurance each loci for each subject animal was duplicated within the plate. In order to size products after genotyping in the sequencer a ROX500 standard was added to each well within the plate.

3.2.5 Statistical analysis of microsatellites

3.2.5.1 Viewing allele sizes – electropherogram analysis

Electropherogram profiles of above listed loci were viewed and sized using GeneMapper® software version 5.0 (Applied Biosystems). Alleles were sized in accordance with the 'Microsatellite Analysis Method' and its affiliated default parameters in the programme. Peak Quality was set at 140 minimum peak height for homozygous results and 85 for heterozygous, maximum peak width 1.5, peak detection for all fluorescent labels was set at 50. Alleles were first located by zooming in to the approximated size bands previously visualised on the agarose gels. Allele peaks were not considered to be true alleles if the amplification value was less than 50. In addition, if a peak was deemed to be too wide (in excess of 1.5bp as per the default values within the programme), which results in a smooth apex and gradual inclines on both the positive and negative slopes then it was also disregarded.

3.2.5.2 Identification of possible genotyping errors

In order to establish whether any possible genotyping errors were present within the population data, results were investigated using Micro-checker version 2.2.3 (van Oosterhout et al. 2004). This programme allows for the detection of null alleles and the presence of stutter peaks. Stutter peaks are detected by analysis of the largest and smallest of allele sizes within a given loci. They will be flagged if there is a deficiency of heterozygotes within the population who exhibit allele sizes that differ by only one repeat and also if there is an excess of homozygotes who exhibit larger allele sizes, as the probability of stutter within larger alleles is greater than evinced in smaller ones (van Oosterhout et al. 2004). Null alleles are detected based on observed frequencies of homozygotes within a study population (Brookfield 1996; Chakraborty et al. 1992). If an apparent excess of homozygotes is observed in comparison to observed heterozygotes then there is a possibility that one allele has not been amplified within an individual.

3.2.5.3 Measures of genetic diversity

The first step to analyse genetic diversity within the two populations was to test whether all loci were in Hardy-Weinberg Equilibrium. The test assumes that the group under analysis is undergoing random mating and that factors of selection, mutation and migration are absent (Crow 1988). Therefore, if the above parameters are met, genotype frequencies will not change from one generation to the next. Thus, loci that do not fall within the Hardy-Weinberg equilibrium may indicate events within the population such as inbreeding or population differentiation. Tests of all loci for Hardy-Weinberg equilibrium were carried out using Arlequin, version 3.5.1.3 (Excoffier & Lischer 2010). The underlying methodologies within the software are based on Fisher's exact test but the two-by-two table is replaced by an extended

version of a triangular table of random size (Guo & Thompson 1992). The utilisation of an exact test was preferred here as the sample size is small (Balding 2006; Bartlett 1937; Guo & Thompson 1992). Results were then tested using the Markov-chain random walk algorithm (Guo & Thompson 1992) by the recommended 100,000 forecasted chain lengths and 100,000 dememorization steps. Loci that were deemed to be out of Hardy-Weinberg proportions were discarded. In addition to this, any locus found to be monomorphic was also discarded. This was performed so as not to skew data in further downstream analysis as a great number of software applications function on the basis that loci are mutating on a neutral basis, unless specifically stipulated and monomorphic loci do not provide a sufficient level of information to ascertain parentage or variability (Excoffier & Heckel 2006).

Arlequin (version as above) was also used to calculate observed heterozygosity (H_o) (calculated as the number of heterozygotes within the population divided by the total number of individuals) and expected heterozygosity (H_e). Expected heterozygosity is calculated within the programme according to the following formula:

$$He = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right)$$

Where n is the number of gene copies in the sample, k is the number of haplotypes and p_i is the sample frequency of the *i*-th haplotype. Both aforementioned measures of heterozygosity are calculated for each locus.

Heterozygosity was also measured at the individual level. Amplification was not successful for every locus within every individual therefore, a standardised version

of heterozygosity was calculated (H_s). Standardised heterozygosity was calculated as the individual heterozygosity divided by the overall mean heterozygosity of all typed loci (Coltman et al. 1999). Thus calculations were according to the following formula:

$$Hs = \frac{H}{\overline{x}H_{li}}$$

Where l is a microsatellite loci genotyped for individual i.

The Individual heterozygosity (H_s) measure of genetic diversity is purported to be a useful measure for assessing inbreeding on a more recent timescale (Coulson et al. 1998). A further test of genetic variability and inbreeding, devised by Coulson and colleagues (1998) is mean d^2. This is also a measure that applies to the individual within a population, but gives deeper insight of inbreeding events deeper within the pedigree. The test functions by calculating the squared difference of allele lengths within a locus which is then averaged over the entire loci within the data set. The calculation for mean d^2 is carried out as follows:

$$Mean\ d^2 = \sum_{i=1}^{n} \frac{(i_{a-}\ i_b)^2}{n}$$

However, as with measures of Individual Heterozygosity, this calculation has also been standardised which limits loci that are more polymorphic than others making a greater contribution to results procured from the above illustrated mean d^2 equation. The amended calculation is referred to as scaled mean d^2 and incorporates the variance at each locus within the data set prior to giving the overall mean d^2 (Coulson et al. 1999). The test is carried out with the following calculation:

$$Mean\ d^2_{SCALED} = \sum_{I=1}^{N} \left[\frac{\left(\frac{(i_a - i_b)^2}{\sigma_i^2} \right)}{n} \right]$$

A further test carried out to assess inbreeding levels is based on the estimation of parental similarity and is known as Internal Relatedness (IR) (Amos et al. 2001). This measure of inbreeding is based on alleles shared between individuals. It assesses the frequency of alleles within a population and unique or rarer alleles give greater weighting within the result that with the more common alleles. The calculation is based on a concept by Queller and Goodnight (1989). The modification made by Amos and colleagues (2001) allows for comparisons between alleles within a locus as opposed to between pairs of individuals. The calculation is carried out as follows:

$$IR = \frac{(2H - \sum fi)}{(2N - \sum fi)}$$

Where *H* is the number of homozygous loci, *N* represents the number of loci and f_i is the frequency of the ith allele within the genotype. A highly positive value may be indicative of inbreeding and conversely, a highly negative value is indicative of highly outbred individuals.

To test correlation between the different indices of genetic diversity a Pearson coefficient (*r*) was calculated in Excel version 14.5.1 according to the following formula where x and y are the samples means of each index array:

$$r = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sqrt{\sum(x - \overline{x})^2 \sum(y - \overline{y})^2}}$$

3.2.5.4 Measures of genetic variation between the two groups

Further genetic variation measures were taken at the locus level and compared between the two populations. The first was allelic richness and was calculated using the software programme FSTAT version 2.9.3.2 (Goudet 2001). It was performed using a rarefaction method which calculates the number of alleles at a given locus independently of the sample size. This was important as the captive-born group is larger than the wild-born. Allelic richness is calculated using the following formula:

$$Allelic\ Richness = \sum_{i-1}^{n}\left[1 - \frac{\binom{2N-N_i}{2n}}{\binom{2N}{2n}}\right]$$

Where N_i is the number of alleles of size *i* among the 2N genes

A further inbreeding statistic was also calculated which is referred to as Wright's F_{IS} (Wright 1951) using FSTAT version 2.9.3.2 (Goudet 2001). This statistic is calculated by comparing both the observed and expected levels of heterozygosity within a population (and by locus) by the following formula:

$$F_{IS} = \frac{H_s - H_i}{H_s}$$

Where H_i is the observed heterozygosity of an individual and H_s is the expected heterozygosity of an individual.

The final two measures of genetic diversity were carried out to ascertain if there was any differentiation between the two populations of captive-born and wild-born. The first was a measure of F_{ST} based on the infinite allele model (Weir & Cockerham 1984) and calculated using Arlequin version 3.5.1.3 (Excoffier & Lischer 2010). Here population differentiation is calculated as a count of differing alleles between two haplotypes. The formula that the software utilises is as follows:

$$\widehat{d}\,xy = \sum_{i-1}^{L} \delta_{xy}\ (i)$$

Where $\delta_{xy}\ (i)$ is the Kronecker function, thus equal to one if the alleles of the *i-th* locus are identical for both haplotypes and equal to zero if not.

The other measure of population differentiation also calculated using Arlequin version 3.5.1.3 (Excoffier & Lischer 2010) was an analogue of Slatkin's R_{ST} (1995) calculated using the stepwise mutation model. This methodology counts the sum of the squared number of repeat differences between two haplotypes and calculated as follows:

$$\widehat{d}\,xy = \sum_{i-1}^{L} (a_{xi} - a_{yi})^2$$

Where a_{xi} is the number of repeats of the microsatellites for the *i-th* locus.

3.2.5.5 The pedigree of Hylobates moloch

Parentage assignment for the *H. moloch* population was assayed using Cervus version 3.0.3 (Kalinowski, Taper & Marshall 2007). The programme uses a likelihood-method approach to assign offspring to a prospective parent. The likelihood method used within this version of the Cervus software is a modification of

the approaches of both Marshall *et al.* (1998) and Morrissey and Wilson (2005). Use of a likelihood based methodology as opposed to the exclusion method was preferred as the former approach allows for genotyping errors, null alleles and the possibility of mutations within the dataset (Jones & Arden 2003). Within Cervus, the most likely parent or parent pair is calculated in accordance with calculations of a logarithm of the likelihood ratio (LOD). Genotypes of offspring and potential parents are analysed and the likelihood of a potential parent of a particular offspring is divided by the likehood of the individuals under analysis being completely unrelated. Offspring are then assigned to a potential parent with the LOD score, which if it is a high score indicates that the likelihood is greater that the match is correct. The number of loci used within the analysis totalled 15, but the minimum typed loci was stated as 10 (with proportion of loci typed stated as 0.95). This allows for a degree of error that may be present within the genotypic dataset. As it was known that not all sires and dams were present within the sample population, a figure to indicate this of 0.9 was used for the proportion of parents sampled.

3.3 Results of Microsatellite genetic diversity and pedigree analysis

After analysis in Micro-checker (van Oosterhout et al. 2004) four loci were indicative of the presence of null alleles: D6S2859, D3S1768, D17S1290 and D12S391. In addition to the possibility of null alleles three of the aforementioned four loci deviated from the expected Hardy-Weinberg proportions. Therefore, they were also removed from further analyses and do not appear in results reported below. The locus D12S391 was found to be in Hardy-Weinberg equilibrium, and although the locus was flagged as potentially containing null alleles, the result was not significant thus was retained for further analysis.

3.3.1 Genetic diversity measures at the group and locus level

The mean count of alleles within each group is very similar, as is the value of mean allelic richness (Table 9). Both observed and expected heterozygosity values are greater within the captive-born group then in the wild-born. The mean F_{IS} value for the wild individuals is a positive number. This can be illustrative of a deficiency of heterozygotes within the group, and thus a possibility of inbreeding. The captive-born group exhibit a negative F_{IS} value which suggests the opposite, and that there maybe an excess of heterozygotes within the group.

The mean d^2 values show a high value for the wild-born group and a value of less than half of that within the captive-born group. However, both groups, and in particular the wild-born, procured extremely high standard deviation figures which suggest results are highly variable. The Scaled Mean d^2, Individual Heterozygosity and Standardized Heterozygosity values are all greater in captive-born individuals. The Internal Relatedness value for the Wild group is a positive number whilst it has yielded a negative number for the Captive group (Table 10).

Table 9 - Number of alleles (A), allelic richness (AR), allele size ranges, observed heterozygosity (H_o), expected heterozygosity (H_e) and F_{IS} for the final selected panel of 15 loci in wild and captive groups

Locus		Wild-born	Captive-born	Total per locus
MFGT21	Sample number	3	7	10
	No. Alleles	3	3	4
	Allelic Richness	3.00	2.40	
	Size Range	130-139	130-135	
	H_o	0.333	0.857	
	H_e	0.733	0.582	
	F_{IS}	0.600	-0.532	
MFGT18	Sample number	7	13	20
	No. Alleles	3	5	5
	Allelic Richness	2.359	2.816	
	Size Range	86-98	82-104	
	H_o	0.429	0.538	
	H_e	0.538	0.646	
	F_{IS}	0.217	0.172	
DXS2506	Sample number	5	12	17
	No. Alleles	3	4	4
	Allelic Richness	2.857	3.112	
	Size Range	281-289	277-289	
	H_o	0.600	0.583	
	H_e	0.711	0.71377	
	F_{IS}	0.172	0.189	
D6S2972	Sample number	5	12	17
	No. Alleles	5	6	7
	Allelic Richness	3.762	3.242	
	Size Range	123-137	123-133	
	H_o	0.800	0.833	
	H_e	0.800	0.714	
	F_{IS}	0.000	-0.176	

Locus		Wild-born	Captive-born	Total per locus
	Sample number	8	13	21
	No. Alleles	6	6	7
	Allelic Richness	3.374	3.610	
D6S2833	Size Range	233-247	230-247	
	H_o	0.750	0.923	
	H_e	0.683	0.778	
	F_{IS}	-0.105	-0.195	
	Sample number	4	10	14
	No. Alleles	4	4	4
	Allelic Richness	3.679	3.26	
D6S2792	Size Range	129-144	129-144	
	H_o	0.750	0.600	
	H_e	0.821	0.742	
	F_{IS}	0.100	0.200	
	Sample number	8	12	20
	No. Alleles	3	3	3
	Allelic Richness	2.509	2.308	
D6S1691	Size Range	184-190	184-190	
	H_o	0.375	0.333	
	H_e	0.567	0.507	
	F_{IS}	0.354	0.353	
	Sample number	7	13	20
	No. Alleles	3	4	4
	Allelic Richness	2.774	2.974	
	Size Range	171-187	171-187	
	H_o	0.714	0.923	
	H_e	0.670	0.705	
D5S820	F_{IS}	-0.071	-0.327	

Locus		Wild-born	Captive-born	Total per locus
	Sample number	6	13	19
	No. Alleles	4	8	9
	Allelic Richness	3.561	3.843	
D5S1457	Size Range	132-152	126-152	
	H_o	0.500	0.846	
	H_e	0.803	0.803	
	F_{IS}	0.400	-0.056	
	Sample number	6	12	18
	No. Alleles	7	4	7
	Allelic Richness	4.515	2.998	
D19S714	Size Range	226-260	226-252	
	H_o	1.000	1.000	
	H_e	0.879	0.707	
	F_{IS}	-0.154	-0.443	
	Sample number	5	13	18
	No. Alleles	3	3	3
	Allelic Richness	2.200	2.334	
D18S72	Size Range	308-314	308-314	
	H_o	0.400	0.462	
	H_e	0.378	0.542	
	F_{IS}	-0.067	0.153	
	Sample number	8	13	21
	No. Alleles	5	6	6
	Allelic Richness	3.863	3.626	
	Size Range	203-234	203-324	
	H_o	0.750	0.769	
	H_e	0.825	0.788	
D14S306	F_{IS}	0.097	0.024	

Locus		Wild-born	Captive-born	Total per locus
	Sample number	5	13	18
	No. Alleles	4	4	5
	Allelic Richness	3.067	3.336	
D13S765	Size Range	214-228	210-228	
	H_o	0.800	0.846	
	H_e	0.644	0.760	
	F_{IS}	-0.280	-0.119	
	Sample number	7	12	19
	No. Alleles	11	8	14
	Allelic Richness	5.505	4.291	
D12S391	Size Range	179-272	217-272	
	H_o	0.857	0.667	
	H_e	0.967	0.859	
	F_{IS}	0.122	0.231	
	Sample number	7	13	20
	No. Alleles	8	7	10
	Allelic Richness	4.611	4.148	
D11S295	Size Range	281-307	281-305	
	H_o	0.857	1.000	
	H_e	0.890	0.849	
	F_{IS}	0.040	-0.186	
	Mean A (s.d.)	4.80 ± 2.336	5.0 ± 1.732	
	Mean AR (s.d.)	3.442 ± 0.615	3.220 ± 0.920	
Totals	Mean H_o (s.d)	0.661 ± 0.207	0.745 ± 0.204	
	Mean H_e (s.d)	0.727 ± 0.154	0.713 ± 0.105	
	Mean F_{IS} (s.d)	0.095 ± 0.231	-0.047 ± 0.265	

Table 10 – Inbreeding measures for the captive and wild groups. Standard deviation figures are shown in brackets.

	Wild-born	Captive-born
N	8	13
Mean d^2	227.981 (± 202.792)	100.513 (± 49.303)
Scaled Mean d^2	0.019 (± 0.014)	0.026 (± 0.014)
Individual Heterozygosity	0.688 (± 0.037)	0.747 (± 0.017)
Standardised Heterozygosity	0.940 (± 0.370)	0.995 (± 0.180)
Internal Relatedness	0.044 (± 0.346)	-0.081 (± 0.185)

The greatest correlation between above mentioned genetic diversity indices are between the standardised heterozygosity and internal relatedness values which are illustrative of a near perfect negative correlation. The least correlation is evinced between internal relatedness and the scaled mean d^2 results.

Table 11 - Pearson coefficient (correlation) between genetic diversity statistics. ** *p<0.01*

	Scaled Mean d^2	**Standardised Heterozygosity**
Internal Relatedness	-0.559**	-0.962**
Standardised Heterozygosity	0.588**	

Table 12 - Measures of population differentiation, R_{ST} and F_{ST}.

	Captive-born (R_{ST})	**Captive-born (F_{ST})**
Wild-born	0.006 (*NS*)	-0.022 (*NS*)

Neither value of F_{ST} and R_{ST} signify any statistically significant genetic deviation between the captive and wild-born groups (Table12).

The values of standardised heterozygosity for each individual and mean kinship values are shown in Table 13. Standardised heterozygosity values are generally high and more in alignment with mean kinship values than in the results from mtDNA.

Table 13 - Mean kinship (MK) and standardised heterozygosity values per individual. Sorted by MK values.

Primate Ref	Mean Kinship	Standardised Heterozygosity
H.mol8	**-**	**1.210**
H.mol17	-	1.073
H.mol21	-	1.028
H.mol4	**0.0000**	**1.072**
H.mol3	**0.0000**	**0.698**
H.mol6	**0.0000**	**0.154**
H.mol1	**0.0083**	**1.195**
H.mol5	**0.0083**	**0.891**
H.mol7	**0.0208**	**1.210**
H.mol15	0.0604	1.252
H.mol16	0.0688	0.678
H.mol2	**0.0875**	**1.165**
H.mol13	0.1083	0.775
H.mol20	0.1167	1.065

H.mol19	0.1167	0.800
H.mol14	0.1271	1.163
H.mol12	0.1271	0.926
H.mol11	0.1271	0.805
H.mol9	0.1354	1.132
H.mol10	0.1479	1.163
H.mol18	0.1521	1.073

3.3.2 Pedigree analysis

In Table 14, cells that are shaded in candidate mother and father columns represent missing samples from either sire or dam (or both) according to zoological family records. Assignments of candidate parents for these offspring have low LOD scores and therefore reiterate assigned parents are incorrect. For Hmol10 however the candidate mother has a positive LOD score which contravenes pedigree records. Candidate fathers highlighted in red do not match zoological pedigree records. Confidence levels (as shown by LOD scores) are only high for 5 other candidate parents.

Table 14 - Results of parentage analysis. Cells highlighted in grey represent sires or dams not present in this dataset. Text in red indicates a different parent from zoological records

Offspring ID	Loci typed	Candidate mother	LOD score	Candidate father	LOD score
H.mol9	14	H.mol18	-4.87E+00	H.mol2	-2.19E+01
H.mol10	15	H.mol18	6.41E-02	H.mol2	-1.43E+01
H.mol11	15	H.mol18	-1.32E+01	H.mol2	-9.35E+00
H.mol12	13	H.mol18	-4.61E+00	H.mol2	3.98E+00
H.mol13	14	H.mol15	9.56E-01	H.mol7	-8.60E+00
H.mol14	15	H.mol8	-1.90E+01	H.mol7	1.03E+00
H.mol15	15	H.mol1	-1.60E+01	H.mol5	-2.31E+01
H.mol16	14	H.mol8	-4.38E-01	H.mol17	-1.35E+01
H.mol17	13	H.mol8	-1.14E+00	H.mol5	-8.72E+00
H.mol18	15	H.mol8	-1.63E+01	H.mol2	-1.62E+01
H.mol19	11	H.mol18	3.94E+00	H.mol17	-3.07E+00
H.mol20	14	H.mol18	1.70E+00	H.mol17	1.45E-01
H.mol21	13	H.mol1	-1.62E+01	H.mol7	-1.88E+01

3.4 Discussion – Does genomic DNA diversity follow the same patterns observed in mitochondrial DNA within the Hylobates moloch groups?

It is a common issue that when researching endangered species, especially within a captive environment that the sample numbers are small. Thus, the question arises within studies such as here where 21 individuals were analysed, just how much biologically informative knowledge a small group can yield with regards to making management decisions based on genetically derived data for conservation purposes. Carrying out analyses to ascertain genetic variation within a small group may not identify all character states that may be present within a wider population and yield results of low variation as a consequence of this. Walsh (2000) stated that to capture genetic polymorphism within a population in a population of infinite size, a sample size of 59 individuals would be required to yield a confidence value that 95% of character states have been captured within the population. For a population of a finite size, this number can be reduced by 20% to derive results that mirror those from a finite population. Crandall *et al.* (2000), however, cite a smaller number to achieve the same objectives as Walsh. They base their guidelines to achieve 95% of genetic diversity within a population on sample sizes of between 20 to 50 individuals, which they believe would encapsulate ancestral alleles present within a small group.

Four of the microsatellite loci had observed heterozygosity levels in excess of 0.8 and five had levels between 0.6 to 0.8 in both the captive and wild-born groups (Table 9). Despite these results, confidence levels derived from parentage analysis were low even in progeny assigned to parents as per studbook records. According to Harrison *et al.* (2012) if sample data does not contain all candidate parents, or if

genotyping errors may be present, then success within differing parentage programs (each with a different inference method) increases linearly with the number of loci. However, as indicated by the shaded boxes (Table 14) low assignment may be indicative of a lack of samples from the parents which has been experienced in other studies (e.g. Gerber et al., 2000; Trong et al., 2013) and it has lead to low assignment rates of offspring.

For analyses of genetic diversity, the number of microsatellites used also varies in the literature. For example in a study of genetic differentiation of European harbour seals (*Phoca vitulina vitulina*) 7 microsatellites were analysed (Goodman 1998), 8 were employed in a study of the German wildcat (*Felis silvestris*) (Eckert et al. 2010), 9 for understanding orang-utan (*Pongo*) sub populations (Kanthaswamy, Kurushima & Smith 2006), and 10 for both Hainan Eld deer (*Cervus eldi hainanus*) (Zhang et al. 2008) and great reed warblers (*Acrocephalus arundinaceus*) (Hansson et al. 2000). However, Slate and Pemberton (2002) trialled a large panel consisting of 71 microsatellites, to test the robustness of inbreeding measures such as heterozygosity and mean d^2 in a population of wild red deer (*Cervus elaphus*), a widespread and abundant species. Their observations concluded that a very minimum of 10 microsatellite markers were required to draw biologically meaningful information from results, particularly if associations are being evaluated between the genetic diversity levels observed from the microsatellite genotypes and fitness traits within the species under assay.

Of the panel of 32 microsatellites within the *Hylobates moloch* assay only 19 loci were successfully amplified, representative of 56.25% of the total set. The success rate here was higher than a previous study of *Hylobates lar* where only 8 microsatellite loci were successfully amplified with the adequate credentials for analysis (i.e. loci were polymorphic and yielded reproducible results) out of a panel of 47 microsatellites previously located within humans (Chambers et al. 2004).

Three loci (D31768, D6S2859 and D17S1290) deviated from Hardy-Weinberg proportions which manifested as an excess of homozygotes at the aforementioned loci in comparison to expected proportions. There are a number of factors that may have effected this, the first is known as the Whalund effect (Sinnock 1975). This hypothesis analyses allele frequencies within a single population and when an excess of homozygotes is detected it may be as a consequence of a genetic division within the group. The occurrence of sub-populations within an apparent single population may arise for example, owing to adaptation to local environmental conditions. The Wahlund effect is a viable theory as after a number of generations within captivity, individuals may adapt to the more relaxed conditions in captivity (Montgomery et al. 2010). However, computations of population differentiation (F_{ST} and R_{ST} as shown in Table 12) for the *H. moloch* population did not support this theory as no genetic subdivision between wild and captive-born was detected.

Values of F_{ST} have been reported to be lower than R_{ST} values (Jarne & Lagoda 1996), which was shown here. The rationale behind this is explained as the underlying mutation models that the theories assume. Values derived from F_{ST} are

based on the Infinite Allele Model, which assumes that states observed within a population have been derived directly as a consequence of inheritance from previous descendants. However, R_{ST} performs under the stepwise mutation model, which is described by some (Meirmans & Hedrick 2011) to be better suited to the multi-allelic nature of microsatellites and also incorporates the fact that there is a bias towards microsatellites increasing in size rather than undergoing reductions (Chambers & MacAvoy 2000; Meirmans & Hedrick 2011). R_{ST} does take into account the sample size, which is a positive attribute here as the size is small. However, F_{ST} is described as being more applicable to studies whereby genetic differentiation between populations is not thought to be extreme.

A second theory that may yield deviations from Hardy-Weinberg proportions is that the population is undergoing non-random mating and thus is experiencing inbreeding depression. As with the Wahlund effect, this is also a viable theory considering the sample population under assay which is a captive, endangered and CITES Appendix I listed species. Owing to these factors there is a small group of available mating partners available and thus there is a possibility that inbreeding is taking place. The mtDNA analysis in the previous chapter illustrated that at the mtDNA level, genetic diversity was markedly lower in the captive-born group and there was evidence of genetic differentiation between the two groups as calculated by F_{ST}. When compared however, with values of $F_{IS,}$ derived from the microsatellite genomic markers the comparison between the two groups are not statistically significant. Thus genomic results were not concordant with mtDNA results for population differentiation. It is highly probable that the wild-born F_{IS} value is subject

to sample bias as the number of individuals within the group is small and furthermore, not all individuals within the group are typed at every locus analysed.

A further explanation, which was deduced to be the most likely explanation for deviations from Hardy-Weinberg proportions was the presence of null alleles. This was detected within all loci as previously mentioned using the Micro-Checker software. All three loci (D31768, D6S2859 and D17S1290) were found to have null alleles present within the data set and furthermore that this was significant ($p<0.01$). Therefore, although loci outside of Hardy-Weinberg proportions may be included in further analyses within studies if they are deemed to have occurred from a biologically significant reason such as due to Wahlund effects or inbreeding (Goodman 1998), they were discarded in this study, as null alleles appeared to be the most probable cause.

The 15 loci were representative of half of the 22 chromosomes archetypal of *H. moloch* (Müller, Hollatz & Wienberg 2003) and thus were deemed to give a broad overview of genetic variation for this study. A caveat should be noted here however, that loci are categorized as per the human microsatellite classification (e.g. D3S1768 is situated on chromosome 3). It is known that a microsatellite locus (and thus its flanking sequences) in one species may not appear on an analogous chromosome in another species. This has been illustrated in genetic mapping studies in primates such as within rhesus macaques (*Macaca mulatta*) (Rogers et al. 2006), baboons (*Papio hamadryas*) (Rogers et al. 2000) and even to other mammals such as the domestic cat (*Felis catus*) (Menotti-Raymond et al. 1999).

3.4.1 Genomic diversity of *Hylobates moloch*

The mean number of alleles over all loci were similar not only between the wild-born and captive-born groups but also in comparison to two other species within the *Hylobates* genus. Whittaker (2005) studied microsatellites within several populations of Kloss's gibbons (*H. klossii*) and found that the average number of alleles within all loci was 2.3. However, sample numbers utilised were very low, for example one of the populations was represented by only one individual, thus there is a possibility that this number would increase with a larger number of individuals. A study of white-handed gibbons (*H. lar*) within 8 microsatellites and a minimum of 43 individuals typed per locus, the overall mean allele count was found to be 7.0 (Chambers et al. 2004). Thus it would seem that the gibbon species with the widest geographic distribution, *H. lar,* maintains the most microsatellite alleles and this deduced from just 8 microsatellites, in comparison to the two endemic species of *H. Klossii* and *H. moloch*. An analysis of three groups of other species of primate in a captive environment, gorilla (*Gorilla gorilla gorilla*) and two sub-species of orang-utan (*Pongo pygmaeus abelii* and *Pongo pygmaeus pygmaeus*) consisting of 104, 21 and 23 individuals respectively were found to have 6.48, 3.28 and 3.2 alleles averaged over all loci (Zhang et al. 2001). Thus, the *H. moloch* groups, both Captive and Wild illustrated a greater number of alleles than the captive orang-utans, but not quite as many as found within the gorilla population.

The captive environment has been stated to be a cause of decline in microsatellite heterozygosity (Montgomery et al. 2010). Within some empirical studies, observations have been made that a marked difference between populations analysed at microsatellite markers within captive populations to their conspecific wild

populations. For example in commercially reared Artic charr (*Salvelinus alpinus*) overall variability at 8 microsatellites was lower than within their wild conspecifics (Ditlecadet et al. 2006). Similarly, a decline in allelic richness was observed in captively bred mallards (*Anas platyrhynchos*) versus wild born individuals with a value of 16.554 in the wild group and 10.088 in the captive (Čížková et al. 2012). The results of the two groups of *H. moloch* here are very similar to those observed in wild and captively bred lesser kestrels (*Falco naumanni*) (Alcaide et al. 2010). The number of alleles and allelic richness were similar in captively bred and wild born individuals. Whilst this may be a consequence of careful management within the captive breeding programme, there is also the possibility that in these two species' genetic variability has not been impacted as they have been bred within the captive environment within a limited number of generations. The possibility of decreasing genetic variability has been postulated to occur after a number (which is not a set number and will alter from species to species in accordance with life history traits) of generations within the captive environment (Frankham 2008). This has been observed within an empirical study of the Mallorcan midwife toad (*Alytes muletensis*) (Kraaijeveld-Smit et al. 2006). Here genetic variation comparisons between wild born and captively bred individuals of up to 8 generations were very similar. However, after 8 generations of captive breeding a noticeable drop in both heterozygosity and allelic richness were observed.

As with the mean count of alleles per group, the values for allelic richness were also similar. The use of allelic richness to ascertain genetic diversity is described to be of 'key relevance' by Rodrigáñez *et al.* (2008) in conservation programmes. The rationale is that a depth of allelic richness allows a greater ability of selection

response for future generations of species to come, rather than solely focusing on the frequency of a specific allele within a population. Thus, the desired outcome of allelic richness is for the value to be high, but of course is dependent on the number of alleles. Values are low when compared to Western lowland gorilla populations (*Gorilla gorilla gorilla*) which varied between 4.97 to 5.60 with the latter value referring to a small sample size of 13 individuals (Le Gouar et al. 2009). This study focused on populations that may have undergone a genetic bottleneck as a consequence of high mortality from outbreaks of Ebola (Le Gouar et al. 2009). Allelic richness was greater in one population after the outbreak, which may indicate that rare alleles were maintained in the surviving individuals. A study of a captive population of the same species, found that values were lower ranging from 3.30 to 5.19 (Nsubuga et al. 2010).

Mean H_e values of 0.727 and 0.713 for the Wild and Captive groups respectively once again are illustrative of little differentiation occurring between the two aforementioned groups. Values of H_o, however are greater in the Captive group with 0.745 versus 0.661 within the Wild individuals. This higher value may be as a consequence of a larger sample group within the captive individuals, however, and also as a consequence of incomplete typing within all loci for all individuals. The Captive group exceeded expectations of the level of heterozygotes, which is a positive attribute as it illustrates that retention of heterozygosity within a number of individuals has been maintained. Both values of observed and expected heterozygosity from the highly endangered Mediterranean monk seal (*Monachus monachus*) (Pastor et al. 2004) can serve to illustrate that the values obtained here for *H. moloch* are propitious. Samples sizes were larger than the *H. moloch* groups

analysed here but an equal number of microsatellites were typed (15) yet the mean H_e value was found to be 0.41 and although no mean value of H_o was given the values ranged from 0.07 to 0.63 with most loci featuring at the bottom end of this scale. These values in conjunction with other findings within the study lead to the conclusion that this species exhibits very low genetic diversity. Owing to negative anthropogenic actions, an endemic cervid species the Chinese water deer (*Hydropotes inermis inermis*) is becoming rare and under threat in its wild habitat, and thus a number of captive populations have been established to aid in it's conservation (Hu et al. 2007). The observed heterozygosity was found to be 0.531 in the captive populations. The lesser kestrel (*F. naumanni*) captive breeding programme produced similar results in their analysis of 8 microsatellites to the *H. moloch* groups both with regard to the heterozygosity levels observed and the wild group yielded a slightly lower value than the captive with 0.64 and 0.68 respectively (Alcaide et al. 2010). The white handed gibbon analysis (*H. lar*) found that H_e ranged from 0.49 to 0.857 and had an H_o of between 0.739 to 0.957 (Chambers et al. 2004). Thus observed heterozygosity is greater within the *lar* species, a second diversity measure that illustrates a greater genomic variability than in comparison to the *moloch* gibbons. Which may be as a consequence of the wider distribution that this gibbon species enjoys in comparison to the endemic island nature of the *moloch* gibbon.

3.4.2 Values of $_{SCALED}$ Mean $d^{2,}$, Standardized Heterozygosity (SH) and Internal Relatedness (IR), to analyse inbreeding levels

The scaled version of *Mean* d^2 was calculated so as to allow for the inclusion of the variance at each locus, which provides a more balanced estimate of all loci under

assay rather than highly polymorphic loci contributing more than the less polymorphic loci. This reversed the results from the Mean d^2 calculations with the captive-born group yielding a value of 0.026 versus 0.019 in the wild-born group. The value within the Wild group is most likely lower as the variance within each loci for this group were all much greater than in the Captive dataset which intimates that there is a higher degree of variability within the Wild group. The values in the *H. moloch* groups are similar to that observed within a study of zebra finches (Forstmeier et al. 2012) with a known pedigree. The standardised Mean d^2 value for all birds analysed was 0.0189, yet when 47 individuals known to be highly inbred were excluded this value lowered to 0.0028. Thus, if this metric is designed to illustrate more outbred individuals by a greater value of Mean d^2 then this result is unclear. This measure of inbreeding is not usually used as a stand alone statistic as it is utilised here, but often analysed in conjunction with a fitness trait with the basis that a greater Mean d^2 correlates with a greater level of fitness. Studies that have found positive links between Mean d^2 and survival have been made in red deer calves (*Cervus elaphus*) whereby females illustrated an elevated first winter survival rate in individuals with a higher Mean d^2 (Coulson et al. 1999). Coltman *et al.* (1998) found a correlation between harbour seal pups (*Phoca vitulina*) and survival to weaning and high Mean d^2 levels, which as above was assumed to be as a consequence of the surviving individuals being more outbred than those who did not survive. Further fitness links with this metric have also been drawn within harp seals (*Phoca groenlandica*) (Kretzmann et al. 2006). In the harp seal study variability of loci was high (9 to 22 alleles) but also homozygosity in individuals was very low (Kretzmann et al. 2006). But, as much as there may be studies drawing statistically significant fitness links with the Mean d^2 measure, which is deduced to be as a

consequence of inbreeding, there are also many studies that find no correlation at all. This has been reported in the grey seal (*Halichoerus* grypus) (Bean et al. 2004), Lippizan horses (*Equus caballus*) (Curik et al. 2003), zebra finches (*Taeniopygia guttata*) (Forstmeier et al. 2012), dice snake (*Natrix tessellata*) (Gautschi et al. 2002), and the Chinese native chicken (*Gallus gallus domesticus*) (Liu et al. 2006). It is more suited to deeper events of inbreeding and also to detecting speciation events (Neff 2004; Slate & Pemberton 2002). The metric also functions on the stepwise mutation model which does allow for a gain or a loss of a repeat within the microsatellite length, but this theory only allows for a single and not multiple repeat alteration. Thus, this would confound results utilising the Mean d^2 approach to ascertain inbreeding as there is a possibility that an allele within a population has culminated at a particular size larger or smaller by more than repeat than its ancestral state within one generation owing to either recombination events or DNA slippage (or possibly a combination of both).

The observed values of heterozygosity increased within both groups once standardised (SH). These values are illustrative of very little inbreeding within either group and would be meeting zoological objectives of maintaining high levels of heterozygosity. This is then corroborated with values of mean kinship. The results are not in complete harmony, with mean kinship increasing with standardised heterozygosity decreasing, however SH on the whole is high. There is one anomaly that appears in H.mol6 where the lowest SH value is achieved but has the greatest MK value. As with Mean d^2, standardised multi locus heterozygosity and indeed heterozygosity values have been studied in many cases in conjunction with fitness traits and are heralded by some to be a more informative and draw more significant

links than with Mean d^2. A strong correlation between heterozygosity levels and semen quality has been observed in the Iberian lynx (*Lynx pardinus*) (Ruiz-López et al. 2012). Individuals exhibiting low levels of heterozygosity (0.333 or 0.730 for standardised mean heterozygosity) also yielded a low semen quality which was described as "extreme" by the researchers, and a potential risk for the species overall fitness and reproductive parameters. A study into heterozygosity and growth levels in soay sheep (*Ovis aries*) found that no correlation existed, however a link between heterozygosity and the reproductive success of males was found (Di Fonzo et al. 2011). A similar observation was made in two Scandinavian brown bear (*Ursus arctos*) populations (Zedrosser et al. 2007). Measures of inbreeding and variability were measured at 18 microsatellite loci by application of heterozygosity and internal relatedness metrics. Both values were found to be correlated with reproductive success within male brown bears.

The values for internal relatedness for the *H. moloch* groups differed in that the captive-born individuals exhibited a negative value of IR, illustrative of an outbred group however the wild-born group yielded a positive value, which would allude to this group being more inbred. As the IR metric utilises the frequency of homozygote alleles it would appear to be a useful measure to ascertain levels of variability within the group, with high frequencies of homozygous alleles suggesting high levels of relatedness between individuals within the group. Despite the positive IR value within the wild-born individuals, other measures have illustrated that there is genetic variation within this group. An explanation for the negative IR results is that the smaller sample group in comparison to the captive-born has influenced the outcome of results. Furthermore, this value is very low which can be illustrated by reporting a

value derived from the common shrew (*Sorex araneus*) which is a species thought to undergo regular breeding with related individuals (Välimäki , Hinten & Hanski 2007). Within the populations assayed the greatest value of IR reached 0.261 (*n=21*), and the correlation was made that those with the greatest IR values lost more trials set within the study. Thus it was deduced that a competitive ability was held by those who were less inbred, an observation made within other species also, such as within the brown bears (*Ursus arctos*) (Zedrosser et al. 2007).

3.5 Conclusions

From a panel of 32 microsatellite loci tested within the *Hylobates moloch* groups, a resultant total of 15 were successfully amplified and were found to be in Hardy-Weinberg equilibrium and thus allele frequencies observed were presumed not to have been disturbed by factors such as genetic drift and selection. Despite yielding negative results in IR and F_{IS} in the wild-born group, the result is likely to have been susceptible to the low number of *moloch* gibbons (*n=8*). Values of heterozygosity once standardised, which accounted for incomplete typing data for some individuals, were almost identical between wild and captive-born groups. The same was also found in the overall mean number of alleles for each group. The allelic richness, a measure of uniqueness with regards to private alleles in each group, is also extremely similar. Hence, although splitting the groups can provide very useful data with regards to the genetic processes that generations in captivity may occur, the small group size for the wild-born individuals has confounded some results here. Therefore, the statistical measures that take into account sample size, or those that standardise data are preferable to make assumptions as to results procured.

Intra-species levels of variability in the two *H. moloch* groups are very similar confirmed by both R_{ST} and F_{ST} which is indicative that no deviation has taken place from the wild-born genetics to captive-born. The inter-species comparisons show a moderate and comparative level with other primate species, including those within the *Hylobates* genus. With regards to the latter primate comparisons, it was interesting to note that both endemic and island dwelling species of *moloch* and *klossii* gibbons showed a slightly lower level of variability than the more geographically widespread *lar* species. This genetic differentiation between island and mainland species is widely reported in a number of different species (Frankham 1997). Despite this, however, the standardised values of heterozygosity in both *H. moloch* groups meet genetic management objectives often set within zoological institutions of 90% of heterozygosity retention (Ballou et al. 2010).

The results of the pedigree analysis highlighted the difficulties in obtaining strong confidence levels for parental assignment when the sample dataset is small and incomplete. From a total of 26 individuals, only 10 parental samples were available. Parental pairings matched zoological records in all but two individuals. The candidate fathers for H.mol14 and H.mol15 suggested a different father than records state.

4 Chapter four - Analysis of an adaptive marker Class II of the Major Histocompatibility Complex

The MHC is responsible for coding proteins involved in immune response and performs vital functionality with regards to infection, autoimmunity and inflammation (Horton et al. 2008). It is a multigene family present within all Gnathostome taxa (Babik 2010), located on the short arm of chromosome 6 in humans (Horton et al. 2008) and is described as the most polymorphic genomic region within vertebrates (Babik & Radwan 2007). Within the MHC, the classical subdivision of genes encompasses three classes, two of which are the main immunological subgroups known as class I and class II (Piertney & Oliver 2006). Whilst class III is involved in immune functionality, it does not code for antigen presenting molecules (Ujvari & Belov 2011). MHC class I genes are expressed on almost all nucleated somatic cells, the molecules of which are comprised of a single polypeptide chain produced from a singular gene (Miller, Belov & Daugherty 2006). Class I molecules function intracellularly, binding invading pathogens, chiefly viral, within the cell's cytoplasm (Bernatchez & Landry 2003). Class II molecules have a narrower distribution among somatic cells with expression occurring primarily on antigen-presenting cells such as B cells and macrophages, and their architectural form is a heterodimer with the differing chains coded for by separate genes (Miller, Belov & Daugherty 2006). Class II molecules monitor pathogenic activity external to the cell. Their role in presenting antigens to helper T cells at the cell surface that have been procured from the binding of pathogenic peptides such as bacteria is vital to the adaptive immune system process (de Groot et al. 2009). Further to classification of MHC genes into classes in accordance with their structure and immune function, there are further groupings of isotypes which in humans are designated HLA-DR, DQ and DP

(HLA is an acronym for Human Leukocyte Antigen) (Doxiadis et al. 2008). Each grouping is then coded by one or more *A* or *B* genes (de Groot et al. 2009). This overall structure has also been described in chimpanzees (*Pan troglodytes*) and macaques (*Macaca mulatta* and *Macaca fascicularis*) (de Groot et al., 2009; Doxiadis et al., 2006; Doxiadis et al., 2007). The locus selected for analysis for this study of captive *H. moloch* was the class II *DRB* region.

4.1.1 Choice of region for analysis – The MHC class II *DRB* Exon 2

Despite both class I and class II genes playing integral roles within an organism's immune response, both innate and acquired there are delineations of their respective actions as previously mentioned. Whilst class I molecules function within the remit of innate immune response, they also serve as a bridge to the acquired immune system, for example via their interaction with killer-cell immunoglobulin-like receptors (Parham 2005). Class II molecules however, function within the acquired immune system and their effectiveness with regards to recognizing and binding extracellular pathogens is greatly dependent on the variability of antigen binding sites (ABS), also referred to as peptide binding regions (PBR). Within class II genes, the *DR* region is described as the most complex in humans (de Groot et al. 2009). This phenomenon, however, is largely communicated via genes of the *DRB* loci than from *DRA* genes as the latter have proved to be conserved within humans (Doxiadis et al. 2008). Thus it was preferable to assay the *DRB* locus and the second exon was further selected as it is known to contain the PBR.

4.1.2 Hypotheses of driving forces behind MHC variation

The high variability of the MHC is as a consequence of strong selective pressures (Sutton et al. 2011). The reason why this genomic region merits such a description is rooted in the ways that variability manifests in different species. Variation is not solely attributed to sequence variation at PBRs for example, it stems from differential counts of gene copies, number of alleles and variation within these loci at the sequence level themselves (Eimes et al. 2011; de Groot et al. 2009; Doxiadis et al. 2007). Furthermore, the configuration of how differing alleles are positioned within differing haplotypes culminates in an additional echelon of genetic diversity at this region (Doxiadis et al. 2007).

Just as the manifestations of variability are diverse at the MHC both at the intra- and inter-species levels, the hypotheses that seek to elucidate them are numerous. Two prominent theories of how MHC polymorphism is maintained are via parasite induced selection and sexual selection, both of which may very well function in conjunction with one another (Drury 2010).

The theorem of pathogen induced selection primarily received little focus as links between MHC genes and autoimmune disease within humans had already been established (Piertney & Oliver 2006). However, as experimental studies in both humans and other species on regions within the MHC increased, the interest in the relationship between susceptibility to disease or parasitic infection and specific MHC genes grew. Subsequently, hypotheses were then created to explain new findings. The first theory was that of ‘over-dominance’ which inferred that fitness levels are

greater in individuals heterozygous for MHC alleles, as a consequence of their potential to present a greater array of antigens (Hughes & Nei 1988). A further model is that of 'negative frequency-dependent selection' which denotes that an individual harbouring a rare allele or perhaps a new one within a population has a fitness advantage over individuals possessing more common isotypes. If this novel allele is selected and proliferates into future generations, then frequency increases within the population (Slade & McCallum 1992). A final concept is that of 'diversifying selection'. This theorem incorporates both temporal and spatial variations that may be observed for differing genotypes within a population. It incorporates selective pressures that may change at differing times of the year and thus only individuals that possess the necessary alleles to counteract these threats will proliferate (Hill 1991). Under this concept, genetic variability is maintained as a constant but via fluctuating frequencies of differing alleles that are beneficial for the ecological context at a specific time.

The hypotheses that have been devised that explore the rationale behind such decisions of MHC type within a mate are similar to those that drive diversity from pathogenic pressures. The original impetus within this field of study stemmed from Lewis Thomas in 1974 who hypothesised that humans may be able to detect underlying HLA genes via human odour and thus influence their decision on selecting "histocompatibility donors" (Beauchamp & Yamazaki 1997). Laboratory tests in mice observed that in a number of cases mice will choose to breed with a mouse possessing an MHC haplotype differing from their own (Yamazaki et al. 1988). This observation suggested that odour cues originating from MHC genotypes led mice to avoid inbreeding. Further studies, again within mice, suggested that

individuals were able to detect conspecifics infected with a specific tumour virus (Beauchamp & Yamazaki 2003). This was purported to be as a consequence of the immune response cascade of reactions stimulated by MHC class II genes, from which T cell apoptosis led to specific odours. So, from laboratory studies with mice it appears that the MHC is involved in mate choice with regards to inbreeding avoidance and disease avoidance. Beyond the laboratory studies however, different species have been hypothesised to use several different strategies, which have the potential to equip offspring with favoured genes to ensure survival (Schwensow et al. 2008). One approach is referred to as dissassortative mating, whereby potential mates are chosen that harbour differing MHC types then their own, which has the potential to yield maximal MHC types within offspring (Bernatchez & Landry 2003). A second hypothesis is the 'good genes' approach, which is where an individual selects a mate that harbour specific alleles which may bestow offspring with a particular pathogenic resistance (Drury 2010). There are some deviations from these concepts that appear to be specific to the species or population under assay. Two empirical studies conducted in differing bird species surmised two different strategies of choosing a mate. A study at the PBR of a class I, exon 3 MHC region within a wild population of house sparrows (*Passer domesticus*) suggested that male sparrows were unsuccessful in securing a female mate if they harboured too few alleles at this locus but also if alleles were vastly different from their own (Bonneaud et al. 2006). The conclusions drawn were that females attempted not only to diversify MHC genes in progeny by increasing allele numbers, but also to avoid males that may disrupt co-adapted genes within their own loci. A similar preference but for locally adapted MHC alleles was found within MHC class II genes of the Great Snipe (*Gallinago media*), where mating success for males was not linked to

harbouring rare alleles, but rather those common to a specific habitat in the region (Ekblom et al. 2010). This, it may be assumed is a tactic to provide a form of defence against a known threat for future progeny. A further example of maintaining co-adapted genes was reported in three-spined stickleback fish (*Gasterosteus aculeatus*) (Eizaguirre et al. 2009). In this study, females were reported to select males for reproduction who harboured intermediate allelic diversity at MHC class II*B* genes, suggesting a preference for genes that would complement the female's own genetic makeup at this region. Furthermore, males that harboured a specific haplotype were more successful in finding a mate, as this linked group of alleles was hypothesised to have resistance against a parasite local to the population's habitat. Studies of a number of different primate species however have concluded that male reproductive success is linked with heterozygosity, as opposed to harbouring rare alleles. This was found in grey mouse lemurs (*Microcebus murinus*) (Schwensow, Eberle & Sommer 2008), mandrills (*Mandrillus sphinx*) (Setchell et al. 2010), rhesus macaques (*Macaca mulatta*) (Sauermann et al. 2001) and in microsatellites of ring-tailed lemurs (*Lemur catta*) (Charpentier, Boulet & Drea 2008).

There is however scepticism surrounding the link between MHC make up and how an animal chooses a mate. The detection of MHC linked odours for choosing a mate requires an olfaction system to be able to discern the different odour profiles produced by different genotypes. In mammals there is the main olfactory system and the vomeronasal (VNO) system, within which two superfamilies of pheromone receptors are present and believed to ascertain odours involved in mate choice (Horth 2007). However, in Catarrhines the olfactory bulb is smaller than in most other mammals (Kim et al. 2014) and the accessory bulb which is believed to

process cues with regards to sexual selection is entirely absent as is the VNO (Alport 2004). It is thought that in Catarrhines, the loss of these olfactory systems was as a consequence of the development of improved visual systems rendering the former more redundant (Kim et al. 2014). This has lead to questioning as to the level of contribution that MHC odour profiles contribute to the process of choosing a mate and thus the validity of carrying out costly genetic analyses. Furthermore, there are other cues which may be argued play a greater role in informing a potential mate of fitness levels, such as via sexual secondary sexual characteristics or odour profiles from scent glands (Setchell et al. 2010). However, birds are believed to be anosmatic and the study of house sparrows (*Passer domesticus*) could find no correlation between mate choice success and visual fitness cues such as body mass (Bonneaud et al. 2006). Yet there was a strong correlation between a partners MHC diversity and moreover, in relation to the bird's own MHC make up.

4.1.3 The value of analyses of MHC class II genes for captive breeding and reintroduction programmes

Analysis of MHC class II genes, particularly at the PBR's, can potentially provide extremely valuable information for captive breeding and reintroduction programmes. Analyses of neutral markers, such as microsatellite loci are used with on the basis that they are related to fitness levels within an individual or population. The genes within class II of the MHC however are involved in an individual's immune response and thus they may also be regarded as having a direct influence on fitness levels of an organism (Trowsdale 2011). With regards to captive breeding, the selection process to pair individuals would benefit from knowledge of the levels of each individual's genetic variability at this vital region. The information would also provide

details of specific alleles and thus if rare alleles are harboured by a particular individual, information that is not available from proxy measures of genetic variability.

The subject area of mate choice within the zoological environment has drawn a heightened level of focus within the last few years, with a symposium on the subject held in the United States in 2010 (Asa, Traylor-Holzer & Lacy 2011). One of the major reasons for the interest is due to the subject of sustainability of zoo populations, or the lack of it. The problem lies within unsuccessful breeding attempts, largely attributed to pair incompatibility (Asa, Traylor-Holzer & Lacy 2011). Furthermore, breeding of endangered species utilising the Mean Kinship objectives within breeding programmes often requires translocation of animals owing to small population numbers and desirability of inbreeding avoidance. Moving animals, such as large mammals can be stressful for the organism in question, and costly and the success of pairings is an unknown factor. So, if information can be derived from genetic data prior to translocation that has the potential to increase success of compatibility in addition to inbreeding avoidance then it would be most helpful. For example, if rare alleles are more preferable in a species, a male with rare alleles may be selected and has the possibility to have an increased chance of acceptability by the female. In some species, such as within primates, alterations of social groups such as introducing a new potential new mate, can have significant negative effects on behaviour often manifesting in aggression (Hosey 2005). Thus, attempting to pair individuals that are not compatible has the potential to yield a two-fold negative outcome in that breeding will be unsuccessful, and individuals may also be subjected to stress and aggression in the process. A caveat should also be

added here however, it is possible that if any events of adaptation to the captive environment have occurred at the MHC owing to increased or decreased exposure to pathogenic pressures, the actual process of selecting a mate has the potential to change in accordance with this. Thus, hypothetically speaking if a parasite residing in a specific species wild habitat influences mate choice for partners that harbour the necessary immune genes to combat infection, this behaviour may not be mirrored in a captive environment. A study of MHC linked mate choice behaviour in rhesus macaques (*Macaca mulattta*) found that males were more successful in producing progeny if they were heterozygous at the DQB1 locus (Sauermann et al. 2001). However, the correlation was only found in free ranging populations and not in a captive population studied, where individuals receive regular veterinary care which includes bi-annual anti-helminth medication

An issue highlighted by Frankham (2005) is how losses of genetic variability within small populations decrease a population's capacity to environmental change. Individuals bred in captivity and destined for reintroduction are undoubtedly going to experience changes within their environment. Alterations within environmental conditions are known to exert varying pressures on populations and thus ability to adapt is vitally important. Examples of such responses cover a wide variety of taxa such as *Acacia depranolobium,* a tree found within the African savannah was found to have increased thorn length in trees that were subjected to herbivory by local goats, thus had formed a defence to protect itself (Young 1987). A thirty year study into the evolutionary changes of Darwin's finches illustrates how changes in abiotic factors such as climate and thus trophic availability has led to phenotypic alterations in order for their continued survival (Grant & Grant 2003). Indeed, the 14 finch

species now residing on the Galápagos Islands have evolved from a common ancestor 2 to 3 million years ago and have experienced alterations in beak size so as to take advantage of differing trophic resources on differing islands and at differing periods of the year (Grant & Grant 2003). The dramatic increase in the frequency of the melanic form of a number of moth species (*Biston betularia*, *Odontoptera bidentata*, and *Apamea crenata*) within industrialised areas of the United Kingdom is illustrative of an alteration in pigmentation so as to be camouflaged against trees that have been affected by pollution (Cook, Sutton & Crawford 2005). The recent genome wide study of a *Nomascus leucogenys* individual revealed the genes that have been key players in forelimb development of the gibbon species, which after periods of positive selection has culminated in the long forelimbs enabling the brachiating movement within gibbons observed today (Carbone et al. 2014). These examples are illustrative of natural selection, external environmental conditions exerting pressure on differing organisms that has then been translated into phenotypic variation in order to increase chances of survival.

If ecological conditions serve as selective pressure then alterations in MHC genes may occur as a consequence to the captive environment. At the MHC, two hypotheses may be postulated. The first is that a reduction in MHC class II variability may occur as a consequence of the relaxed environment in captivity. A zoological institution will give veterinary care to a sick animal to ensure the best health and welfare, thus relieving one of the pressures that would be present within a wild type environment. Within the wild scenario, the individual will not have the chance to transmit genetic material to the next generation, however within the captive environment the opposite may be true and thus it is not just the fittest that

survive but also the weak. The second hypothesis is that some pathogenic pressures may actually be greater in the captive environment, owing to factors such as food contamination and possible closer proximity of living enclosures enabling easier transmission. The pathogenic threats within the captive environment may in fact surpass those present within the wild environment, and thus increase variation at class II genes over a number of generations.

It is essential, however, that individuals destined for release back into their wild environment after habitation in captivity harbour the ability to react to novel pathogens. These may be localised and specific to their wild habitat, but they may also arise as a consequence of climate change (Burek, Gulland & O'Hara 2008; La Porta et al. 2008). Not only can climate change alter abiotic environmental conditions, but these can then cause alterations in biotic interactions and life cycles of the organisms inhabiting affected areas. One such example is the prevalence of a chytrid fungus *Batrachochytrium dendrobatidis* that has been the root cause of a number of amphibian extinctions (Pounds et al. 2006). One of the hypotheses that sought to find the aetiology of such great losses of amphibian numbers was that alterations in climatic conditions allowed for optimum conditions for the fungal organism to mutate and exhibit pathogenic characteristics (Fagotti & Pascolini 2007).

4.1.4 Choice of MHC DRB typing method

The differing manifestations of polymorphism at the MHC such as allelic variation, and copy number variation render typing of this region difficult (Babik 2010; Doxiadis

et al. 2007). The complexities of correctly encapsulating different alleles and furthermore how they may be positioned in haplotypes are difficulties in themselves but they can also differ among the same species (Babik 2010). One issue highlighted by Babik in his review of MHC typing methods (2010) was the use of primers required to amplify a desired locus. There is no current information available as to the MHC genomic organisation within *H. moloch*, rendering the creation of Sequence Specific Primers a difficult and costly task. Among the different approaches listed by Babik (2010) is the utilisation of microsatellites linked to MHC genes. The same issue remains however, as with the selection of primers to amplify the desired regions that the microsatellite must be identified and primers designed. In addition, the microsatellite itself must be sufficiently polymorphic to identify differing MHC alleles in accordance with microsatellite allele lengths. A microsatellite marker deemed to be a promising candidate for use within the *H. moloch* groups was published by Doxiadis *et al.* (2007). A highly complex dinucleotide microsatellite marker known as D6S2878 was utilised in the aforementioned study to facilitate haplotyping of rhesus macaques (*Macaca mulatta*) at the MHC class II DRB exon 2 region. The microsatellite is located within intron 2 of the *DRB* region and thus adjacent to the highly polymorphic exon 2 containing the PBR. This microsatellite had already been found within the human HLA-DRB genes which enabled the identification of 5 region configurations of differing haplotypes comprised of differing alleles with high allelic variation (de Groot et al. 2009; Doxiadis et al. 2007). Within the macaques however, allelic variation was lower, but more than 30 region configurations were identified. This approach to use D6S2878 to explore DRB exon 2 variability has also been applied within chimpanzees (*Pan*

troglodytes troglodytes and *Pan troglodytes verus*) (de Groot et al. 2009) and long-tailed macaques (*Macaca fascicularis*) (de Groot et al. 2008).

Prior studies utilising D6S2878 as a means to identify *DRB* alleles and haplotypes have stated that the sensitivity of this marker is so great it is able to distinguish highly similar alleles via the nature of both the content and length variability of the microsatellite (de Groot et al. 2008; de Groot et al. 2009; Doxiadis et al. 2007; Doxiadis et al. 2010). As opposed to a more traditional microsatellite where two allele peaks manifest in heterozygotes and a single peak for homozygotes, constellations for this marker can be numerous, owing to its close association to differing *DRB* alleles at potentially differing loci.

The chimpanzee research into DRB exon 2 utilised human primers which was successful, however the study by Doxiadis and colleagues (2007) on macaques (*Macaca mulatta*) utilised species specific primers developed for the species. As *Hylobates moloch* shares membership of the primate superfamily 'Hominoidae' with humans (Müller, Hollatz & Wienberg 2003), it was decided that primers would be based on human HLA (Doxiadis et al. 2007).

4.2 Materials and Methods

4.2.1 Sample population

The sample population assayed at the MHC class II exon 2 region comprised of 21 individuals as per chapter 2.2.1.

4.2.2 DNA extraction

DNA was extracted as per chapter 2.2.2.

4.2.3 Parentage ascertainment

In order to ensure accuracy for microsatellite-*DRB* haplotyping, results from microsatellite analyses in the previous chapter were utilised. This was to observe that alleles are segregating in accordance with Mendelian inheritance parameters and thus D6S2878 microsatellite results can be correctly assigned within familial units.

4.2.4 D6S2878 – *DRB* genotyping

All laboratory procedures from this point were carried out at the Biomedical Primate Research Centre, Rijswijk, Netherlands.

In order to encapsulate both the microsatellite D6S2878 and exon 2 of the MHC class II DRB region, a forward primer located at the end of exon 2 (5' HLA-DRB-STR_VIC: GAG AGC TTC ACA GTG CAG C) (Applied Biosystems, USA) and a reverse primer located within intron 2 (3' HLA-DRB-STR: GAG AGG ATT CTA AAT GCT CAC) (Invitrogen, UK) were utilised (Doxiadis et al. 2007). The forward primer

was labelled with a VIC fluorescent tag for detection during capillary electrophoresis and subsequent genotyping analysis. The resultant PCR reaction components were as described by Doxiadis *et al.* (2007) and were performed in a 25µl reaction volume as follows: 1 unit of Platinum Taq polymerase (Invitrogen, UK), 0.1 µM of VIC labelled forward primer, 0.1 µM of the aforementioned reverse primer, 2.5mM $MgCl_2$, 0.2mM of each dNTP, 1xPCR buffer (At x10 concentration contains 200 mM Tris HCl (pH 8.4), 500 mM KCl) (Invitrogen, UK) and 100ng template DNA. Lukas and Vigilant (2005) highlighted a number of approaches to minimise erroneous results when utilising faecal samples for studies within the MHC. This study adhered to the majority of suggestions, such as quantification of DNA and replication of PCR for each individual.

PCR cycling conditions were as follows: initial denaturation for 5 min at 94°C, then 5 cycles of 1 min at 94°C, 45 sec at 58°C, 45 sec at 72°C. Then 25 cycles commencing with 45 sec at 94°C, 30 sec at 58°C and 45 sec at 72°C. The final extension period was set at 72°C for a period of 30 min. Resultant PCR products were then run in a 1.4% ethidium bromide gel. This process was then repeated for quality control purposes. PCR products that exhibited clear, singular and bright bands under UV light were then selected and extracted and purified as per the protocol provided within the GeneJET Extraction Kit (Thermo Scientific, USA). Resultant PCR products were then prepared for genotyping via capillary electrophoresis in accordance with manufacturer's guidelines and analysed on an ABI 3130xl sequencer (Applied Biosystems). Genotyping analysis of microsatellite peaks were performed using GeneMapper® software version 5.0 (Applied Biosystems) as per the method described in chapter 3.2.5.1.

4.2.5 Cloning of PCR products and sequencing

4.2.5.1 PCR preparation for cloning

A second PCR was then carried out to procure the target gene segment as described above but to be utilised for cloning to obtain the differing alleles. The fluorescently labelled forward primer utilised for DRB-STR genotyping was replaced by a generic unlabelled forward primer (5' DRB-STR CGT GTC CCC ACA GCA CGT TTC). The reverse primer was as used in D6S2878-*DRB* genotyping. The PCR reaction components totalling 100µl per reaction were as follows: 4 units of Platinum Taq polymerase (Invitrogen, UK), 0.2 µM of forward primer, 0.2 µM of reverse primer, 2.5mM $MgCl_2$, 0.2mM of each dNTP, 1xPCR buffer II (At x10 concentration contains 200 mM Tris HCl (pH 8.4), 500 mM KCl) (Invitrogen, UK) and 100ng template DNA. The PCR cycling conditions were also as per the protocol for D6S2878-*DRB* genotyping.

4.2.5.2 Preparation of competent cells

The first stage of the cloning procedure was to prepare day culture bacterial colonies. Preparation of the agar was carried out as follows by mixing 32g Luria Broth Agar (Invitrogen, UK) with 1L of distilled water. This was then sterilised at 121°C for 15 minutes then allowed to cool to 55°C. A tetracycline solution was then made with 50mg tetracycline and 10ml 50% ethanol (thus resulting in a 5mg/ml solution). For each individual plate, 30ml of LB broth was mixed with 60µl of tetracycline solution. Bacteria, *E.coli* XL1-Blue (a tetracycline resistant strain) was then streaked onto the agar and stored in order to allow for colonies to develop.

4.2.5.3 Ligation and transformation

The vector used for cloning *DRB* exon 2 alleles was the pJET 1.2 vector (Thermo Scientific, USA) which can be transformed by competent *E.coli* XL1-Blue strains of bacteria. It functions via a lethal gene that is disrupted once a DNA amplicon is successfully inserted, thus only allowing bacterial cells that harbour the insert to propagate.

As the PCR product from section 4.2.5.1 results in an amplicon with sticky ends, the protocol carried out for ligation commenced with a blunting reaction as per the instructions provided with the CloneJET PCR Cloning Kit (Thermo Scientific, USA). Workflow for ligation that followed was as per the protocol described with the exception of the final step where the ligation mix was put on a thermomixer at 22°C for two hours.

The transformation procedure of the recombinant plasmids with the *E.coli* XL1-blue colonies was carried out as per the protocol described in the TransformAid Bacterial Transformation Kit (Thermo Scientific, USA). The only exception was that at step 6 of the protocol 10µl and 15µl (for strong bands and weak bands respectively) of ligation mixture were added to the pelleted cells in place of the 5µl stated within the instructions. This was carried out to attempt to maximise results as DNA was extracted from faecal samples. Transformed cells were plated on pre warmed LB agar plates that had been prepared as with the above mentioned protocol for *E.coli* XL1-Blue preparation. As the vector pJET 1.2 contains an ampicillin resistant gene (*β*-lactamase) agar plates were prepared with this. Ampicilline at a concentration of

50mg/mg. Ampicilline solution was added to agar broth base solution at a concentration of 50µg/ml.

4.2.5.4 Isolation of DRB exon 2 alleles

After colonies containing the DNA insert were allowed to cultivate overnight they were then picked and added to a 96 deep well plate containing 1.25ml of ampicilline LB medium (with a concentration of 1ml of ampicilline to 500ml of LB medium). For bands that had been clear and bright 48 colonies were picked for that individual and 96 were selected for individuals that had exhibited weaker bands. The plates were then covered with an airpore strip and placed in an incubator at 37°C and mixed overnight.

The protocol that then followed was a modification from Qiaprep® Miniprep kit (Qiagen). Deep well plates were removed from the incubator and centrifuged for 10 mins at 3600rpm. The supernatant together with sticks utilised to pick colonies were discarded. An aliquot of 100µl of P1 reagent (Qiagen) was added to each deep well and then vortexed until the pellet was loosened. This was then followed by the addition of 100µl of P2 reagent (Qiagen) and plates were gently shaken. 100µl of P3 (Qiagen) was then added to each well within a 5 minute time frame from the addition of P2 and plates were then placed on ice for 5 mins. 50µl of isopropanol was added to 1ml 96 DeepWell™ (Thermo Scientific) plates and then a 96 well filter plate was placed on top (Nunc™, Thermo Scientific). Then 250µl of the lysate was added to the filter plate and both plates were then centrifuged at 3800rpm for 5 mins. Then filterplates were removed and the DeepWell™ plates were covered with a coverstrip and centrifuged at 3800rpm for a further 45 mins. The resultant

supernatant was then discarded, with the pellets remaining in each well. Plates were then placed upside down on paper. To remove residual isopropanol pellets were then washed by the addition of 300µl of 70% ethanol and then centrifuged for 5 mins at 4000rpm. The ethanol was then discarded and plates placed upside down on paper and centrifuged for 2 mins at 900rpm. To dissolve the pellet an amount of 60µl of PCR water was added to each well and incubated at 37°C for 5 to 10 mins. Then finally each plate was vortexed. Plates were then centrifuged once more for 1 min.

4.2.5.5 Sequencing

Preparation for sequencing was carried out with the following steps. An aliquot of 6µl of DNA was added to the respective wells within the 96 well plate. A sequencing master mix consisting of the following per reaction, was then prepared with 1µl of Big Dye™ v3.1 (Applied Biosystems), 1µl forward for pJET 1.2 (CloneJET PCR Cloning Kit, Thermo Scientific), 2µl 5 x Big Dye™ v3.1 buffer (Applied Biosystems). Then 4µl of the sequencing master mix was added to each well and gently mixed. Plates were then placed in the PCR machine to mix solutions together with the following cycling conditions: 25 cycles of 94°C for 10 secs, 50°C for 10 secs, 60°C for 2 mins, then finally 4°C for 2 mins.

A denaturation mix was prepared to precipitate DNA and to help specify base pairs so as to minimise adhesion of DNA to the Big Dye™. A 30µl mix per reaction was prepared with 1µl PCR water, 1µl of 7.5M ammonium acetate and 28µl of 100% ethanol. An amount of 30µl of denaturation mix was added to each well and then plates were centrifuged for 30 mins at 3800rpm. The supernatant was discarded

and then 40µl of 70% ethanol was added to each well. Plates were then placed upside down on paper and centrifuged for 1 min at 1000rpm. The pellet was then re-suspended by the addition of 30µl of PCR water to each well. Plates were then gently mixed by vortexing following by a final round in the centrifuge for 1 min at 1000rpm. The plates were then placed in the PCR machine and a 2 min heatshock at 94°C was carried out to break DNA strands. Plates were then placed in the 3130xl Genetic Analyzer for sequencing (Applied Biosystems).

4.2.6 Data analysis

Upon completion of capillary electrophoresis, files containing nucleotide sequences of *Hylobates moloch DRB* exon 2 and the microsatellite D6S2878 were imported into the software SeqMan Pro version 12.1.0 (DNASTAR) for analysis. Sequences were manually aligned for each subject using HLA DRB1*0101 as a consensus sequence from which to work. The software is pre-loaded with vector sequences such as the pJET1.2 (Thermo Scientific) utilised here. This is a useful tool to easily identify vector nucleotide sequences from target amplicon sequences. Sequence alignments were then analysed by using the BLAST® (Basic Local Alignment Search Tool, National Centre for Biotechnology Information) tool that is embedded within the programme to identify similar alleles already identified and published within the GenBank programme.

Manually aligned sequences from SeqMan Pro software containing the DRB exon 2 segment (excluding the microsatellite) consisting of 263 base pairs were then entered into the software programme MacVector™ version 13.0.7 (Oxford Molecular Group). In order to identify loci and corresponding alleles for *H. moloch* sequences,

MHC class II DRB sequences and the corresponding D6S2878 sequences from human (HLA) and chimpanzee (Patr) and one sequence from a common marmoset (Caja) were aligned within MacVector™ using the embedded ClustalW multiple sequence alignment option, with default settings. All sequences were provided by Dr. Gaby Doxiadis from BPRC, Netherlands. MHC sequences may also be downloaded from the Immuno Polymorphism Database (EMBL-EBI) (Robinson et al. 2014a&b).

To further understand the lineage and identify genes and alleles for *H. moloch* sequences all aligned sequences were added into the software programme MEGA version 5.2.2 (Tamura et al. 2013) in order to construct a phylogenetic tree. The first step carried out to construct a tree was to select the best model that corresponded to underlying nucleotide sequences. This was performed in MEGA and the model with the lowest Bayesian Information Criterion was selected as per the programme recommendations. The resultant model to estimate evolutionary distances was the Kimura 2-parameter method (Kimura 1980) with an added rate variation modelled with a gamma distribution of 0.9 as per programme recommendation parameters. This evolutionary model analyses homologous nucleotide sequences incorporating different weightings of occurrences of transitions to transversions, and assumes that transversion substitutions are less frequent. The methodology utilised to construct the phylogenetic tree was the neighbour joining method (Saitou & Nei 1987) which has been employed for other analyses of this genetic region (eg. Doxiadis *et al.* 2007). This algorithm creates tree topologies based on underlying matrices of pairwise distances between nucleotide sequences. The method identifies pairs of operational taxonomic units by way of minimum evolution between differing tree

branches whose lengths are minimized at each step of clustering (Saitou & Nei 1987). A bootstrap test of 1000 replicates was also carried out to ascertain the percentage of resultant trees that clustered together in the final topology selected.

4.2.6.1 Tests for diversity and selection

To ascertain nucleotide diversity (π) within sequences the software programme DNaSP version 5.0 was used (Librado & Rozas 2009).

To test for evidence of possible selection within the MHC class II exon 2 fragment, rates for synonymous (d_S) and non-synonymous (d_N) substitutions were calculated within MEGA version 5.2.2 (Tamura et al. 2013). This was performed using the methodology of Nei and Gojobori (1986) with application of the Jukes Cantor correction (1969). The variance estimation was performed with 1000 bootstrap replications. The method is a simpler approach than in some other methodologies whereby differing weighting methods are applied if a codon bears a synonymous versus a non-synonyous substitution. It also ignores differential weightings applied in accordance with the number of evolutionary pathways that may occur between a pair of codons. In the Nei and Gojobori (1986) method, weighting for all pathways is equal. The application of Jukes and Cantor (1969) also simplifies the computations as this approach gives equal weighting to both transitional and transversional substitutions, and thus all nucleotide alterations may occur at equal frequencies. Although these combined methodologies are simplified in their approach, they yield comparable results with more complicated versions (Nei & Gojobori 1986). In addition, these methods are useful for studies of MHC sequences which are

potentially highly polymorphic and thus harbour a higher number of substitutions than other areas of the genome.

A simple test for positive selection was then performed by calculating the d_N / d_S ratio. If the result is greater than unity then the sequences under assay are assumed to be under positive selection. To further scrutinize the d_N / d_S ratio a one-tailed z test was also performed in MEGA version 5.2.2. Neutral selection assumes that d_S is equal to d_N and positive selection assumes that d_N is greater than d_S and both use the Nei-Gojobori (1986) method with the Jukes Cantor correction (1969) for calculations. As with d_N and d_S calculations, the calculation was performed with bootstrap variance estimates using 1000 replications. The Z- test calculation is computed as follows:

$$Z = \frac{d_N - d_S}{\sqrt{Var\ (d_S) + Var\ (d_N)}}$$

4.3 Results for MHC Class II DRB exon 2 analysis

4.3.1 MHC alleles and hypothesised haplotypes

From the 21 individuals analysed within the *H. moloch* population a total of 14 MHC class II *DRB* exon 2 alleles were detected. These have been deposited in GenBank with the accession numbers: KJ701253 – KJ701266. The alleles are named and designated in accordance with Immuno Polymorphism Database – MHC nomenclature rules (Ellis et al. 2006) and are shown in Table 15. In accordance with the nomenclature rules, alleles for *H. moloch* are named *Hymo.*

Table 15 - New *Hylobates moloch* (*Hymo*) MHC class II DRB exon 2 allele designations.

*Hymo-DRB1*04:01*	*Hymo-DRB*W096:01*
*Hymo-DRB1*04:02*	*Hymo-DRB*W096:02*
*Hymo-DRB1*04:03*	*Hymo-DRB*W097:01*
*Hymo-DRB1*04:04*	*Hymo-DRB*W100:01*
*Hymo-DRB*W094:01*	*Hymo-DRB*W098:01*
*Hymo-DRB*W094:02*	*Hymo-DRB*W098:02*
*Hymo-DRB*W095:01*	*Hymo-DRB*W099:01*

Alleles in Table 15 designated with the letter 'W' are stated as workshop numbers as there are currently no HLA lineage equivalents (Doxiadis et al. 2000). This is illustrated in Figure 5 where the phylogeny of *Hymo,* chimpanzee (*Patr*) and human (*HLA*) are shown. The majority of *Hymo* sequences form their own clades and hence do not share lineages with either primate species with which sequences are aligned. There are however, four sequences that form a clade with the human allele

*HLA-DRB1*04:01:01* and have thus been associated with this allele group. The only other *Hymo* allele that forms a clade with both chimpanzee and human alleles is that of *Hymo-DRB*W095*01.* However, the genetic distance within the *moloch* species is too far from the *HLA* and *Patr* sequences and thus does not receive a homologous designation. Table 16 incorporates the D6S2878 microsatellite sequence and further illustrates the similarities between the *H. moloch,* chimpanzees and humans. It shows the compound nature of the D6S2878 microsatellite and how it modifies in accordance with particular *DRB* alleles, some length variability is found even within the same *DRB*-linked DNA repeat. For example alleles designated *Hymo-DRB*W095:01*, *Hymo-DRB*W094:02* and *Hymo-DRB*DRB1*04:01* all have slight variations in length in the first section (GT) of the D6S2878 microsatellite.

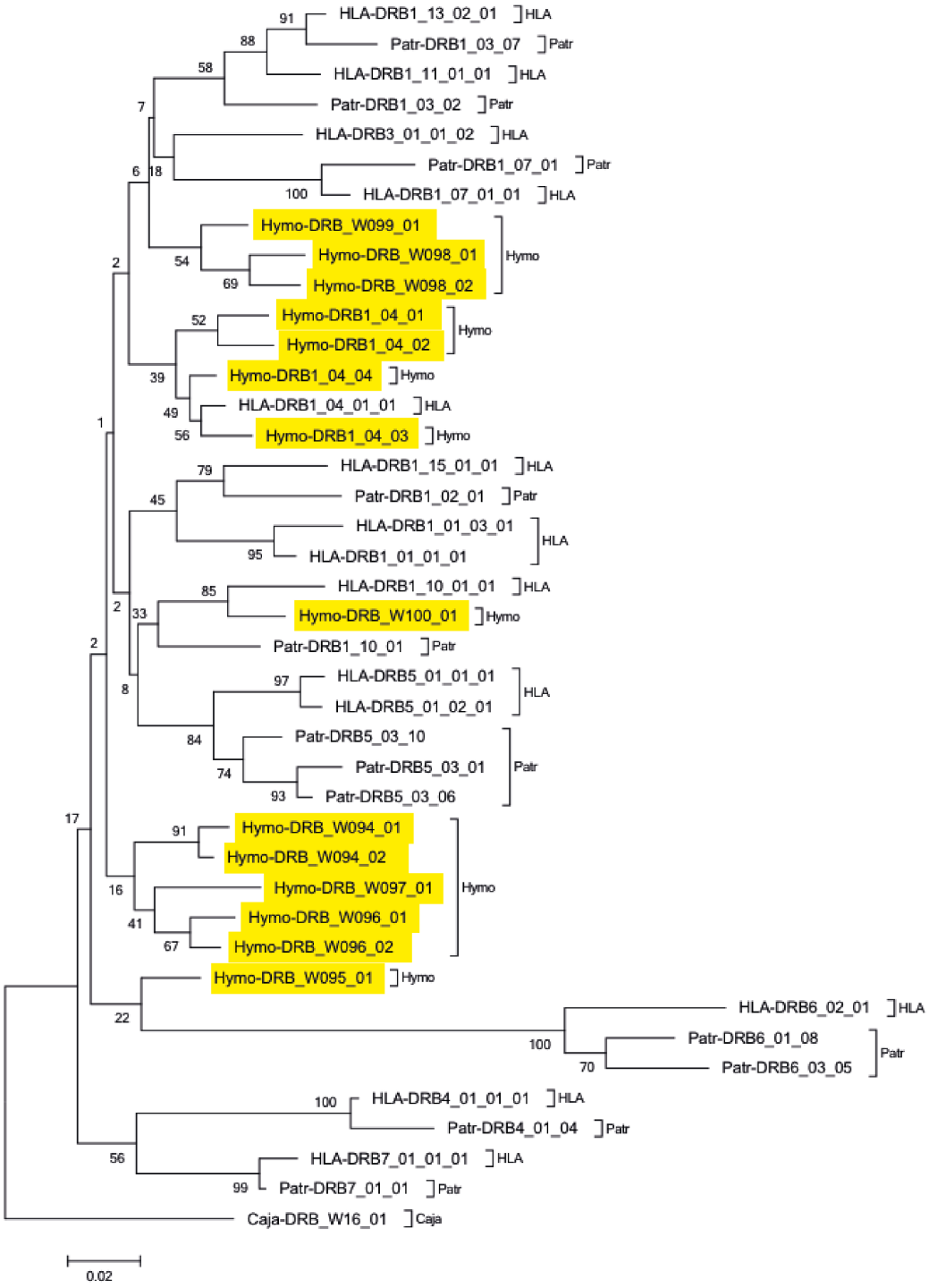


Figure 5 - Neighbour joining phylogenetic tree of *Hylobates moloch* (*Hymo*) D6S2878 and *DRB* exon 2 alleles aligned with chimpanzee (*Patr*) and human (*HLA*) sequences. The percentage of trees after 1000 bootstrap replications are shown next to the branches. The root sequence is from a common marmoset (*Caja*).

Table 16 –– Aligned DRB exon 2 alleles with their respective D6S2878 microsatellite constituent parts. Numbers in brackets correspond to number of repeats. Part 2 highlighted sections

DRB alleles	Part 1 (GT)n	Part 2 mixed	Part 3 (GA)n	Part 4 (GC)n
HLA-DRB1*13:02:01	(GT)20,30,31	(GA)10-12CA(GA)3CA	(GA)6	(GC)2
Patr-DRB1*03:07	(GT)11-13		(GA)17-23	(GC)2
HLA-DRB1*11:01:01	(GT)22-26	(GA)5CA(GA)3CA	(GA)6	(GC)2
Patr-DRB1*03:02	(GT)23-28	(GA)5AA	(GA)6	(GC)4
HLA-DRB3*01:01:02	(GA)3(GT)13,14	(GA)9,10GGAA(GA)2CA(GA)3GG	GA	(GC)3
Patr-DRB1*07:01	(GT)2GGTT(GT)14	(GA)4GC	(GA)2	CCGC
HLA-DRB1*07:01:01	(GT)11	(GA)8GC	(GA)2	CCGC
HLA-DRB1*15:01:01	(GT)15-20	(GA)5-6CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
Patr-DRB1*02:01	GTGA(GT)19-22	(GA)12-14CA(GA)4CA(GA)3GGAA	(GA)6-7	(GC)2
HLA-DRB1*01:03:01	(GT)16	AAGAAA	(GA)4	(GC)3
HLA-DRB1*01:01:01	(GT)16,17	AAGAAA	(GA)4	(GC)3
Hymo-DRB*W098:01	(GT)6CT(GT)8,9		(GA)14	GC
Hymo-DRB*W098:02	GTGA(GT)5GA(GT)7	(GA)2(CA)2	(GA)4	GC
Hymo-DRB*W099:01	(GT)6,7CT(GT)13		(GA)11-13	GC
Hymo-DRB1*04:01	(GT)21-24	GA(CA)2	(GA)4	GC
Hymo-DRB1*04:02	(GT)19	GA(CA)2	(GA)4	GC
Hymo-DRB1*04:04	(GT)21	GA(CA)2	(GA)4	GC
HLA-DRB1*04:01:01	(GT)21-22		(GA)15,16	(GC)2
Hymo-DRB1*04:03	(GT)12	(GA)2(CA)2	(GA)4	GC
HLA-DRB1*10:01:01	(CT)2(GT)16		(GA)8	(GC)3
Hymo-DRB*W100:01	(GT)4GA(GT)2TT(GT)5	(GA)12GGAA	GA	GC
Patr-DRB1*10:01	(GT)9-10	(GA)11-12)CA(GA)4AA	(GA)5	(GC)3
HLA-DRB5*01:01:01	(GT)18-24	(GA)5-8GGAA(GA)4CA(GA)2GG	GA	(GC)3
HLA-DRB5*01:02:01	(GT)22	(GA)8GGAA(GA)4CA(GA)7GGAA(GA)4CA(GA)2GG	GA	(GC)3
Patr-DRB5*03:10	(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
Patr-DRB5*03:01	(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
Patr-DRB5*03:06	(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
Hymo-DRB*W096:01	(GT)20	(GA)2(CA)2	(GA)4	GC
Hymo-DRB*W096:02	(GT)13	GA(CA)2	(GA)4	GC
Hymo-DRB*W097:01	GC(GT)20	GACA	(GA)7	GC
Hymo-DRB*W095:01	(GT)20-25	GG(CA)2	(GA)4	GC
Hymo-DRB*W094:01	(GT)18	GG(CA)2	(GA)4	GC
Hymo-DRB*W094:02	(GT)18,19	GG(CA)2	(GA)4	GC
Patr-DRB6*01:08	(GT)4	GAGGGCA(GG)2TC(GG)3GCAG	(GA)6	
Patr-DRB6*03:05	(GT)4	GGGAGGA(GG)3GTGGA(GG)2CGATAGG	(GA)8	
HLA-DRB6*02:01	(GT)13-24		(GA)11-14	
HLA-DRB4*01:01:01	GTAT(GT)9-11	(GA)9-14(CAGA)1,2GGAA	(GA)5	GC(GT)1,2(GC)1,2
Patr-DRB4*01:04	GTAT(GT)4-5	(GA)8-9(CAGA)2(GA)10CAGATGAA(GA)3AA	GA	GCGT(GC)2
HLA-DRB7*01:01:01	(GT)2TT(GT)3T	(GA)5(CA)6		
Patr-DRB7*01:01	(GT)2TT(GT)3T	(GA)6(CA)5		

Table 17 illustrates hypothesised haplotypes present within the *H. moloch* population as defined by the primary stage of genotyping and then confirmed by sequencing. Where a question mark is present it signifies that a microsatellite was detected from the genotyping but the alleles could not be confirmed by sequencing. This configuration intimates that there are 11 haplotypes, with haplotype 3 appearing with two configurations as they only differ by one allele at the first locus. Thus they are designated as 3a and 3b. The alleles *DRB*W094:02* differs from *DRB*W094:01* by only 3 base pairs resulting in 1 non-synonymous amino acid alteration. Figure 6 shows the pedigree data of how each MHC haplotype has been inherited.

Table 17 - *Hymo* DRB haplotypes defined by both exon 2 sequencing and DRB-D6S2878 microsatellite (STR) genotyping.

Hap.	1st locus	STR	2nd Locus	STR	3rd Locus	STR
1	*DRB*W094:02*	155	*DRB*W096:01*	159		
2	*DRB*W094:02*	155	*DRB*W099:01*	169	*DRB*W098:01*	161
3a	*DRB*W094:02*	155	*DRB*W099:01*	169	*DRB*W097:01*	166
3b	*DRB*W094:01*	155	*DRB*W099:01*	169	*DRB*W097:01*?	166
4	*DRB*W095:01*	160	*DRB1*04:01*	163	?	146
5	*DRB*W095:01*	166	*DRB*W099:01*	169		
6 *	*DRB*W096:02*	149	*DRB*W100:01*	162	*DRB*W098:02*	143
7	*DRB1*04:01*	165	*DRB*W099:01*	167	*DRB*W098:01*	163
8 *	*DRB1*04:02*	157	?	160		
9	*DRB1*04:03*	143	*DRB*W099:01*	169	*DRB*W098:01*	161
10	*DRB1*04:04*	161	*DRB*W099:01*	169	*DRB*W098:01*	163

Figure 6 - Pedigree of MHC haplotypes. Shaded boxes represent individuals not sampled within this study and estimated haplotypes inferred from parents

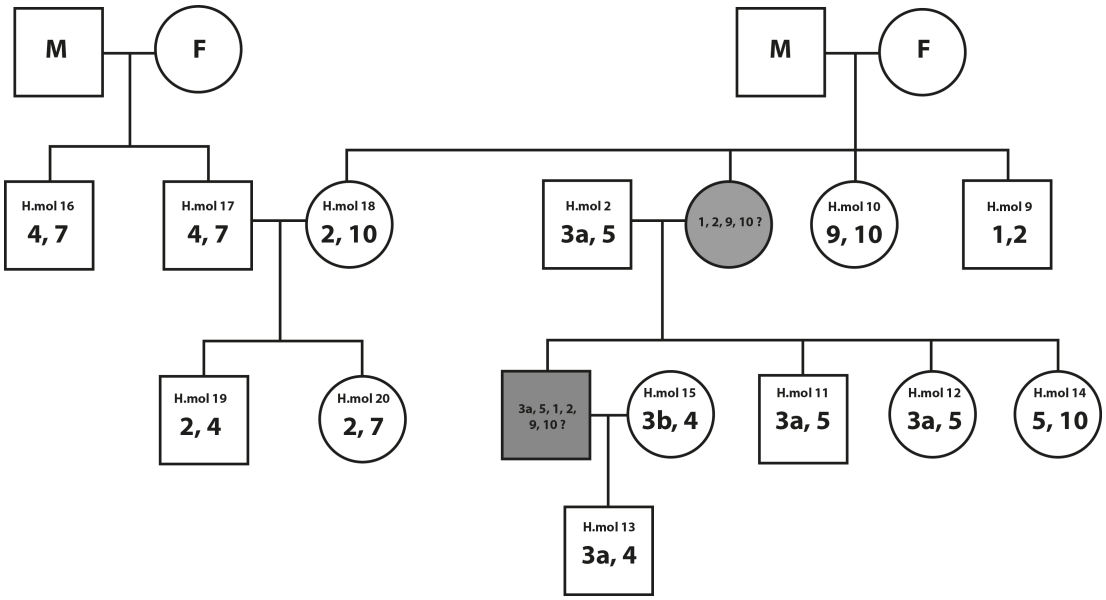


Table 18 shows the MK values and the respective MHC haplotypes per individual. There are three unique haplotypes which as seen in mtDNA are not all attributable to individuals with the lowest MK values. The latter part of the table however shows more common haplotypes within the groups with increasing MK values which shows that the MK approach has worked.

Table 18 - Mean kinship (MK) values and MHC haplotypes per individual. Sorted by MK value. Highlighted cells represent unique haplotypes.

Primate Ref	Mean Kinship	MHC Haplotypes	
H.mol8	**-**	**7**	**8**
H.mol21	-	9	10
H.mol17	-	4	7
H.mol6	**0.0000**	**1**	**7**
H.mol4	**0.0000**	**-**	**-**
H.mol3	**0.0000**	**1**	**4**
H.mol5	**0.0083**	**1**	**3a**
H.mol1	**0.0083**	**6**	**?**
H.mol7	**0.0208**	**3a**	**5**
H.mol15	0.0604	3b	4
H.mol16	0.0688	4	7
H.mol2	**0.0875**	**3a**	**5**
H.mol13	0.1083	3a	4
H.mol20	0.1167	2	7
H.mol19	0.1167	2	4
H.mol14	0.1271	5	10
H.mol12	0.1271	3a	5
H.mol11	0.1271	3a	5
H.mol9	0.1354	1	2
H.mol10	0.1479	9	10
H.mol18	0.1521	2	10

4.3.2 MHC class II DRB exon 2 measures of diversity

Figure 7 shows how allele frequencies are distributed within the Captive and Wild groups. The most common allele to both groups is *DRB*W099:01,* but the frequency is greatest with in the Captive group. There are alleles private to each group, *DRB*W094:01, DRB1*04:03* and *DRB1*04:04* within the Captive individuals and *DRB1*W094:02, DRB*W096:02, DRB1*W098:02* and *DRB*W100:01* within the Wild group.

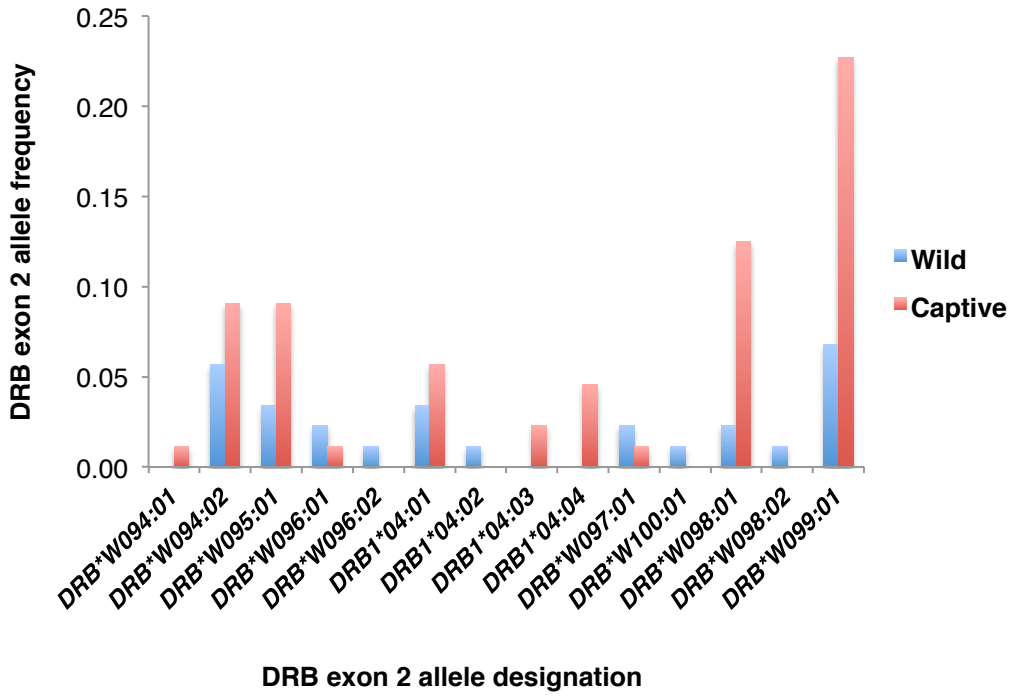


Figure 7 - Allele frequencies in the *Hylobates moloch* population

Figure 8 shows how haplotypes are distributed in the *H. Moloch* population. The distribution of haplotype 3a is equal within both Captive and Wild groups. There are fewer unique haplotypes in the Wild group as evinced by haplotypes 6 and 8, however the Captive group maintain haplotypes 2, 3b, 9 and 10.

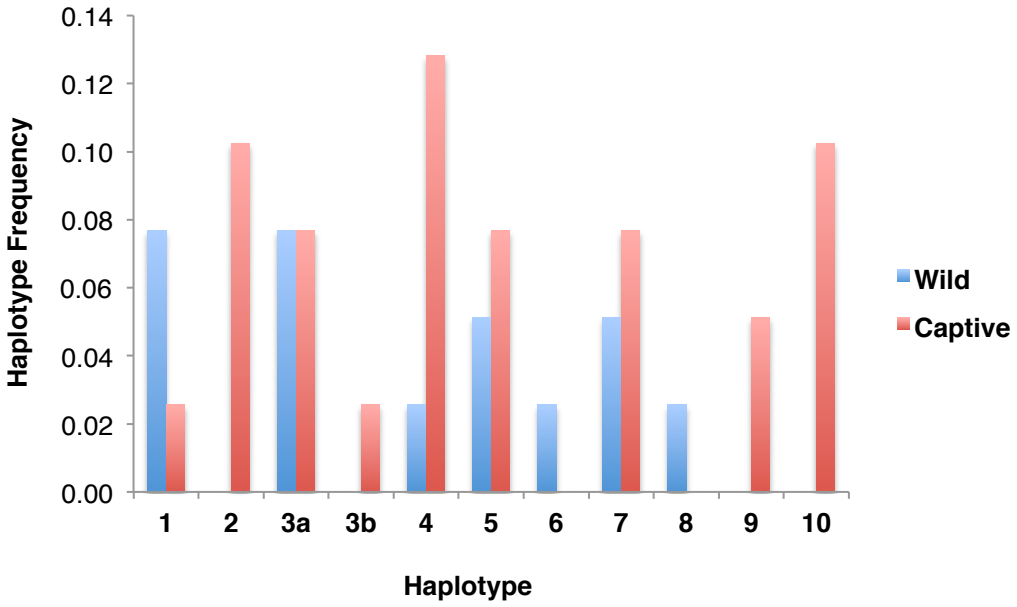


Figure 8 - Haplotype frequencies in the *Hylobates moloch* population

Table 19 illustrates the summary statistics of allelic variation and the number of haplotypes observed within each group. Despite a lower number of individuals (*n*=7) a greater number of alleles are observed with the Wild group. Diversity at the nucleotide sequence level is also greater within the Wild group by both number of polymorphic sites and π. The Captive group harbour a greater number of purported haplotypes with a total of 9.

The results for signs of selection at the MHC *DRB* exon 2 loci are illustrated in Table 20. There is a higher rate of non-synonymous substitutions within alleles found within the Wild group. The *dN>dS Z* tests for positive selection were not significant.

Table 19 - Measure of sequence diversity within *Hylobates moloch* population. *n* is the number of individuals; π is nucleotide diversity.

	n	*No. alleles*	*No. polymorphic sites*	$\pi \pm$ (s.d)	*No. haplotypes*
Wild	7	11	47	0.064 ± 0.006	7
Captive	13	10	36	0.054 ±0.005	9

Table 20 - Rates of synonymous (*dS*) and non-synonymous (*dN*) substitutions (± S.E) in DRB exon 2 for *Hylobates moloch* population, including a positive selection ($d_N > d_S$) one tailed z-test.

	n	d_N	d_S	d_N / d_S	*Z*	*p*
Wild	7	0.061 ± 0.012	0.088 ± 0.025	0.693	-1.080	*NS*
Captive	13	0.049 ± 0.012	0.082 ± 0.021	0.596	-1.409	*NS*

4.4 Discussion

4.4.1 Architecture of *Hylobates moloch* MHC class II DRB exon two, alleles and haplotypes

This study is the first exploration into the genetic make up of the *Hylobates moloch* species within the MHC. Furthermore, the MHC class II DRB exon 2 alleles were amplified and sequenced from non-invasively collected faecal samples. Although this is not the first study to use this source type of DNA (e.g. Lukas et al. 2004; Wan et al. 2006) it is promising that sufficient quantity and quality of DNA was present.

Although there are 8 individuals in the wild-born group, results were only procured from 7 individuals and failure of the target amplicon was as a consequence of the low DNA content extracted for that particular individual. As stated in the materials and methods section, 48 or 96 clones were sequenced per individual. However, not all recombinant clones were successful in each individual and therefore in some cases the number of successfully sequenced amplicons was lower. However, in the case of the lowest number of sequences obtained (7) a total of 5 alleles were detected which exceeded the lowest count of 3 alleles that was evinced in a different individual. Unsuccessful sequencing results were deemed to be those that were markedly shorter in length than comparable sequences and of course those that provided no information at all. Similar low counts of resultant sequences from recombinant clones were observed in a study of the bank vole (*Myodes glareolus*) (Babik & Radwan 2007) yet the overall allele count procured was 15.

Within the *H. moloch* population as a whole, the lowest number of alleles present per individual was three and the greatest number was five. Allele designation for *Hymo* alleles was performed by analysis of levels of similarity of *DRB* exon two together with adjacent D6S2878 microsatellite sequences, with recognized MHC *DRB* alleles. In addition, designation was also carried out by how alleles clustered within the phylogenetic tree with human (*HLA*) and chimpanzee (*Patr*) sequences (as shown in Figure 5) (Doxiadis et al. 2000). The only DRB region that *Hymo* sequences clustered with and showed sufficient sequence similarity to known *HLA* alleles was within *DRB1*. Therefore four alleles belonging to the same group as *HLA-DRB1*04:01:01* were designated as *Hymo-DRB1*04:01* to *Hymo-DRB1*04:04* with the latter number designation indicating that although they cluster with the **04* allele group each individual sequence differs by at least one nucleotide substitution that has resulted in a non-synonymous substitution. There are only two other *Hymo* alleles that share branches with *HLA* sequences. The first is designated *Hymo-DRB*W095:01* that has its own branch linked to two *Patr-DRB6* and one *HLA-DRB6* gene, however the sequence disparity as illustrated by the shortened branch length rendered identification of the *Hymo* locus as unknown. Furthermore, *HLA-* and *Patr-DRB* 6 genes have been identified as pseudogenes as a consequence of both a missing component within exon one that codes for the leader peptide and the presence of stop codons (Doxiadis et al. 2000). By utilising *HLA-DRB1*01*01*01* as a consensus sequence from which to work all *Hymo* alleles were analysed within the Open Reading Frame Finder (NCBI 2014) to search for the presence of stop codons. If the *Hymo* allele does indeed belong to the *DRB6* locus group it was interesting to note that no stop codons were detected. Although it is not possible to state for certain as mRNA analysis was not possible owing to the utilisation of faecal

samples for this study, the absence of stop codons may indicate that this allele is expressed within the *Hylobates moloch* species. Furthermore, the *DRB* associated microsatellite sequence for this allele contained a fourth part (GC) which is absent within the pseudogenes both in the human and chimpanzee sequences. The second allele that forms a separate clade with an *HLA* allele is *Hymo-DRB*W100:01.* Despite sharing this clade, this allele was also classified with a workshop designation, however from its placement it may be deduced that this is also an allele from the *DRB1* locus. All remaining *Hymo* alleles form clades that indicate a separate lineage from the *HLA* and *Patr* sequences as they solely contain *H. moloch* designations.

The overall haplotypic structure for the *moloch* species appears to reflect a structure more akin to that found in humans than that reported for macaques (*Macaca mulatta)* (Doxiadis et al. 2007) and chimpanzees (*Pan troglodytes spp.*) (de Groot et al. 2009). The number of loci per haplotype has a maximum count of three, whereas, the macaques (*M.mulatta*) and chimpanzees (*P.troglodytes spp.*) exhibited haplotypes that contained up to a maximum of 6 loci (de Groot et al. 2009; Doxiadis et al. 2007). A striking difference however between *H. moloch* and human *DRB* exon two architecture is the number of alleles present per locus. For example, the human species has been reported to maintain 542 alleles at the *DRB1* gene alone (Busch, Waser & DeWoody 2008).

4.4.2 MHC Class II *DRB* variability

There is no specific published delineation with regards to *DRB* allele counts within a population or species that categorises them as low or high. Nevertheless, there are reports where allele counts within populations clearly illustrate a depauperate allele count at the *DRB* region (Sommer, Schwab & Ganzhorn 2002). The causes and effects of this occurrence are varied.

In 1992, Slade observed that MHC diversity in southern elephant seals (*Mirounga leonina*) was low. It was hypothesised that the saline aquatic habitat hindered pathogen survival and in response the seals required a limited number of alleles to combat possible infections. Although Slade (1992) referred to low sequence polymorphism within his study, Hoelzel *et al.* (1999) explored allele count and variability at the *DQB* locus in both the same and three other species of seal and the maximum number of observed alleles within a population of 109 individuals totalled just 8 sequences. Low allele counts have also been observed within populations that are ecologically isolated either by physical barriers or as a consequence of their own social and mating systems. For example, the critically endangered Malagasy giant jumping rat (*Hypogeomys antimena*), a monogomous rodent species that lives a solitary lifestyle was found to have just 5 alleles (identified from 22 individuals) at the *DRB* region (Sommer, Schwab & Ganzhorn 2002). As with the prior mentioned marine mammals, reduced exposure to parasites, in this instance as a consequence of reduced lateral transfer of pathogens from their mode of living is one of the reasons postulated to contribute to this low allele count. The endangered Galápagos penguin (*Spheniscus mendiculus*) (Bollmer, Vargas & Parker 2007) and island populations of black-footed rock-wallabys (*Petrogale lateralis lateralis*)

(Mason, Browning & Eldridge 2011) have also been observed to harbour low allele counts at the *DRB* and *DAB* MHC regions respectively. In both studies, the populations in question have a narrow ecological niche and in intra-species comparisons with their conspecifics whose distributions are not limited to the confines of an island, they enjoyed a greater MHC allele count (Bollmer, Vargas & Parker 2007; Mason, Browning & Eldridge 2011). It may be surmised that a contributing factor to a contraction in the number of alleles of island inhabitants mirrors the examples within the marine mammals and the Malagasy rodent, in that a greater scope of alleles at the MHC class II are not required as the suite of pathogens that interact with these species is smaller in comparison to conspecifics within more exposed habitats. Another explanation may also be that gene flow is restricted in such constrained habitats. *Hylobates moloch* is an endemic island species and in the previous chapter regarding microsatellites a reduced level of alleles was observed when compared to mainland *H. lar* species (Chambers et al. 2004). There are no data available to ascertain if this is also the case at the MHC *DRB* region however.

Babik *et al.* (2009) studied two great crested newt populations (*Triturus cristatus*) inhabiting refugial and post-glacial areas. The post-glacial population harboured just two *DAB* alleles, in comparison to the refugial population that maintained 24 alleles. One of the interesting factors of the study, was that the populations with the least number of alleles had proliferated for more than 10,000 years. The grey seal (*Halichoerus grypus*) is reported to harbour a low number of MHC class II alleles at the *DQB* locus (de Assunção-Franco et al. 2012). Within a large breeding population of grey seals residing on the Isle of May in the UK, a total of 5 alleles

were detected in pups and adults. It was found that seals had a greater chance of survival if they had all 5 *DQB* alleles, however the strongest predictor of survivorship was found in one allele in particular. Of 284 dead pups assayed the majority did not possess this allele and thus it was deduced that it plays a vital role in combatting pathogenic infection. Thus, in this example it is the specificity of the allele itself that appears to confer resistance. A similar observation was made in the talas tuco-tuco rodent (*Ctenomys talarum*) where of the 9 alleles identified within 87 individuals, it was as a consequence of a particular group of alleles (referred to as A within the study) that affected susceptibility of individuals to two pathogenic parasites (Cutrera, Zenuto & Lacey 2011).

It seems that there are species that illustrate low variability at the MHC class II, yet do not appear to be affected and then others that show signs of fitness vulnerability as a consequence. There also cases where specific alleles have been shown to be of benefit (de Assunção-Franco et al. 2012). This is particularly pertinent for captive breeding and perhaps even more so for reintroduction programmes. From examples provided, there appears to be habitat specific alleles that function proficiently within their areas, likely as a consequence of the pathogens present in these locations. However, for animals that are reared in captivity and earmarked for release back into the wild, pathogen pressures possibly present within the captive environment may differ to that in the wild. Although some examples such as the great crested newt proliferated over hundreds of generations with just 2 MHC class II alleles an alteration in ecological conditions and therefore possible pathogen presence, within their habitat could result in an increase in mortality. So to reference the results from analyses of the *Hylobates moloch* population as a whole (both

captive and wild), a total of 14 alleles were detected from 20 of the 21 individuals. This count may mean that individuals are equipped to respond to a number of different pathogens whether in captivity or their endemic habitat. However, when the count of alleles is viewed within each the two groups of Captive and Wild *H. moloch,* there are 10 alleles within the Captive group (*n*=13), and 11 alleles within the Wild group (*n*=7). Although it appears to only be a loss of one allele, this lower count within the Captive group was found in almost double the number of individuals than that within the Wild. Further reductions of allele count in future generations born in the captive environment would not be a desired trend. Both Captive and Wild groups harbour rare alleles, and this occurs not only within their respective groupings, but in some cases only occurring within one individual. The wild alleles not found within the Captive group are: *DRB*W100:01*, *DRB*W098:02*, *DRB*W096:02* and *DRB1*04:02*. Captive alleles not found within the wild group are: *DRB*W094:01*, *DRB1*04:03* and *DRB1*04:04*. The frequencies of all alleles within both groups are illustrated in Figure 7. Although it is encouraging that the group as a whole have a moderate allele count, taking into consideration the differential of rare alleles between the Captive and Wild groups would be deemed to be of value when selecting individuals for breeding. As yet, the role of each allele is unknown (i.e the binding capabilities harboured within each allele) and furthermore for *H. moloch* the pathogenic threats within their native habitat are unknown.

DRB alleles are transmitted from generation to generation via haplotypes. The haplotype frequencies within the Captive and Wild groups are illustrated in Figure 8. Two haplotypes are specific to the wild-born group, one of which contains 75% of alleles only found within one individual. The four haplotypes specific to the captive-

born group are composed of alleles present in other haplotypes shared in both groups. This occurrence of identical alleles found at differing loci was observed in greater prairie chickens (*Tympanuchus cupido*) and researchers called the phenomenon “drift- across-loci hypothesis” (Eimes et al. 2011). The overall effect of this occurrence is that copy number variation of alleles within the population is reduced. Haplotypes 9 and 10 contain derivations of the *DRB1* gene only found within captive individuals and haplotype 3b contains *DRB*W094:01* which is also only found within the captive group and furthermore, only found within one individual. Haplotype 3b differs from 3a via this rare allele only via one non-synonymous mutation altering the amino acid from Tryptophan to Valine. Although only found within one individual the variant of this allele was found in multiple recombinant clones derived from two separate PCR reactions and therefore, unlikely to be a PCR artefact.

What was apparent from both the D6S2878 genotyping results and the inferred haplotypes, is the father for H.mol15 that is recorded within zoological records differs to results obtained here. This information was also discovered in the microsatellite pedigree analysis. As haplotypes are inherited on a Mendelian basis, it is expected that a haplotype from the father be present within progeny. However, H.mol15 maintains haplotypes 3b and 4, yet the father has 3a and 5. This may be as a consequence of mutation but when viewed in conjunction with microsatellite results the entry may be incorrect. A further explanation may be owing to contamination of faecal samples, or incorrect labelling of samples at the collection stage. However, if upon further analysis, which is advocated here, a differing father from current records is confirmed then studbook entries should be amended. This is of great

importance, as breeding decisions will be made based on an assumption of relatedness between particular individuals that may be incorrect.

With regards to haplotype counts per group, the Captive group yield a greater overall sum however the differential between the two groups is small (2) when considering the number of individuals assayed (wild *n*=7, captive *n*=13). What is observed is high frequencies of two haplotypes in particular (4 and 10) within the Captive group as a consequence of Mendelian inheritance from related individuals. Whilst common alleles within a population have been reported to be maintained via sexual selection of partners (Bonneaud et al. 2006), which may be effected as a means to combat a common pathogen, or as a strategy to avoid the disruption of co-adapted genes this can not be confirmed here. Caution should be taken by breeding individuals with identical haplotypes if it results in a loss of rare alleles. The mean kinship values are greatest in individuals who harbour common haplotypes. However, it is not entirely fault free as unique haplotypes are present within individuals who have an MK value that is greater than 0.000.

Copy number variation has been detected in some species such as the Tasmanian devil (*Sarcophilus harrisii*) (Siddle et al. 2010) and rhesus macaque (*Macaca mulatta*) (Otting et al. 2005) within the MHC class I region. At the levels observed it was postulated to be a mechanism against pathogenic infections by having varied gene numbers and combinations as the actual nucleotide sequence variation evinced was low. Although there is evidence in *H. moloch* of possible duplication of

loci, with particular alleles appearing in multiple haplotypes, the extent was not as great as that within the aforementioned species.

Within the 14 alleles observed within the *H. moloch* population as a whole, the classification of some *DRB* alleles are indicative of a shared lineage. For example there are four variants of the *DRB1*04* allele group, two of *DRB*094, *096 and *098.* The captive-born group maintain 3 of the 4 *DRB1* gene variants and amino acid substitutions between all four alleles total 9 variations. Similarly both variants of the *Hymo-DRB*W094* allele are found in the captive-born group whereas only one (*Hymo-DRB*W094:02*) is present within the wild-born. Although each allele variant is a different form of immune defence from the Captive group it may be deduced that the *DRB* alleles present within this group are derived from fewer loci than within the Wild group. This is only a hypothesis however, as the loci for the *Hymo* alleles are as yet unidentified.

This nucleotide diversity at the MHC class II *DRB* region is much greater than evinced in the mitochondrial DNA analysis of the *H. moloch* population, which was expected owing to the normally high levels of polymorphism within this genetic region. In isolated populations of brown trout (*Salmo trutta*) nucleotide diversity averaged over all populations assayed was 0.0543 derived from 24 alleles (Campos, Posada & Morán 2006). The nucleotide diversity observed within the brown trout is almost identical to the Captive group of *H. moloch*. This example of a population living in isolation can in some cases be akin to maintaining a population within the captive environment. This correlation is made and is deemed to be applicable to the

H. moloch population as the number of individuals available for captive breeding can be small, in fact it is just 48 individuals for the *moloch* gibbons. Thus in some cases breeding partners are limited in number, rather like isolated populations. The Galápagos penguin (*Sheniscus mendiculus*) exhibited very low variability at the class II *DRB* region maintaining just 3 alleles and perhaps not surprisingly nucleotide diversity was just 0.013. With 14 alleles observed and a greater value of π the *moloch* gibbons assayed here showed a greater genetic diversity. An interesting observation within a species previously shown to harbour a low count of alleles within the MHC, was in the Bengal tiger (*Panthera tigris tigris*) (Pokorny et al. 2010). The study in question is one of the few studies that approach gene variability between captive and wild populations of the same species. They observed that within captive (n=5) and wild (n=11) Bengal tigers only 4 alleles were maintained at exon 2 of the *DRB,* however the diversity of sequences was greater than evinced with *H. moloch* with a value of 0.097 (averaged over both captive and wild groups). As observed with *H. moloch* with there was no significant differentiation between the two groups. This sequence variability may be the species' mechanism for counteracting pathogenic threats and hence why nucleotide diversity appeared to be high with a low allele count. With a similar number of individuals assayed, the *H. moloch* population as a whole may utilise a greater number of alleles rather than sequence polymorphisms to perform the same function.

4.4.3 Tests for selection using d_N and d_S ratios

For both the Wild and Captive *H. moloch* groups rates of non-synonymous substitutions were lower than the synonymous results (wild: d_N 0.061 and d_S 0.088, captive: d_N 0.049 and d_S 0.0817). The synonymous substitutions also referred to as

silent substitutions do not culminate in a change of amino acid and are therefore often referred to as a neutral genetic alteration, although selection has been shown to occur at these sites within mammals (Stoletzki & Eyre-Walker 2011). A test of selection utilising these ratios is the d_N / d_S ratio, which if resulting in a value greater than unity, it is purported to indicate that the genetic region under assay is under positive selection. The d_N / d_S ratio for both groups of *H. moloch* did not exceed 1 and were 0.693 for the Wild and 0.596 for the Captive individuals. As results for both groups are lower than unity, the results may be interpreted that there is some evidence of negative or purifying selection occurring at this genomic region within the *H. moloch* population as a whole. In one sense, this is not entirely unfeasible as the MHC is a very important genomic region with regards to immune response and as Stoletzki and Eyre-Walker (2011) noted, selection to reduce mutation at such important regions may maintain alleles required for optimum fitness or avoid fixation of alleles that may be detrimental for fitness. Although the ratios may indicate purifying selection, it is not deemed that this is a phenomenon occurring within the study population here. The first reason is that the sample group is small and a much wider array of individuals would be required to ascertain this, a factor highlighted by Ellengren (2005). Furthermore, the results were not statistically significant. The d_N / d_S ratio is stated to be used with caution as it was created for inter-species comparisons of genetic regions and not for within-species studies (Kryazhimskiy & Plotkin 2008). The rationale behind this being that comparisons of evolutionary fixation events are more reliable than if utilised with one species as substitutions are deemed to be true fixation events and not transient polymorphisms. However, this ratio is used often for intra-species, for example in the bank vole (*Myodes glareolus*) (Babik & Radwan 2007), great crested newt (*Triturus cristatus*) (Babik et al. 2009),

striped mouse (*Rhabdomys pumilio*) (Froeschke & Sommer 2005), chacma baboon (Huchard et al. 2006), frogs (three families, *Centrolenidae, Hylidae* and *Ranidae*) (Kiemnec-Tyburczy et al. 2012) and black-footed rock wallaby (*Petrogale lateralis lateralis*) (Mason, Browning & Eldridge 2011).

4.5 Conclusions

There is no significant differentiation between the wild and captive-born groups of *H. moloch* with regards to allele and haplotype frequencies. The levels of allelic diversity are lower in the captive-born versus the wild-born groups as demonstrated by a lower count of polymorphic sites and from nucleotide diversity. This is owing to an increased frequency of particular haplotypes within this captive individuals. With the knowledge of which haplotype each individual has, this information can be incorporated into pairing decisions with the aim of maintaining diversity and rare alleles.

There was no evidence that the captive environment has impacted on this genetic region with tests for selection at the MHC *DRB* exon 2 not statistically significant. With a limited number of generations of *H. moloch* in the sample group this is not a surprising outcome, but data provided can serve as a baseline from which to work for future progeny born in captivity. The $d_N > d_S$ z-test yielded negative values within both groups. The value was greater in the Captive group and by looking at the rate of non-synonymous mutations observed within this group, which is markedly lower than in the Wild group it is evident why such a negative value has been obtained as synonymous rates within the groups are very similar.

The information within the MHC work reinforced findings within the microsatellite section in that the sire for h.mol15 may not be h.mol7 as stated in zoological records.

5 Chapter five - Summation of Findings and Recommendations

5.1 Mean kinship versus genetic analysis

Evaluating mean kinship values versus information obtained from genetic analysis it is evident that the proxy measure is both inexpensive and a good basis from which to start selecting individuals for captive breeding. The results for the nuclear marker of microsatellites yielded comparable results which could mean that future work could exclude this step and save both time and money. However, the genetic analysis did highlight a factor that renders data within the mean kinship as incorrect when one of the fathers appeared to be incorrectly stated in the pedigree. This not only affects the individual in question, but has a knock on effect on all MK values thereafter as the measure incorporates how the group as a whole are related to one another. For the MHC analysis, commonality of haplotypes was reflected in individuals with greater MK values which once again shows the usefulness of the MK method. However, it was not without fault and if the values were taken as a means to select individuals for breeding and reintroduction there would be a possibility that rare alleles be lost as they appeared in individuals with intermediate MK scores. The mtDNA results were not in line with prediction of MK values. This particular marker is only inherited maternally and it is perhaps for this reason MK has not worked here particularly well.

5.2 General summation

Evidence within the mtDNA region showed a contraction of variability in captive-born individuals. Although this does not yet reflect in nuclear DNA (as calculated by microsatellites) it is not a trend that a breeding programme would wish to see. But

as Morin *et al.* (2004) stated, a better understanding of fitness within a population may be derived from markers known to be under selection, such as the MHC *DRB* alleles analysed here as it provides information with regards to selective adaptation at genes that are directly involved in fitness. The 14 alleles found within the two groups confer resistance against pathogenic threats and equipped with the knowledge of each individuals haplotypes at this region, detailed decisions can be made with regards to pairing individuals. This would prevent the loss of rare alleles which may provide a valuable role in the wild environment. The results from the three genetic analyses address not only inbreeding and genetic diversity issues that proxy measures provide, but they provide information for most of the genetic problems that may arise as a consequence of the captive environment. The data may also be applied in post-release monitoring purposes by analysis of faecal samples matched to results obtained here.

Despite the fact that the addition of molecular studies to captive breeding and reintroduction programmes provides both a greater array and quality of information it is not routinely carried out and is underrepresented within the literature. For example, a literature review that focused on the genetic aspects of reintroduction programmes found just 15% of over 450 papers referenced the subject between the years of 1990 and 2005 (Seddon, Armstrong & Maloney 2007). A further publication found that from a literature search of *ex-situ* conservation programmes that incorporated genetic analyses between the years 1979 to 2010, only 188 studies had been published (Witzenberger & Hochkirch 2011).

Ultimately, captive breeding and more importantly, reintroduction programmes are two conservation strategies carried out to safeguard endangered species so that they may survive for future generations to come. They are complex, costly and time-consuming. Costs can vary from species to species, from € 14,467 published for the lacertid lizard (*Psammodromus algirus*) the costs of which were stated to cover a captive breeding a subsequent 50 day post-release monitoring programme (Santos et al. 2009). The California condor (*Gymnogyps californianus*) re-introduction programme which by 1993 had been in place for 14 years had up to that point cost in the region of $20,000,000 (Cohn 1993). Despite the apparently low costing in the first example above, expenses from the condor programme are more akin to the average running costs which have been stated to be in the region of $500,000 per annum (Snyder et al. 1996). When such large investments are made, ensuring high survival rates is key. Therefore, although it is acknowledged that budgets will have to include an extra provision, but by implementing genetic analyses as carried out within this study of *H. Moloch* breeding and conservation managers would procure a depth of knowledge of the individuals and indeed the species that surpasses proxy measures of inbreeding, or studbook data.

5.3 Mitochondrial DNA

Further study of the mtDNA molecule can be of value to the wider scientific community as this is a marker of choice for phylogenetic studies and thus be applied to studies of primate evolution and phylogeny. This is pertinent to *Hylobates* as there are differing opinions as to the number of species pertaining to the genus. The HV-1 region of the mtDNA is a highly mutable genetic region and thus provides information as to genetic variability within a differing time scale than within nuclear

DNA. From a laboratory point of view, this molecule is easy to manipulate. Although it provides genetic information from a uniparental perspective, contractions of variability at this molecule may be indicative of inbreeding beyond the mtDNA region, although this was not evinced here at least within the microsatellites. Results from the *moloch* gibbons illustrated that just two haplotypes within individuals born within the captive environment were maintained. The comparison with wild individuals confirmed that this is not as a consequence of a life history trait but rather from decisions taken within the captive breeding process. This was also evinced in the results showing that the captive-born and wild-born groups had undergone a significant level of population differentiation at this genetic region, despite the fact that they are managed as one entire population. Thus, females may be chosen for future generations who do not maintain these two haplotypes and improve levels of genetic variability at this region. Although the captive breeding programme within tigers (Luo et al. 2008) has been carried out with the aims of achieving greater variability at this region, it is not advocated here that this be carried out to the detriment of losing variability at other genomic regions, the breeding for particular alleles as discussed by Lacy (2000). Information from both microsatellites and the MHC *DRB* region should also be considered as below.

5.4 Microsatellites, the second neutral marker

The differentiation between the wild and captive-born groups at the mtDNA level did not appear to have occurred with genomic DNA. This was deduced from the second neutral marker analysed of microsatellites that were taken from 11 chromosomes of the 22 present with the *H. moloch* species. Using the Mendelian inherited microsatellites, the desired levels of heterozygosity that captive breeding

programmes aim to conserve can physically be checked. Standardized heterozygosity was fairly high in both wild and captive-born groups. The molecular information gathered here also yielded a vital piece of information with regards to the species studbook, as results indicated that two fathers may be incorrectly assigned in zoological records.

5.5 The major histocompatibility complex class II DRB region

The MHC class II *DRB* region is extremely important for adaptive immune responses within individuals (de Groot et al. 2009; Mason, Browning & Eldridge 2011). The second exon analysed here contains the peptide binding region for extra-cellular pathogenic binding (Doxiadis et al. 2007). Therefore, the information derived from this genetic region was not only beneficial from a variability viewpoint, with high levels desired, but also owing to its direct links to fitness. The alleles present within the wild-born individuals can now be monitored so as to ensure that they are promulgated in future generations and this is most important for individuals who will take part in the re-introduction programme. As seen in mtDNA, high frequencies of particular haplotypes were also seen in the MHC results. If this were to continue there is the possibility that rarer alleles be lost within the captive population, that may be of great importance within the wild habitat.

5.6 Recommendations

Owing to the large amount of investment required to orchestrate and run a captive breeding and reintroduction programme, it is surprising that genetic aspects are not incorporated on a more frequent basis owing to the fact that it is known that

inbreeding and losses of diversity have the potential to negatively impact on success rates. Thus, it is recommended here that molecular analysis of individuals within captive breeding programmes if incorporating release strategies to the wild, be included on a routine basis so as to maximise fitness levels for those in the programme and to wild conspecifics. For the *moloch* groups analysed here, a continuation of the strategy employed to include all individuals in the programme to provide information at the individual level but also to monitor deviations in genomic DNA from wild-born to captive-born individuals observed here. Although data was insufficient to draw conclusions with regards to what type of mate choice mechanism *H. moloch* employs, further work is advocated to be carried out as this has the benefit to pair individuals with less aggression or stress and also be more akin to the decision making that would occur in their natural habitat.

6 Bibliography

Adams, MS & Villablanca, FX 2007, 'Consequences of a genetic bottleneck in California condors: A mitochondrial DNA perspective', in A Mee, LS Hall (eds.), *Series in Ornithology: Condors in the 21st Century*, University of Califnornia Press, California.

Ahlering, MA, Hedges, S, Johnson, A, Tyson, M, Schuttler, SG & Eggert, LS 2011, 'Genetic diversity, social structure, and conservation value of the elephants of the Nakai Plateau, Lao PDR, based on non-invasive sampling', *Conservation Genetics*, **vol 12**, pp. 413-422.

Alcaide, M, Negro, JJ, Serrano, D, Antolin, JL, Casado, S & Pomarol, M 2010, 'Captive breeding and reintroduction of the lesser kestrel Falco naumanni A genetic analysis using microsatellites', *Conservation Genetics*, **vol 11**, pp. 331-338.

Alport, LJ 2004, 'Comparative analysis of the role of olfaction and the neocortex in primate intrasexual competition', *The Anatomical Record Part A*, **vol 28**, pp. 1182-1189.

Amos, W & Harwood, J 1998, 'Factors affecting levels of genetic diversity in natural populations', *Philosophical Transactions of the Royal Society of London B*, **vol 353**, pp. 177-186.

Amos, W, Worthington Wilmer, J, Fullard, K, Burg, TM, Croxall, JP, Bloch, D & Coulson, T 2001, 'The influence of parental relatedness on reproductive success', *Proceedings of the Royal Society of London B*, **vol 268**, pp. 2021-2027.

Andayani, N, Morales, JC, Forstner, MJ, Supriatna, J & Melnick, DJ 2001, 'Genetic variability in mtDNA of the silvery gibbon: Implications for the conservation of a critically endangered speices', *Conservation Biology*, **vol 15**, pp. 770-775.

Andrabi, SMH & Maxwell, WMC 2007, 'A review on reproductive biotechnologies for conservation of endangered mammalian species', *Animal Reproduction Science*, **vol 99**, pp. 223-243.

Arandjelovic, M, Head, J, Kühl, H, Boesch, C, Robbins, MM, Maisels, F & Vigilant, L 2010, 'Effective non-invasive monitoring of mutiple wild western gorilla groups', *Biological Conservation*, **vol 143**, pp. 1780-1791.

Asa, CS, Traylor-Holzer, K & Lacy, RC 2011, 'Can conservation-breeding programmes be improved by incorporating mate choice?', *International Zoo Yearbook*, **vol 45**, pp. 203-212.

Asa, CS, Traylor-Holzer, K & Lacy, RC 2011, 'Mate choice as a potential tool to increase population sustainability', *WAZA Magazine*, **vol 12**, pp. 23-25.

Aspinall Foundation 2010, *Indonesia-Saving the Endangered Javan Gibbon*, viewed January 2010,

"http://www.aspinallfoundation.org/aspinall/conservation-work/view/164/indonesia"

Babik, W 2010, 'Methods for MHC genotyping in non-model vertebrates', *Molecular Ecology Resources*, **vol 10**, pp. 237-251.

Babik, W, Pabijan, M, Arntzen, JW, Cogălniceanu, D, Durka, W & Radwan, J 2009, 'Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation', *Molecular Ecology*, **vol 18**, pp. 769-781.

Babik, W & Radwan, J 2007, 'Sequence diversity of MHC class II DRB genes in the bank vole *Myodes glareolus'*, *Acta Theriologica*, **vol 52**, pp. 227-235.

Balding, DJ 2006, 'A tutorial on statistical methods for population association studies', *Nature Reviews Genetics*, **vol 7**, pp. 781-791.

Ballou, JD & Lacy, RC 1995, 'Identifying genetically important individuals for management of genetic variation in pedigreed populations', in JD Ballou, M Gilpin, TJ Foose (eds.), *Population Management for Survival and Recovery. Analytical Methods and Strategies in Small Population Conservation*, Columbia University Press, New York.

Ballou, JD, Lees, C, Faust, LJ, Long, S, Lynch, C, Lackey, LB & Foose, TJ 2010, 'Demographic and genetic management of captive populations', in DG Kleiman, KV Thompson, CK Baer (eds.), *Wild Mammals in Captivity. Principles and Techniques for Zoo Management*, 2nd edn, University of Chicago Press Ltd., London.

Bartlett, MS 1937, 'Properties of sufficiency and statistical tests', *Proceedings of the Royal Society London A*, **vol 160**, pp. 268-282.

Bean, K, Amos, W, Pomeroy, PP, Twiss, SD, Coulson, TN & Boyd, IL 2004, 'Patterns of parental relatedness and pup survival in the grey seal (*Halichoerus grypus*)', *Moleular Ecology*, **vol 13**, pp. 2365-2370.

Beauchamp, GK & Yamazaki, K 1997, 'HLA and mate selection in humans: Commentary', *American Journal of Human Genetics*, **vol 61**, pp. 494-496.

Beauchamp, GK & Yamazaki, K 2003, 'Chemical signalling in mice', *Biochemical Society Transactions*, **vol 31**, pp. 147-151.

Beauclerc, KB, Johnson, B & White, BN 2010, 'Genetic rescue of an inbred captive population of the critically endangered Puerto Rican crested toad (*Peltophryne lemur*)', *Conservation Genetics*, **vol 11**, pp. 21-32.

Belay, G & Mori, A 2006, 'Intraspecific phylogeographic mitochondrial DNA (D-loop) variation of Gelada baboon, *Theropithecus gelada*, in Ethiopia', *Biochemical Systematics and Ecology*, **vol 34**, pp. 554-561.

Bernatchez, L & Landry, C 2003, 'MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years?', *Journal of Evolutionary Biology*, **vol 16**, pp. 363-377.

Birky, Jr, CW 1995, 'Biparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution', *Proceedings of the National Academy of Sciences, USA*, **vol 92**, pp. 11331-11338.

Bollmer, JL, Vargas, FH & Parker, PG 2007, 'Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*)', *Immunogenetics*, **vol 59**, pp. 593-602.

Bonneaud, C, Chastel, O, Federici, P, Westerdahl, H & Sorci, G 2006, 'Complex MHC-based mate choice in a wild passerine', *Proceedings of the Royal Society B*, **vol 273**, pp. 1111-1116.

Bradley, BJ, Boesch, C & Vigilant, L 2000, 'Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces', *Conservation Genetics*, **vol 1**, pp. 289-292.

Bradshaw, WE & Holzapfel, CM 2008, 'Genetic response to rapid climate change: Its seasonal timing that matters', *Molecular Ecology*, **vol 17**, pp. 157-166.

Britt, A, Welch, C & Katz, A 2003, 'Can small, isolated primate poulations be effectively reinforced through the release of individuals from a captive population?', *Biological Conservation*, **vol 115**, pp. 319-327.

Brockelman, WY, Reichard, U, Treesucon, U & Raemakers, JJ 1998, 'Dispersal, pair formation and social structure in gibbons (*Hylobates lar*)', *Behavioural Ecology and Sociobiology*, **vol 42**, pp. 329-339.

Brookfield, JFY 1996, 'A simple new method for estimating null allele frequency from heterozygote deficiency', *Molecular Ecology*, **vol 5**, pp. 453-455.

Brooks, TM, Mittermeier, RA, Fonseca, GAB, Gerlach, J, Hoffman, M, Lamoreaux, JF, Mittermeier, CG, Pilgrim, JD & Rodrigues, ASL 2006, 'Global biodiveristy conservation priorities', *Science*, **vol 313**, pp. 58-61.

Burek, KA, Gulland, FMD & O'Hara, TM 2008, 'Effects of climate change on Arctic marine mammal health', *Ecological Applications*, **vol 18**, pp. 126-134.

Buschiazzo, E & Gemmell, NJ 2010, 'Conservation of human microsatellites across 450 million years of evolution', *Genome Biology and Evolution*, **vol 2,** pp. 153-165.

Busch, JD, Waser, PM & DeWoody, JA 2008, 'Characterization of expressed class II MHC sequences in the banner-tailed kangaroo rat (*Dipodomys spectabilis*) reveals multiple DRB loci', *Immunogenetics*, **vol 60**, pp. 677-688.

Bushar, LM, Maliga, M & Reinart, HK 2001, 'Cross-species amplification of Crotalus horridus microsatellites and their application in phylogenetic analysis', *Journal of Herpetology*, **vol 35**, pp. 532-537.

Caballero, S, Santos, MCDO, Sanches, A & Mignucci-Giannoni, AA 2013, 'Initial description of the phylogeography, population structure and genetic diversity of Atlantic spotted dolphins from Brazil and the Caribbean, inferred from analyses of mitochondrial and nuclear DNA', *Biochemical Systematics and Ecology*, **vol 48**, pp. 263-270.

Campos, JL, Posada, D & Morán , P 2006, 'Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of brown trout (*Salmo trutta*)', *Conservation Genetics*, **vol 7,** pp. 515-530.

Carbone, L, Harris, RA, Gnerre, S, Veeramah, KR, Lorente-Galdos, B, Huddleston, H, Meyer, TJ, Herrero, J, Roos, C, Aken, B, Anaclerio, F, Archidiacono, N, Baker, C, Barrell, D, Batzer, MA, Beal, K, Blancher, A, Bohrson, CL, Brameier, M, Campbell, MS, et al. 2014, 'Gibbon genome and the fast karyotype evolution of small apes', *Nature*, **vol 513**, pp. 195-201.

Chakraborty, R, De Andrade, M, Daiger, SP & Budowle, B 1992, 'Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications', *Annals of Human Genetics*, **vol 56**, pp. 45-57.

Chambers, GK & MacAvoy, ES 2000, 'Microsatellites: Consensus and controversy', *Comparative Biochemistry and Physiology, Part B*, **vol 126**, pp. 455-476.

Chambers, KE, Reichard, UH, Möller, A, Nowak, K & Vigilant, L 2004, 'Cross-species amplification of human microsatellite markers using noninvasive samples from white-handed gibbons (*Hylobates lar*)', *American Journal of Primatology*, **vol 64**, pp. 19-27.

Chan, Y-C, Roos, C, Inoue-Murayama, M, Inoue, E, Shih, C-C, Pei, KJ-C & Vigilant, L 2010, 'Mitochondrial genome sequences effectively reveal the phylogeny of Hylobates gibbons', *PLoS ONE*, **vol 5**, pp. 1-9.

Chapin III, FS, Zavaleta, ES, Eviner, VT, Naylor, RL, Vitousek, PM, Reynolds, HL, Hooper, DU, Lavorel, S, Sala, OE, Hobbie, SE, Mack, MC & Diaz, S 2000, 'Consequences of changing biodiversity', *Nature*, **vol 405**, pp. 234-242.

Charpentier, MJE, Boulet, M & Drea, CM 2008, 'Smelling right: The scent of male lemurs advertises genetic quality and relatedness', *Molecular Ecology*, **vol 17**, pp. 3225-3233.

Chaves, PB, Paes, MF, Mendes, SL, Strier, KB, Louro, ID & Fagundes, V 2006, 'Noninvasive genetic sampling of endangered muriqui (Primates, *Atelidae*): Efficiency of fecal DNA extraction', *Genetics and Molecular Biology*, **vol 29**, pp. 750-754.

Chen, XJ & Butow, RA 2005, 'The organization and inheritance of the mitochondrial genome', *Nature Reviews Genetics*, **vol 6**, pp. 815-825.

Chow, S, Suzuki, N, Brodeur, RD & Ueno, Y 2009, 'Little population structuring and recent evolution of he Pacific saury (*Cololabis saira*) as indicated by mitochondrial and nuclear DNA sequence data', *Journal of Experimental Marine Biology and Ecology*, **vol 369**, pp. 17-21.

Christie, MR, Marine, ML & Blouin, MS 2012, 'Genetic adaptation to captivitiy can occur in a single generation', *Proceedings of the National Academy of Sciences*, **vol 109**, pp. 238-242.

Čížková, D, Javůrková , V, Champagnon, J & Keisinger, J 2012, 'Duck's not dead: Does restocking with captive bred individuals affect the genetic integrity of wild mallard (*Anas platyrhynchos*) population?', *Biological Conservation*, **vol 152**, pp. 231-240.

Cohn, JP 1993, 'The flight of the Califnornian condor', *BioScience*, **vol 43**, pp. 206-209.

Coltman, DW, Bowen, WD & Wright, JM 1998, 'Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites', *The Proceedings of the Royal Society of London B*, **vol 265**, pp. 803-809.

Coltman, DW, Pilkington, JG, Smith, JA & Pemberton, JM 1999, 'Parasite-mediated selection against inbred Soay sheep in a free-living, island population', *Evolution*, **vol 53**, pp. 1259-1267.

Cook, LM, Sutton, SL & Crawford, TJ 2005, 'Melanic moth frequencies in Yorkshire, and old English industrial hot spot', *Journal of Heredity*, **vol 96**, pp. 522-528.

Coote, T & Bruford, MW 1996, 'Human microsatellites applicable for analysis of genetic variation in Apes and Old World Monkeys', *The Journal of Heredity*, **vol 87**, pp. 406-410.

Coughlan, J, Mirimin, L, Dillane, E, Rogan, E & Cross, TF 2006, 'Isolation and characterization of novel microsatellite loci for the short-beaked common dolphin (*Delphinus delphis*) and cross-amplification in other cetacean species', *Molecular Ecology Notes*, **vol 6**, pp. 490-492.

Coulson, T, Albon, S, Slate, J & Pemberton, J 1999, 'Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in red deer calves', *Evolution*, **vol 53**, pp. 1951-1960.

Coulson, TN, Pemberton, JM, Albon, SD, Beaumont, M, Marshall, TC, Slate, J, Guinness, FE & Clutton-Brock, TH 1998, 'Microsatellites reveal heterosis in red deer', *Proceedings of the Royal Society of London B*, **vol 265,** pp. 489-495.

Crandall, KA, Bininda-Emonds, ORP, Mace, GM & Wayne, RK 2000, 'Considering evolutionary processes in conservation biology', *Trends in Ecology and Evolution*, **vol 15**, pp. 290-295.

Crouau-Roy, B & Clisson, B 2000, 'Evolution of an Alu DNA element of type Sx in the lineage of primates and the origin of an associated tetranucleotide microsatellite', *Genome*, **vol 43**, pp. 642-648.

Crow, JF 1988, 'Eighty years ago: The beginnings of population genetics', *Genetics*, vol 119, pp. 473-476.

Crow, JF & Kimura, M 1970, *An Introduction to Population Genetics Theory*, Harper & Row, New York, USA.

Curik, I, Zechner, P, Solkner, J, Achman, R, Bodo, I, Dovc, P, Kavar, T, Marti, E & Brem, G 2003, 'Inbreeding, microsatellite heterozygosity, and morphological traits in Lipizzan horses', *Journal of Heredity*, **vol 94**, pp. 125-132.

Cutrera, AP, Zenuto, RR & Lacey, EA 2011, 'MHC variation, multiple simultaneous infections and physiological condition in the subterranean rodent *Ctenomys talarum*', *Infection, Genetics and Evolution*, **vol 11**, pp. 1023-1036.

Dallman, R & Geissman, T 2009, 'Individual and geographical variability in the songs of wild silvery gibbons (*Hylobates moloch*) on Java, Indonesia', in S Lappan, DJ Whittaker (eds.), *The Gibbons: New Perspectives on Small Ape Socioecology and Population Biology*, Springer Science, New York, USA.

de Assunção-Franco, M, Hoffman, JI, Harwood, J & Amos, W 2012, 'MHC genotype and near-deterministic mortality in grey seals', *Scientific Reports*, **vol 2**, pp. 1-3.

de Groot, N, Doxiadis, GGM, de Vos-Rouweler, AJM, de Groot, NG, Verschoor, EJ & Bontrop, RE 2008, 'Comparative genetics of a highly divergent DRB microsatellite in different macaque species', *Immunogenetics*, vol 60, pp. 737-748.

de Groot, N, Heijmans, CMC, de Groot, N, Doxiadis, GGM, Otting, N & Bontrop, RE 2009, 'The chimpanzee Mhc-DRB region revisited: Gene content, polymorphism, pseudogenes and transcripts', *Molecular Immunology*, **vol 47,** pp. 381-389.

de Groot, NG, Heijmans, CMC, de Groot, N, Otting, N, de Vos-Rouweler, AJM, Remarque, EJ, Bonhomme, M, Doxiadis, GM, Crouau-Roy, B & Bontrop, RE 2008, 'Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics', *Molecular Ecology*, **vol 17**, pp. 2074-2088.

de Jong, MA, Wahlberg, N, van Eijk, M, Brakefield, PM & Zwaan, BJ 2011, 'Mitochondrial DNA signature for range-wide populations of *Bicyclus anynana* suggests a rapid expansion from recent refugia', *Plos ONE*, **vol 6**, pp. 1-5.

Di Fonzo, MMI, Pelletier, F, Clutton-Brock, TH, Pemberton, JM & Coulson, T 2011, 'The population growth consequences of variation in individual heterozygosity', *PloS ONE*, **vol 6**, pp. 1-8.

Di Rienzo, A, Peterson, AC, Garza, JC, Valdes, AM, Slatkin, M & Freimer, NB 1994, 'Mutational processes of simple-sequence repeat loci in human populations', *Procedings of the National Academy of Sciences USA*, **vol 91**, pp. 3166-3170.

Ditlecadet, D, Dufresne, F, Le François , NR & Blier, PU 2006, 'Applying microsatellites in two commercial strains of Artic charr (*Salvelinus alpinus*): Potential for a selective breeding program', *Aquaculture*, **vol 257**, pp. 37-43.

Doxiadis, GGM, de Groot, N, Claas, FHJ, Doxiadis, IIN, van Rood, JJ & Bontrop, RE 2007, 'A higly divergent microsatellite facilitating fast and accurate DRB haplotyping in humans and rhesus macaques', *Proceedings of the National Academy of Sciences*, **vol 104**, pp. 8907-8912.

Doxiadis, GGM, de Groot, N, de Groot, NG, Doxiadis, IIN & Bontrop, RE 2008, 'Reshuffling of ancient peptide binding motifs between HLA-DRB multigene family members: Old wine served in new skins', *Molecular Immunology*, **vol 45**, pp. 2743-2751.

Doxiadis, GGM, de Groot, N, de Groot, NG, Rotmans, G, de Vos-Rouweler, AJM & Bontrop, RE 2010, 'Extensive DRB region diversity in cynomolgus macaques: Recombination as a driving force', *Immunogenetics*, **vol 62**, pp. 137-147.

Doxiadis, GGM, Otting, N, de Groot, NG, Noort, R & Bontrop, RE 2000, 'Unprecedented polymorphism of Mhc-DRB region configurations in rhesus macaques', *Journal of Immunology*, **vol 164**, pp. 3193-3199.

Doxiadis, GGM, Rouweler, AJM, de Groot, NG, Louwerse, A, Otting, N, Verschoor, EJ & Bontrop, RE 2006, 'Extensive sharing of MHC class II alleles between rhesus and cynomolgus macaques', *Immunogenetics*, **vol 58,** pp. 259-268.

Drury, JP 2010, 'Immunity and mate choice: A new outlook', *Animal Behaviour*, **vol 79**, pp. 539-545.

Eckert, I, Suchentrunk, F, Markov, G & Hartl, GB 2010, 'Genetic diversity and integrity of German wildcat (*Felis silvestris*) populations as revealed by microsatellites, allozymes and mitochondrial DNA sequences', *Mammalian Biology*, **vol 75**, pp. 160-174.

Eimes, JA, Bollmer, JL, Whittingham, LA, Johnson, JA, van Oosterhout, C & Dunn, PO 2011, 'Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and loss of copy number variation', *Journal of Evolutionary Biology*, **vol 24**, pp. 1847-1856.

Eizaguirre, C, Yeates, SE, Lenz, TL, Kalbe, M & Milinski, M 2009, 'MHC-based mate choice combines good genes and maintenance of MHC polymorphism', *Molecular Ecology*, **vol 18**, pp. 3316-3329.

Ejsmond, MJ & Radwan, J 2011, 'MHC diversity in bottlenecked populations: A simulation model', *Conservation Genetics*, **vol 12,** pp. 129-137.

Ekblom, R, Sæther , SA, Fiske, P, Kålås, JA & Höglund, J 2010, 'Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe', *Genetica*, **vol 138**, pp. 453-461.

Ellengren, H 2005, 'Evolution: Natural selection in the evolution of humans and chimps', *Current Biology*, **vol 15**, pp. R919-R922.

Ellis, SA, Bontrop, RE, Antczak, DF, Ballingall, K, Davies, CJ, Kaufman, J, Kennedy, LJ, Robinson, J, Smith, DM, Stear, MJ, Stet, RJM, Waller, MJ, Walter, L & Marsh, SGE 2006, 'ISAG/IUIS-VIC comparative MHC nomenclature committee report, 2005', *Immunogenetics*, **vol 57**, pp. 953-958.

Ernest, HB, Penedo, MCT, May, BP, Syvanen, M & Boyce, WM 2000, 'Molecular tracking of mountain lions in hte Yosemite Valley region in California: Genetic analysis using microsatellites and faecal DNA', *Molecular Ecology*, **vol 9**, pp. 433-441.

Estoup, A, Jarne, P & Cornuet, J-M 2002, 'Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis', *Molecular Ecology*, **vol 11**, pp. 1591-1604.

Excoffier, L & Heckel, G 2006, 'Computer programs for population genetics data analysis: A survival guide', *Nature Reviews Genetics*, **vol 7**, pp. 745-758.

Excoffier, L & Lischer, HEL 2010, 'Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows', *Molecular Ecology Resources*, **vol 10**, pp. 564-567.

Fagotti, A & Pascolini, R 2007, 'The proximate cause of frog declines?', *Nature*, **vol 447**, pp. E4-E5.

Falconer, DS 1981, *Introduction to Quantitative Genetics*, Longman, New York.

Felsenstein, J 1981, 'Evolutionary trees from DNA sequences: A maximum likeliehood approach', *Journal of Molecular Evolution*, **vol 17**, pp. 368-376.

Forstmeier, W, Schielzeth, H, Mueller, JC, Ellegren, H & Kempenaers, B 2012, 'Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation', *Molecular Ecology*, **vol 21**, pp. 3237-3249.

Francisco, FDO, Santiago, LR & Arias, MC 2013, 'Molecular genetic diversity in populations of the stingless bee *Plebeia remota*: A case study', *Genetics and Molecular Biology*, **vol 36**, pp. 118-123.

Frankham, R 1997, 'Do island populations have less genetic variation than mainland populations?', *Heredity*, **vol 78**, pp. 311-327.

Frankham, R 2005, 'Genetics and extinction', *Biological Conservation*, **vol 126**, pp. 131-140.

Frankham, R 2008, 'Genetic adaptation for captivity in species conservation programs', *Molecular Ecology*, **vol 17**, pp. 325-333.

Frankham, R 2010, 'Challenges and opportunities of genetic approaches to biological conservation', *Biological Conservation*, **vol 143**, pp. 1919-1927.

Frankham, R, Ballou, JD & Briscoe, DA 2010, *Introduction to Conservation Genetics*, Second Edition edn, Cambridge University Press, New York, USA.

Frankham, R, Ballou, JD, Eldridge, MDB, Lacy, RC, Ralls, K, Dudash, MR & Fenster, CB 2010b, 'Predicting the probability of outbreeding depression', *Conservation Biology*, **vol 25**, pp. 465-475.

Freeland, JR 2005, *Molecular Ecology*, John Wiley & Sons Ltd., Chichester, England.

Froeschke, 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*, **vol 22**, pp. 1254-1259.

Fu, YX 1997, 'Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection', *Genetics*, **vol 147**, pp. 915-925.

Galindo, CL, McIver, LJ, McCormick, JF, Skinner, MA, Xie, Y, Gelhausen, RA, Ng, K, Kumar, NM & Garner, HR 2009, 'Global microsatellite contect distinguises humans, primates, animals and plants', *Molecular Biology and Evolution*, **vol 26**, pp. 2809-2819.

Gang, H, Kang, H, SongTao, G, WeiHong, J, XiaoGuang, Q, Yi, R, XueLin, J & BaoGuo, L 2011, 'Evaluating the reliability of microsatellite genotyping from low-quality DNA templates with a polynomial distribution model', *Chinese Science Bulletin*, **vol 56**, pp. 2523-2530.

Gautschi, B, Widmer, A, Joshi, J & Koella, JC 2002, 'Increased frequency of scale anomalies and loss of genetic variation in serially bottlenecked populations of the dice snake, Natrix tessellata', *Conservation Genetics*, **vol 3**, pp. 235-245.

Geissman, T & Nijman, V 2006, 'Calling in wild silvery gibbons (*Hylobates moloch*) in Java (Indonesia): Behaviour, phylogeny and conservation', *American Journal of Primatology*, **vol 68**, pp. 1-19.

Gerber, S, Mariette, S, Streiff, R, Bodénès , C & Kremer, A 2000, 'Comparison of microsatellites and amplified fragment length polymorphisms markers for parentage analysis', *Molecular Ecology*, **vol 9**, pp. 1037-1048.

Glenn, TC, Staton, JL, Vu, AT, Davis, LM, Bremer, JRA, Rhodes, WE, Brisbin, Jr, IL & Sawyer, RH 2002, 'Low mitochondrial DNA variation among American alligators and a novel non-coding region in Crocodilians', *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **vol 294**, pp. 312-324.

Goodman, SJ 1998, 'Patterns of extensive genetic differentiation and variation among European harbour seals (*Phoca vitulina vitulina*) revealed using microsatellite DNA polymorphisms', *Molecular Biology and Evolution*, **vol 15**, pp. 104-118.

Goosens, B, Latour, S, Vidal, C, Jamart, A, Ancrenaz, M & Bruford, MW 2000, 'Twenty new microsatellite loci for use with hair and faecal samples in the chimpanzee (*Pan troglodytes troglodytes*)', *Folia Primatologica*, **vol 71**, pp. 177-180.

Goudet, J 2001, *FSTAT a program to estimate and test gene diversities and fixation indices (version 2.9.3)*, viewed January 2014, "http://www.unil.ch/izea/softwares/fstat.html."

Grant, BR & Grant, PR 2003, 'What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity', *BioScience*, **vol 53**, pp. 965-975.

Griekspoor, A & Groothius, T 2006, *4Peaks*, viewed 25 March 2014, <http://nucleobytes.com/index.php/4peaks

Griffith, SC, Owens, IPF & Thuman, KA 2002, 'Extra pair paternity in birds: A review of interspecific variation and adaptive function', *Molecular Ecology*, **vol 11**, pp. 2195-2212.

Guichoux, E, Lagache, L, Wagner, S, Chaumeil, P, Léger, P, Lepais, O, Lepoittevin, C, Malausa, T, Revardel, E, Salin, F & Petit, RJ 2011, 'Current trends in microsatellite genotyping', *Molecular Ecology Resources*, **vol 11**, pp. 591-611.

Guo, SW & Thompson, EA 1992, 'Performing the exact test of Hardy-Weinberg proportion for multiple alleles', *Biometrics*, **vol 48**, pp. 361-372.

Hübner, K, Gonzalez-Wanguemert, M, Diekmann, OE & Serrão , EA 2013, 'Genetic evidence for polygynandry in the black-striped pipefish *Syngnathus abaster*: A microsatellite-based parentage analysis', *Journal of Heredity*, **vol 104**, pp. 791-797.

Haasl, RJ & Payseur, BA 2011, 'Multi-locus inference of population structure: A comparison between single nucleotide polymorphisms and microsatellites', *Heredity*, **vol 106**, pp. 158-171.

Hansson, B, Bensch, S, Hasselquist, D, Lillandt, B-G, Wennerberg, L & von Schantz, T 2000, 'Increase of genetic variation over time in a recently founded poulation of great reed warblers (*Acrocephalus arundinaceus*) revealed by microsatellites and DNA fingerprinting', *Molecular Ecology*, **vol 9**, pp. 1529-1538.

Harrison, HB, Saenz-Agudelo, P, Planes, S, Jones, GP & Berumen, ML 2012, 'Relative accuracy of three common methods of parentage analysis in natural populations', *Molecular Ecology*, **vol 22**, pp. 1158-1170.

Hasegawa, M, Kishino, H & Yano, T 1985, 'Dating of the human-ape splitting by a molecular clock of mitochondrial DNA', *Journal of Molecular Evolution*, **vol 22**, pp. 160-174.

Hassanin, A, Ropiquet, A, Gourmand, A-L, Chardonnet, B & Rigoulet, J 2007, 'Mitochondrial DNA variability in *Giraffa camelopardalis*: Consequences for taxonomy, phylogeography and conservation of giraffes in West and central Africa', *Comptes Rendues Biologies*, **vol 330**, pp. 265-274.

Hazkani-Covo, E, Zeller, RM & Martin, W 2010, 'Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes', *PloSGenetics*, **vol 6**, p. e1000834.

Hedrick, PW 2005, *Genetics of Populations*, Third Edition, Jones and Bartlett Publishers, Massachusetts, USA.

Hedrick, PW, Miller, PS, Geffen, E & Wayne, R 1997, 'Genetic evaluation of the three captive Mexican wolf lineages', *Zoo Biology*, **vol 16**, pp. 47-69.

Hill, AVS 1991, 'HLA associations with malaria in Africa: Some implications for MHC evolution', in J Klein, D Klein (eds.), *Molecular Evolution of the Major Histocompatibility Complex*, Springer, Berlin.

Hoelzel, AR, Stephens, JC & O'Brien, SJ 1999, 'Molecular genetic diversity and evolution at the MHC DQB locus in four species of Pinnipeds', *Molecular Biology and Evolution*, **vol 16**, pp. 611-618.

Hogg, JT, Forbes, SH, Steele, BM & Luikart, G 2006, 'Genetic rescue of an insular population of large mammals', *Proceedings of the Royal Society of London B*, **vol 273**, pp. 1491-1499.

Horth, L 2007, 'Sensory genes and mate choice: Evidence that duplications, mutations and adaptive evolution alter variation in mating cues and their receptors', *Genomics*, **vol 90**, pp. 159-175.

Horton, R, Gibson, R, Coggill, P, Miretti, M, Allcock, RJ, Almeida, J, Forbes, S, Gilbert, JGR, Halls, K, Harrow, JL, Hart, E, Howe, K, Jackson, DK, Palmer, S, Roberts, AN, Sims, S, Stewart, CA, Traherne, JA, Trevanion, S, Wilming, L, et al. 2008, 'Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project', *Immunogenetics*, **vol 60**, pp. 1-18.

Hosey, GR 2005, 'How does the zoo environment affect the behaviour of captive primates', *Applied Animal Behaviour Science*, **vol 90**, pp. 107-129.

Houlden, BA, Costello, BH, Sharkey, D, Fowler, EV, Melzer, A, Ellis, W, Carrick, F, Baverstock, PR & Elphinstone, MS 1999, 'Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817)', *Molecular Ecology*, **vol 8**, pp. 999-1011.

Huchard, E, Cowlishaw, G, Raymond, M, Weill, M & Knapp, LA 2006, 'Molecular study of Mhc-DRB in wild chacma baboons reveals high variability and evidence for trans-species inheritance', *Immunogenetics*, **vol 58**, pp. 805-816.

Huchard, E, Knapp, LA, Wang, J, Raymond, M & Cowlishaw, G 2010, 'MHC, mate choice and heterozygote advantage in a wild social primate', *Molecular Ecology*, **vol 19**, pp. 2545-2561.

Hughes, A & Nei, M 1988, 'Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection', *Nature*, **vol 355**, pp. 167-170.

Hu, J, Pan, H-J, Wan, Q-H & Fang, S-G 2007, 'Nuclear DNA microsatellite analysis of genetic diversity in captive populations of Chinese water deer', *Small Ruminant Research*, **vol 67**, pp. 252-256.

IUCN 2008, *IUCN Red List of Endangered Species*, viewed 20 January 2010, http://www.iucnredlist.org/details/10550/0 >.

Ivy, JA & Lacy, RC 2010, 'Using molecular methods to improve the genetic management of captive breeding programs for threatened species', in JA DeWoody, JW Bickham, CH Michler, KM Nicols, OE Rhodes, KE Woeste (eds.), *Molecular*

Approaches in Natural Resource Conservation and Management, Cambridge University Press, Cambridge.

Ivy, JA, Miller, A, Lacy, RC & DeWoody, JA 2009, 'Methods and prospects for using molecular data in captive breeding programs: An empirical example using parma wallabies (*Macropus parma*)', *Journal of Heredity*, **vol 100**, pp. 441-454.

Jarne, P & Lagoda, PJL 1996, 'Microsatellites, from molecules to populations and back', *Trends in Ecology and Evolution*, **vol 11**, pp. 424-428.

Jones, AG & Arden, WR 2003, 'Methods of parentage analysis in natural populations', *Molecular Ecology*, **vol 12**, pp. 2511-2523.

Jones, KL, Glenn, TC, Lacy, RC, Pierce, JR, Unruh, N, Mirande, CM & Chavez-Ramirez, F 2002, 'Refining the Whooping Crane studbook by incorporating microsatellite DNA and leg-banding analyses', *Conservation Biology*, **vol 16**, pp. 789-799.

Jukes, TH & Cantor, CR 1969, 'Evolution of protein molecules', in HN Munro (ed.), *Mammalian Protein Metabolism*, Academic Press, New York, USA.

Jule, KR, Leaver, LA & Lea, SEG 2008, 'The effects of captive experience on reintroduction survival in carnivores: A review and analysis', *Biological Conservation*, **vol 141**, pp. 255-263.

Kalinowski, ST, Taper, ML & Marshall, TC 2007, 'Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment', *Molecular Ecology*, **vol 16**, pp. 1099-1106.

Kanthaswamy, S, Kurushima, JD & Smith, DG 2006, 'Inferring Pongo conservation units: a perspective based on microsatellite and mitochondrial DNA analyses', *Primates*, **vol 47**, pp. 310-321.

Karl, SA, Castro, ALF, Lopz, JA, Charvet, P & Burgess, GH 2011, 'Phylogeography and conservation of the bull shark (*Carcharhinus leucas*) inferred from mitochondrial and microsatellite DNA', *Conservation Genetics*, **vol 12**, pp. 371-382.

Kelkar, YD, Strubczewski, N, Hile, SE, Chiaromonte, F, Eckert, KA & Makova, KD 2010, 'What is a microsatellite: A computational and experimental definition based upon repeat mutational behaviour at A/T and GT/AC repeats', *Genome Biology and Evolution*, **vol 2**, pp. 620-635.

Kelkar, YD, Eckert, KA, Chiaromonte, F & Makova, KD 2011, 'A matter of life or death: How microsatellites emerge in and vanish from the human genome', *Genome Research*, **vol 21**, pp. 2038-2048.

Kenney, MC, Chwa, M, Atilano, SR, Falatoonzadeh, P, Ramirez, C, Malik, D, Tarek, M, del Carpio, JC, Nesburn, AB, Boyer, DS, Kuppermann, BD, Vawter, MP, Jazwinski, SM, Miceli, MV, Wallace, DC & Udar, N 2014, 'Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: Implications for population susceptibility to diseases', *Biochimica et Biophysica Acta*, **vol 1842**, pp. 208-219.

Kiemnec-Tyburczy, KM, Richmond, JQ, Savage, AE, Lips, KR & Zamudio, KR 2012, 'Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication', *Heredity*, **vol 109**, pp. 146-155.

Kim, DS, Wang, Y, Oh, HJ, Lee, K & Hahn, Y 2014, 'Frequent loss and alteration of the MOXD2 gene in Catarrhines and whales: A possible connection with the evolution of olfaction', *PlosONE*, **vol 9**, p. e104085.

Kim, SK, Carbone, L, Becquet, C, Mootnick, AR, Li, DJ, Jong , PJ & Wall , JD 2011, 'Patterns of genetic variation within and between gibbon species', *Molecular Biology and Evolution*, **vol 28**, pp. 2211-2218.

Kimura, M 1968, 'Evolutionary rate at the molecular level', *Nature*, **vol 217**, pp. 624-626.

Kimura, M 1980, 'A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences', *Journal of Molecular Evolution*, **vol 16**, pp. 111-120.

Kimura, M & Crow, F 1964, 'The number of alleles that can be maintained in a finite population', *Genetics*, **vol 49**, pp. 725-738.

Kimura, M & Ohta, T 1978, 'Stepwise mutation model and distribution of alleleic frequencies in a finite population', *Proceedings of the National Academy of Science USA*, **vol 75**, pp. 2868-2872.

Kraaijeveld-Smit, FJL, Griffiths, RA, Moore, RD & Beebee, TJC 2006, 'Captive breeding and the fitness of reintroduced species; A test of the responses to predators in a threatened amphibian', *Journal of Applied Ecology*, **vol 43**, pp. 360-365.

Kretzmann, M, Mentzer, L, DiGiovanni Jr., R, Leslie, MS & Amato, G 2006, 'Microsatellite diversity and fitness in stranded juvenile harp seals (*Phoca groenlandica*)', *Journal of Heredity*, **vol 97**, pp. 555-560.

Kruglyak, S, Durrett, RT, Schug, MD & Aquadro, CF 1998, 'Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations', *Proceedings of the National Academy of Sciences USA*, **vol 95**, pp. 10774-10778.

Kryazhimskiy, S & Plotkin, JB 2008, 'The population genetics of dN/dS', *PlosGenetics*, **vol 4,** pp. 1-10.

La Porta, N, Capretti, P, Thomsen, IM, Kasanen, R, Hietala, AM & von Weissenberg, K 2008, 'Forest pathogens with higher damage potential due to climate change in Europe', *Canadian Journal of Plant Pathology*, **vol 30**, pp. 177-195.

Lacobazzi, V, Castegna, A, Infantino, V & Andria, G 2013, 'Mitochondrial DNA methylation as a next-generation biomarker and diagnostic tool', *Molecular Genetics and Metabolism*, **vol 110**, pp. 25-34.

Lacy, RC 2000, 'Should we select genetic alleles in our conservation breeding programs', *Zoo Biology*, **vol 19**, pp. 279-282.

Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ & Higgins, DG 2007, 'Clustal W and Clustal X version 2.0', *Bioinformatics*, **vol 23**, pp. 2947-2948.

Lathuilliere, CM & Crouau-Roy, B 2000, 'Conservation and evolution of microsatellite loci in primate taxa', *American Journal of Primatology*, **vol 50**, pp. 205-214.

Le Gouar, PJ, Vallet, D, David, L, Bermejo, M, Gatti, S, Levréro, F, Petit, EJ & Ménard, N 2009, 'How Ebola impacts genetics of Western lowland gorilla populations', *Plos One*, **vol 4**, p. e8375.

Leclercq, S, Rivals, E & Jarne, P 2010, 'DNA slippage occurs at microsatellite loci without minimal threshold length in humans: A comparative genomic approach', *Genome Biology and Evolution*, **vol 2**, pp. 325-335.

Lewis , OT & Thomas, CD 2001, 'Adaptation to captivity in the butterly *Pieris brassicae (L.)* and the implications for ex situ conservation', *Journal of Insect Conservation*, **vol 5**, pp. 55-63.

Librado, P & Rozas, P 2009, 'DnaSP v5: a software for comprehensive analysis of DNA polymorphism data', *Bioinformatics*, **vol 25**, pp. 1451-1452.

Liu, GQ, Jiang, XP, Wang, JY & Wang, ZY 2006, 'Correlations between heterozygosity at microsatellite loci, mean d2 and body weight in a Chinese native chicken', *Asian-Australasian Journal of Animal Sciences*, **vol 19**, pp. 1671-1677.

Lukas, D & Vigilant, L 2005, 'Reply: Fact, faeces and setting standards for the study of MHC genes using noninvasive samples', *Molecular Ecology*, **vol 14**, pp. 1601-1602.

Lunt, DH & Hyman, BC 1997, 'Animal mitochondrial DNA recombination', *Nature*, **vol 387**, p. 247.

Luo, S-J, Johnson, WE, Martenson, J, Antunes, A, Martelli, P, Uphyrkina, O, Traylor-Holzer, K, Smith, JLD & O'Brien, SJ 2008, 'Subspecies genetic assignments of worldwide captive tigers increase conservation value of captive populations', *Current Biology*, **vol 18**, pp. 592-596.

Luo, S-J, Kim, J-H, Johnson, WE, van der Walt, J, Martenson, J, Yuhki, N, Miquelle, DG, Uphyrkina, O, Goodrich, JM, Quigley, HB, Tilson, R, Brady, G, Martelli, P, Subramaniam, V, McDougal, C, Hean, S, Huang, S-Q, Pan, W, Karanth, UK,

Sunquist, M, et al. 2004, 'Phylogeography and genetic ancestry of tigers (*Panthera tigris*)', *PLoS Biology*, **vol 2**, pp. 2275-2293.

Müller, S, Hollatz, M & Wienberg, J 2003, 'Chomosomal phylogeny and evolution of gibbons (*Hylobatidae*)', *Human Genetics*, **vol 113**, pp. 493-501.

Maddison, DR & Madison, WP 2005, *MacClade 4: Analysis of phylogeny and character evolution.*, viewed 25 March 2014,

"http://macclade.org/index.html" http://macclade.org/index.html >.

Marrero, P, Fregel, R, Cabrera, VM & Nogales, M 2009, 'Extraction of high-quality host DNA from feces and regurgitated seeds: a useful tool for vertebrate ecological studies ', *Biological Research*, **vol 42**, pp. 147-151.

Marshall, TC, Slate, J, Kruuk, LEB & Pemberton, JM 1998, 'Statistical confidence for likelihood-based paternity inference in natural populations', *Molecular Ecology*, **vol 7**, pp. 639-655.

Mason, RAB, Browning, TL & Eldridge, MDB 2011, 'Reduced MHC class II diversity in island compared to mainland populations of the black-footed rock-wallaby (*Petrogale lateralis lateralis*)', *Conservation Genetics*, **vol 12**, pp. 91-103.

McGreevy Jr, TJ, Dabek, L, Gomez-Chiarri, M & Husband, TP 2009, 'Genetic diversity in captive and wild Matschie's Tree Kangaroo (*Dendrolagus matschiei*) from Huon Peninsula, papua New Guinea, based on mtDNA control region sequences', *Zoo Biology*, **vol 28**, pp. 183-196.

McPhee, ME 2003, 'Generations in captivity increases behavioural variance: Considerations for captive breeding and reintroduction programs', *Biological Conservation*, **vol 115**, pp. 71-77.

Meirmans, P & Hedrick, PW 2011, 'Assessing population structure: FST and related measures', *Molecular Ecology Resources*, **vol 11**, pp. 5-18.

Menotti-Raymond, M, David, VA, Lyons, LA, Schäffer , AA, Tomlin, JF, Hutton, MK & O'Brien, SJ 1999, 'A genetic linkage map of microsatellites in the domestic cat (*Felis catus*)', *Genomics*, **vol 57**, pp. 9-23.

Miller, HC, Belov, K & Daugherty, CH 2006, 'MHC class I genes in the Tuatara (*Sphenodon spp.*): Evolution of the MHC in an ancient reptilian Order', *Molecular Biology and Evolution*, **vol 23**, pp. 949-956.

Miller, KA, Chapple, DG, Towns, DR, Ritchie, PA & Nelson, NJ 2009, 'Assessing genetic diversity for conservation management : a case study of a threatened reptile', *Animal Conservation*, **vol 12**, pp. 163-171.

Mishmar, D, Ruiz-Pesini, E, Golik, P, Macaulay, V, Clark, AG, Hosseini, S, Brandon, M, Easley, K, Chen, E, Brown, MD, Sukernik, RI, Olckers, A & Wallace, DC 2003, 'Natural selection shaped regional mtDNA variation in humans', *Proceedings of the National Academy of Sciences USA*, **vol 100**, pp. 171-176.

Monda, K, Simmons, RE, Kressirer, P, Su, B & Woodruff, DS 2007, 'Mitochondrial DNA hypervariable region-1 sequence variation and phylogeny of the Concolor gibbons, Nomascus', *American Journal of Primatology*, **vol 69**, pp. 1285-1306.

Montgomery, ME, Ballou, JD, Nurthen, RK, England, PR, Briscoe, DA & Frankham, R 1997, 'Minimizing kinship in captive breeding programs', *Zoo Biology*, **vol 16**, pp. 377-389.

Montgomery, ME, Woodworth, LM, England, PR, Briscoe, DA & Frankham, R 2010, 'Widespread selective sweeps affecting microsatellites in Drosophila populations

adapting to captivity: Implications for captive breeding programs', *Biological Conservation*, **vol 143**, pp. 1842-1849.

Moreira, MAM, Bonvicino, CR, Soares, MA & Seuánez , HN 2010, 'Genetic diversity of Neotropical primates: Phylogeny, population genetics and animal models for infectious diseases', *Cytogenetic and Genome Research*, **vol 128**, pp. 88-98.

Morin, PA, Chambers, KE, Boesch, C & Vigilant, L 2001, 'Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*)', *Molecular Ecology*, **vol 10**, pp. 1835-1844.

Morin, PA, Luikart, G, Wayne, RK & SNP Workshop Group 2004, 'SNPs in ecology, evolution and conservation', *Trends in Ecology and Evolution*, **vol 19**, pp. 208-215.

Moritz, C, Dowling, TE & Brown, WM 1987, 'Evolution of animal mitochondrial DNA: Relevance for population biology and systematics', *Annual Review of Ecology and Systematics*, **vol 18**, pp. 269-292.

Morrissey, MB & Wilson, AJ 2005, 'The potential costs of accounting for genotypic errors in molecular parentage analyses', *Molecular Ecology*, **vol 14**, pp. 4111-4121.

Muse, SV & Gaut, BS 1994, 'A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the Chloroplast genome', *Molecular Biology and Evolution*, **vol 11**, pp. 715-724.

Nachman, MW, Boyer, SN & Aquadro, CF 1994, 'Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice', *Proceedings of the National Academy of Sciences USA*, **vol 91**, pp. 6364-6368.

Nagaraja, R, MacMillan, S, Kere, J, Jones, C, Griffin, S, Schmatz, M, Terrell, J, Shomaker, M, Jermak, C, Hott, C, Masisi, M, Mumm, S, Srivastava, A, Pilia, G, Featherstone, T, Mazzarella, R, Kesterson, S, McCauley, B, Railey, B, Burough, F, et al. 1997, 'X chromosme map at 75-kb STS resolution, revealing extremes of recombination and GC contect', *Genome Research*, **vol 7**, pp. 210-222.

NCBI 2014, *Open Reading Frame Finder*, viewed 14 July 2014, "http://www.ncbi.nlm.nih.gov/projects/gorf/orfig.cgi"

Neff, BD 2004, 'Mean d2 and divergence time: Transformation and standardizations', *Journal of Heredity*, **vol 21**, pp. 165-171.

Nei, M 1987, *Molecular Evolutionary Genetics*, Columbia University Press, Columbia.

Nei, M & Gojobori, T 1986, 'Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions', *Molecular Biology and Evolution*, **vol 3**, pp. 418-426.

Nguyen, TTT, Ingram, B, Sungan, S, Gooley, G, Sim, SY, Tinggi, D & De Silva , SS 2006, 'Mitochondrial DNA diversity of broodstock of two indigenous mahseer species, *Tor tambroides* and *T. douronensis* (*Cyprinidae*) cultured in Sarawak, Malaysia', *Aquaculture*, **vol 253**, pp. 259-269.

Nichols, JD, Boulinier, T, Hines, JE, Pollock, KH & Sauer, JR 1998, 'Estimating rates of local species extinction, colonization and turnover in animal communities', *Ecological Applications*, **vol 8**, pp. 1213-1225.

Nielson, R 2001, 'Statistical tests of selective neutrality in the age of genomics', *Heredity*, **vol 86**, pp. 641-647.

Nijman, V 2004, 'In-situ and ex-situ status of the Javan gibbon and the role of zoos in conservation of this species', *Contributions to zoology*, **vol 75**, pp. 161-168.

Nsubuga, AM, Holzman, J, Chemnick, LG & Ryder, OA 2010, 'The cryptic genetic structure of the North American captive gorilla population', *Conservation Genetics*, **vol 11**, pp. 161-172.

Nsubuga, AM, Robbins, MM, Roeder, AD, Morin, A, Boesch, C & Vigilant, L 2004, 'Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage', *Molecular Ecology*, **vol 13**, pp. 2089-2094.

Oka, T & Takenaka, O 2001, 'Wild gibbons' parentage tested by non-invasive DNA sampling and PCR-amplified polymorphic microsatellites', *Primates*, **vol 42**, pp. 67-73.

O'Ryan, C, Flamand, JRB & Harley, EH 1994, 'Mitochondrial DNA variation in Black Rhinoceros (*Diceros bicornis*): Conservation management implications', *Conservation Biology*, **vol 8**, pp. 495-500.

Otting, N, de Groot, N, de Vos-Rouweler, AJM, Louwerse, A, Doxiadis, GM & Bontrop, RE 2012, 'Multilocus definition of MHC haplotypes in pedigreed cynomolgus macaques (*Macaca fasicularis*)', *Immunogenetics*, **vol 64**, pp. 755-765.

Otting, N, Heijmans, CMC, Noort, RC, de Groot, NG, Doxiadis, GGM, van Rood, JJ, Watkins, DI & Bontrop, RE 2005, 'Unparalleled complexity of the MHC class I region in rhesus macaques', *Proceedings of the National Academy of Sciences USA*, **vol 102**, pp. 1626-1631.

Overall, ADJ, Byrne, KA, Pilkington, G & Pemberton, JM 2005, 'Heterozygosity, inbreeding and neonatal traits in soay sheep on St Kilda', *Molecular Ecology*, **vol 14**, pp. 3383-3393.

Paine, RT 1971, 'A short-term experimental investigation of resource partitioning in a New Zealand rocky intertidal habitat', *Ecology*, **vol 52**, pp. 1096-1106.

Paine, RT 1980, 'Food webs: Linkage, interaction strength and community infrastructure', *Journal of Animal Ecology*, **vol 49**, pp. 667-685.

Parham, P 2005, 'MHC class I molecules and kirs in human history, health and survival', *Nature Reviews Immunology*, **vol 5**, pp. 201-214.

Pastor, T, Garza, JC, Allen, P, Amos, W & Aguilar, A 2004, 'Low genetic variability in the highly endangered Mediterranean Monk Seal', *Journal of Heredity*, **vol 95**, pp. 291-300.

Pelletier, F, Réale , D, Watters, J, Boakes, EH & Garant, D 2009, 'Value of captive populations for quantitative genetics research', *Trends in Ecology and Evolution*, **vol 24**, pp. 263-270.

Pereira, F, Soares, P, Carneiro, J, Pereira, L, Richards, MB, Samuels, DC & Amorim, A 2008, 'Evidence for variable selective pressures at large secondary structure of the human mitochondrial DNA control region', *Molecular Biology and Evolution*, **vol 25**, pp. 2759-2770.

Perelman, P, Johnson, WE, Roos, C, Seuánez, HN, Horvath, JE, Moreira, MAM, Kessing, B, Pontius, J, Roelke, M, Rumpler, Y, Schneider, MPC, Silva, A, O'Brien, SJ & Pecon-Slattery, J 2011, 'A molecular phylogeny of living primates', *PLoS Genetics*, **vol 7**, pp. 1-17.

Piertney, SB & Oliver, MK 2006, 'The evolutionary ecology of the major histocompatibility complex', *Heredity*, **vol 96**, pp. 7-21.

Piggott, MP & Taylor, AC 2003, 'Extensive evaluation of faecal preservation and DNA extraction methods in Australian native and introduced species', *Australian Journal of Zoology*, **vol 51**, pp. 341-355.

Ping-ping, J, Qiu-lei, L, Sheng-guo, F, Ping, D & Li-ming, C 2005, 'A genetic diversity comparison between captive individuals and wild individuals of Elliot's pheasant (*Syrmaticus ellioti*) using mitochondrial DNA', *Journal of Zheijang University Science*, **vol 6B**, pp. 413-417.

Pokorny, I, Sharma, R, Goyal, SP, Mishra, S & Tiedmann, R 2010, 'MHC class I and MHC class II DRB gene variability in wild and captive Bengal tigers (*Panthera tigris tigris*)', *Immunogenetics*, **vol 62**, pp. 667-679.

Pounds, JA, Bustamante, MR, Coloma, LA, Consuegra, JA, Fogden, MPL, Foster, PN, La Marca, E, Masters, KL, Merino-Viteri, A, Puschendorf, R, Ron, SR, Sánchez-Azofeifa, GA, Still, CJ & Young, BE 2006, 'Widespread amphibian extinctions from epidemic disease driven by global warming', *Nature*, **vol 439**, pp. 161-167.

Pukazhenthi, B, Comizzoli, P, Travis, AJ & Wildt, DE 2006, 'Applications of emerging technologies to the study and conservation of threatened and endangered species', *Reproduction, Fertility and Development* , **vol 18**, pp. 77-90.

Queller, DC & Goodnight, KF 1989, 'Estimating relatedness using genetic markers', *Evolution*, **vol 43**, pp. 258-275.

Ralls, K, Ballou, JD, Rideout, BA & Frankham, R 2000, 'Genetic management of chondrodystrophy in California condors', *Animal Conservation*, **vol 3**, pp. 145-153.

Ramírez-Soriano, A, Ramos-Onsins, SE, Rozas, J, Calafell, F & Navarro, A 2008, 'Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination', *Genetics*, **vol 179**, pp. 555-567.

Ray, JW, King, RB, Duvall, MR, Robinson, JW, Jaeger, CP, Dreslik, MJ, Swanson, BJ & Mulkerin, D 2013, 'Genetic analysis and captive breeding program design for the Eastern Massasauga *Sistrurus catenatus catenatus'*, *Journal of Fish and Wildlilfe Management*, **vol 4**, pp. 104-112.

Reed, DH & Frankham, R 2003, 'Correlation between fitness and genetic diversity', *Conservation Biology*, **vol 17**, pp. 230-237.

Reed, DH, Lowe, EH, Briscoe, DA & Frankham, R 2003, 'Inbreeding and extinction: Effects of rate of inbreeding', *Conservation Genetics*, **vol 4**, pp. 405-410.

Reynolds, J, Weir, BS & Cockerham, CC 1983, 'Estimation of the coancestry coefficient: Basis for a short-term genetic distance', *Genetics*, **vol 105**, pp. 767-779.

Robinson, J, Halliwell, JA, McWilliam, H, Lopez, R & Marsh, SGE 2014a, *Index of /pub/databases/ipd/imgt/hla/fasta/*, viewed 24 June 2014, ftp://ftp.ebi.ac.uk/pub/databases/ipd/imgt/hla/fasta/

Robinson, J, Halliwell, JA, McWilliam, H, Lopez, R & Marsh, SGE 2014b, *Index of /pub/databases/ipd/mhc/nhp*, viewed 24 June 2014, ftp://ftp.ebi.ac.uk/pub/databases/ipd/mhc/nhp

Robinson, NA 1995, 'Implications from mitochondrial DNA for management to conserve the Eastern barred bandicoot (*Perameles gunnii*)', *Conservation Biology*, **vol 9**, pp. 114-125.

Rodrigáñez, J, Barragan, C, Alves, E, Gortázar, C, Toro, MA & Silió , L 2008, 'Genetic diversity and allelic richness in Spanish wild and domestic pig population estimated from microsatellite markers', *Spanish Journal of Agricultural Research*, **vol 6**, pp. 107-115.

Rodriguez-Barreto, D, Consuegra, S, Jerez, S, Cejas, JR, Martin, V & Lorenzo, A 2013, 'Using molecular markers for pedigree reconstruction of the greater amberjack (*Seriola dumerili*) in the absence of parental information', *Animal Genetics*, **vol 44**, pp. 596-600.

Rogers, J, Garcia, R, Shelledy, W, Kaplan, J, Arya, A, Johnson, Z, Bergstrom, M, Novakowski, L, Nair, P, Vinson, A, Newman, D, Heckman, G & Cameron, J 2006, 'An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci', *Genomics*, **vol** 87, pp. 30-38.

Rogers, J & Gibbs, RA 2014, 'Comparative primate genomics: Emerging patterns of genome content and dynamics', *Nature Reviews Genetics*, **vol 15**, pp. 347-359.

Rogers, J, Mahaney, MC, Witte, SM, Nair, S, Newman, D, Wedel, S, Rodriguez, LA, Rice, KS, Slifer, SH, Perelygin, A, Slifer, M, Palladino-Negro, P, Newman, T, Chambers, K, Joslyn, G, Parry, P & Morin, PA 2000, 'A genetic linkage map of the baboon (*Papio hamadryas*) genome based on human microsatellite polymorphisms', *Genomics*, **vol 67**, pp. 237-247.

Rokas, A, Ladoukakis, E & Zouros, E 2003, 'Animal mitochondrial DNA recombination revisited', *Trends in Ecology and Evolution*, **vol 18**, pp. 411-417.

Roos, C & Geissman, T 2001, 'Molecular phylogeny of the major Hylobatid divisions', *Molecular Phylogenetics and Evolution*, **vol 19**, pp. 486-494.

Rubinsztein, DC, Amos, W & Cooper, G 1999, 'Microsatellite and trinucleotide repeat evolution: Evidence for mutational bias and different rates of evolution in different lineages', *Philosophical Transactiobns of the Royal Society of London B*, **vol 354**, pp. 1095-1099.

Rudnick, JA & Lacy, RC 2008, 'The impact of assumptions about founder relationships on the effectiveness of captive breeding strategies', *Conservation Genetics*, **vol 9**, pp. 1439-1450.

Ruiz-López, MJ, Gañan, N, Godoy, JA, Del Olmo, A, Garde, J, Espeso, G, Vargas, A, Martinez, F, Roldán, ERS & Gomendio, M 2012, 'Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals', *Conservation Biology*, **vol 26**, pp. 1121-1129.

Sagvik, J, Uller, T & Olsson, M 2005, 'Outbreeding depression in the common frog, *Rana temporaria*', *Conservation Genetics*, **vol 6**, pp. 205-211.

Sainudiin, R, Durrett, RT, Aquadro, CF & Nielsen, R 2004, 'Microsatellite mutation models: Insights from a comparison of humans and chimpanzees', *Genetics*, **vol 168**, pp. 383-395.

Saitou, N & Nei, M 1987, 'The neighbour-joining method: A new method for reconstructing phylogenetic trees', *Molecular Biology and Evolution*, **vol 4**, pp. 406-425.

Santos, T, Pérez-Tris, J, Roberto, CR, Tellería, JL & Díaz , JA 2009, 'Monitoring the performance of wild-born and introduced lizards in a fragmented landscape: Implications for *ex situ* conservation programmes', *Biological Conservation*, **vol 142**, pp. 2923-2930.

Sato, M & Sato, K 2013, 'Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA', *Biochimica et Biophysica Acta*, **vol 1833**, pp. 1979-1984.

Sauermann, U, Nürnberg, P, Bercovitch, FB, Berard, JD, Trefilov, A, Widdig, A, Kessler, M, Schmidtke, J & Krawczak, M 2001, 'Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques', *Human Genetics*, **vol 108**, pp. 249-254.

Sbisà, E, Tanzariello, F, Reyes, A, Pesole, G & Saccone, C 1997, 'Mammalian mitochondrial D-loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications', *Gene*, **vol 205**, pp. 125-140.

Schmidt, HA & Haeseler, AV 2009, 'Phylogenetic inference using maximum likelihood methods', in P Lemey, M Salemi, A-M Vandamme (eds.), *The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edn, Cambridge University Press, Cambridge, UK.

Schwensow, N, Eberle, M & Sommer, S 2008, 'Compatibility counts: MHC-associated mate choice in a wild promiscuous primate', *Proceedings of the Royal Society B*, **vol 275**, pp. 555-564.

Schwensow, N, Fietz, J, Dausmann, K & Sommer, S 2008, 'MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate', *Evolutionary Ecology*, **vol 22,** pp. 617-636.

Scribner, KT, Gust, JR & Fields, RL 1996, 'Isolation and characterization of novel salmon microsatellite loci: Cross-species amplification and population genetic

applications', *Canadian Journal of Fisheries and Aquatic Sciences*, **vol 53**, pp. 833-841.

Seddon, PJ, Armstrong, DP & Maloney, RF 2007, 'Developing the science of reintroduction biology', *Conservation Biology*, **vol 21**, pp. 303-312.

Seligmann, H, Krishnan, NM & Rao, BJ 2006, 'Possible multiple origins of replication in primate mitochondria: Alternative role of tRNA sequences', *Journal of Theoretical Biology*, **vol 241**, pp. 321-332.

Setchell, JM, Charpentier, MJE, Abbott, KM, Wickings, EJ & Knapp, LA 2010, 'Opposites attract: MHC-associated mate choice in a polygynous primate', *Journal of Evolutionary Biology*, **vol 23**, pp. 136-148.

Setchell, JM & Huchard, E 2010, 'The hidden benefits of sex: Evidence for MHC-associated mate choice in primate societies', *Bioessays*, **vol 32**, pp. 940-948.

Setchell, JM, Vaglio, S, Abbott, KM, Moggi-Cecchi, J, Boscaro, F, Pieraccini, G & Knapp, LA 2011, 'Odour signals major histocompatibility complex genotype in an Old World monkey', *Proceedings of the Royal Society of London B*, **vol 278**, pp. 274-280.

Sharma, PC, Grover, A & Kahl, G 2007, 'Mining microsatellites in eukaryotic genomes', *Trends in Biotechnology*, **vol 25**, pp. 490-498.

Siddle, HV, Marzec, J, Cheng, Y, Jones, M & Belov, K 2010, 'MHC gene coy number variation in the Tasmanian devils: Implications for the spread of a contagious cancer', *Proceedings of the Royal Society B*, **vol 277**, pp. 2001-2006.

Simberloff, D 1998, 'Flagships, umbrellas and keystones: Is single-species management passe in the landscape era?', *Biological Conservation*, **vol 83**, pp. 247-257.

Simonsen, KL, Churchill, GA & Aquadro, CF 1995, 'Properties of statistical tests of neutrality for DNA polymorphism data', *Genetics*, **vol 141**, pp. 413-429.

Sinnock, P 1975, 'The Wahlund effect for the two-locus model', *The American Naturalist*, **vol 109**, pp. 565-570.

Slade, RW 1992, 'Limited MHC polymorphism in the Southern elephant seal: Implications for MHC evolution and marine mammal population biology', *Proceedings of the Royal Society of London B*, **vol 249**, pp. 163-171.

Slade, R & McCallum, H 1992, 'Overdominant vs. frequency-dependent selection at MHC loci', *Genetics*, **vol 132**, pp. 861-862.

Slate, J & Pemberton, JM 2002, 'Comparing molecular measures for detecting inbreeding depression', *Journal of Evolutionary Biology*, **vol 15**, pp. 20-31.

Slatkin, M 1995, 'A measure of population subdivision based on microsatellite allele frequencies', *Genetics*, **vol 139**, pp. 457-462.

Smith, S, Belov, K & Hughes, J 2010, 'MHC screening for marsupial conservation: Extremely low levels of class II diversity indicate population vulnerability for an endangered Australian marsupial', *Conservation Genetics*, **vol 11**, pp. 269-278.

Snyder, NFR, Derrickson, SR, Beissinger, SR, Wiley, JW, Smith, TB, Toone, WD & Miller, B 1996, 'Limitations of captive breeding in endangered species recovery', *Conservation Biology*, **vol 10**, pp. 338-348.

Sommer, S, Schwab, D & Ganzhorn, JU 2002, 'MHC diversity of endemic Malagasy rodents in relation to geographic range and social system', *Behavioural Ecology and Sociobiology*, **vol 51**, pp. 214-221.

St. John, JC 2014, 'Chapter 34: Mitochondrial DNA: Its Transmission from Gametes and Embryos', in J Cibelli, J Gurdon, J Wilmut, R Jaenisch, R Lanza, M West, K Campbell (eds.), *Principles of Cloning*, Second Edition edn, Academic Press, Boston, USA.

Stewart, JB, Freyer, C, Elson, JL & larsson, N-G 2008, 'Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease', *Nature Reviews Genetics*, **vol 9**, pp. 657-662.

Stoinski, TS & Beck, BB 2004, 'Changes in locomotor and foraging skills in captive-born, reintroduced golden lion tamarins', *American Journal of Primatology*, **vol 62**, pp. 1-13.

Stoletzki, N & Eyre-Walker, A 2011, 'The positive correlation between dN/dS and dS in mammals is due to runs of adjacent substitutions', *Molecular Biology and Evolution*, **vol 28**, pp. 1371-1380.

Sutton, JT, Nakagawa, S, Robertson, BC & Jamieson, IG 2011, 'Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes', *Molecular Ecology*, **vol 20**, pp. 4408-4420.

Tajima, F 1983, 'Evolutionary relationship of DNA sequences in finite populations ', *Genetics*, **vol 105**, pp. 437-460.

Tajima, F 1989, 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism', *Genetics*, **vol 123**, pp. 585-595.

Takacs, Z, Morales, JC, Geissman, T & Melnick, DJ 2005, 'A complete species-level phylogeny of the Hylobatidae based on mitochondrial ND3-ND4 gene sequences', *Molecular Phylogenetics and Evolution*, **vol 36**, pp. 456-467.

Tamura, K & Nei, M 1993, 'Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees', *Molecular Biology and Evolution*, **vol 10**, pp. 512-526.

Tamura, K, Stecher, G, Peterson, D, Filipski, A & Kumar, S 2013, 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', *Molecular Biology and Evolution*, **vol 30**, pp. 2725-2729.

Temperley, ND, Webster, LMI, Adam, A, Keller, LF & Johnson, CD 2009, 'Cross-species utility of microsatellite markers in Trichostrongyloid nematodes', *The Journal of Parasitology*, **vol 95**, pp. 487-489.

Thinh, VN, Mootnick, AR, Geissman, T, Li, M, Ziegler, T, Agil, M, Moisson, P, Nadler, T, Walter, L & Roos, C 2010, 'Mitochondrial evidence for multiple radiations in the evolutionary history of small apes', *BioMed Central Evolutionary Biology*, **vol 10**, pp. 1-13.

Tomasco, IH & Lessa, EP 2014, 'Two mitochondrial genes under episodic positive selection in subterranean octodontoid rodents', *Gene*, **vol 534**, pp. 371-378.

Trong, TQ, Bers, NV, Crooijmans, R, Dibbits, B & Komen, H 2013, 'A comparison of microsatellites and SNPs in parental assignment in the GIFT strain of Nile tilapia (*Oreochromis niloticus*): The power of exclusion', *Aquaculture*, **vol 388**, pp. 14-23.

Trowsdale, J 2011, 'The MHC, disease and selection', *Immunology Letters*, **vol 137**, pp. 1-8.

Tsaousis, AD, Martin, DP, Ladoukakis, ED, Posada, D & Zouros, E 2005, 'Widespread recombination in published animal mtDNA sequences', *Molecular Biology and Evolution*, **vol 22**, pp. 925-933.

Tsuji, J, Frith, MC, Tomii, K & Horton, P 2012, 'Mammalian NUMT insertion is non-random', *Nucleic Acids Research*, **vol 40**, pp. 9073-9088.

Tymchuk, WE, Sundström, LF & Devlin, RH 2007, 'Growth and survival trade-offs and outbreeding depression in rainbow trout (*Oncorhynchus mykiss*)', *Evolution*, **vol 61**, pp. 1225-1237.

Tzen, JM, Hsu, H-J & Wang, M-N 2008, 'Redefinition of hypervariable region I in mitochondrial DNA control region and comparing its diversity among various ethnic groups', *Mitochondrion*, **vol 8**, pp. 146-154.

Uhlemann, A-C, Szlezak, NA, Vonthein, R, Tomiuk, J, Emmer, SA, Lell, B, Kremsner, PG & Kun, JFJ 2004, 'DNA phasing by TA dinucleotide microsatellite length determines in vitro and in vivo expression of the gp91pox subunit of NADPH oxidase and mediates protection against severe malaria', *Journal of Infectious Diseases*, **vol 189**, pp. 2227-2234.

Ujvari, B & Belov, K 2011, 'Major histocompatibility complex (MHC) markers in conservation biology', *International Journal of Molecular Sciences*, **vol 12**, pp. 5168-5186.

Valtonen, M, Palo, JU, Ruokonen, M, Kunnasranta, M & Nyman, T 2012, 'Spatial and temporal variation in genetic diversity of an endangered freshwater seal', *Conservation Genetics*, **vol 13**, pp. 1231-1245.

van Oosterhout, C, Hutchinson, WF, Wills, DPM & Shipley, P 2004, 'MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data', *Molecular Ecology Notes*, **vol 4**, pp. 535-538.

Vandamme, A-M 2009, 'Basic concepts of molecular evolution', in P Lemey, M Salemi, A-M Vandamme (eds.), *The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edn, Cambridge University Press, Cambridge, UK.

Varela, MA & Amos, W 2010, 'Heterogeneous distribution of SNPs in the human genome: Microsatellites as predictors of nucleotide diversity and divergence', *Genomics*, **vol 95**, pp. 151-159.

Varela, MA, Sanmiguel, R, Gonzalez-Tizon, A & Martinez-Lage, A 2008, 'Heterogeneous nature and distribution of interruptions in dinucleotides may indicate the existence of biased substitutions underlying microsatellite evolution', *Journal of Molecular Evolution*, **vol 66**, pp. 575-580.

Välimäki , K, Hinten, G & Hanski, I 2007, 'Inbreeding and competitive ability in hte common shres (Sorex araneus)', *Behavioural Ecology and Sociobiology*, **vol 61**, pp. 997-1005.

Vigilant, L & Bradley, BJ 2004, 'Genetic variation in gorillas', *American Journal of Primatology*, **vol 64**, pp. 161-172.

Viguera, E, Canceill, D & Ehrlich, SD 2001, 'Replication slippage involves DNA polymerase pausing and dissociation', *The European Molecular Biology Organization Journal*, **vol 20**, pp. 2587-2595.

Walsh, PD 2000, 'Sample size for the diagnosis of conservation units', *Conservation Biology*, **vol 14**, pp. 1533-1537.

Wasser, SK, Houston, CS, Koehler, GM, Cadd, GG & Fain, SR 1997, 'Techniques for application of faecal DNA methods to field studies of Ursids', *Molecular Ecology*, **vol 6**, pp. 1091-1097.

Weir, BS & Cockerham, CC 1984, 'Estimating F-statistics for the analysis of population structure', *Evolution*, **vol 38**, pp. 1358-1370.

Whittaker, DJ 2005, 'Evolutionary genetics of Kloss's gibbons (*Hylobates Klossii*): Systematics, phylogeography and conservation.', PhD Thesis, The City University of New York, New York.

Whittaker, DJ, Morales, JC & Melnick, DJ 2007, 'Resolution of the Hylobates phylogeny: Congruence of mitochondrial D-loop sequences with molecular, behavioural and morphological data', *Molecular Phylogenetics and Evolution*, **vol 45**, pp. 620-628.

Williams, SE & Hoffman, EA 2009, 'Minimizing genetic adaptation in captive breeding programs: A review', *Biological Conservation*, **vol 142**, pp. 2388-2400.

Witzenberger, KA & Hochkirch, A 2011, 'Ex situ conservation genetics: A review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species', *Biodiversity Conservation*, **vol 20**, pp. 1843-1861.

World Wildlife Fund 2014, 'Living Planet Report 2014', WWF International, Gland, Switzerland.

Worm, B, Barbier, EB, Beaumont, N, Duffy, JE, Folke, C, Halpern, BS, Jackson, JBC, Lotze, HK, Micheli, F, Palumbi, SR, Sala, E, Selkoe, KA, Stachowicz, JJ & Watson, R 2006, 'Impacts on biodiversity loss on ocean ecosystem services', *Science*, **vol 314**, pp. 787-790.

Wright, S 1951, 'The genetical structure of populations', *Annals of Eugenics*, **vol 15**, pp. 323-353.

Yamazaki, K, Beauchamp, GK, Kupniewski, D, Bard, J, Thomas, L & Boyse, EA 1988, 'Familial imprinting determines H-2 selective mating preferences', *Science*, **vol 240**, pp. 1331-1332.

Young, TP 1987, 'Increased thorn length in *Acacia depranolobium* - an induced response to browsing', *Oecologica*, **vol 71**, pp. 436-438.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Kidd, JR, Deinard, AS, Ruder, O, Kidd, KK & Li, W-H 2003, 'Low nucleotide diversity in chimpanzees and bonobos', *Genetics*, **vol 164**, pp. 1511-1518.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Ryder, O & Li, W-H 2004, 'Nucleotide diversity in Gorillas', *Genetics*, **vol 166**, pp. 1375-1383.

Zedrosser, A, Bellemain, E, Taberlet, P & Swenson, JE 2007, 'Genetic estimates of annual reproductive success in male brown bears: The effects of body size, age, internal relatedness and population density', *Journal of Animal Ecology*, **vol 76**, pp. 368-375.

Zenger, KR, Eldridge, MDB & Johnston, PG 2005, 'Phylogenetics, population structure and genetic diversity of the endangered southern brown bandicoot (*Isoodon obesulus*) in south-eastern Australia', *Conservation Genetics*, **vol 6**, pp. 193-204.

Zhang, D-X & Hewitt, GM 1996, 'Nuclear integrations: challenges for mitochondrial DNA markers', *Trends in Ecology and Evolution*, **vol 11**, pp. 247-251.

Zhang, Y-W, Morin, PA, Ryder, OA & Zhang, Y-P 2001, 'A set of tri- and tetra-nucleotide microsatellite loci useful for population analyses in gorillas (*Gorilla gorilla gorilla*) and orangutans (*Pongo pygmaeus*)', *Conservation genetics*, **vol 2**, pp. 391-395.

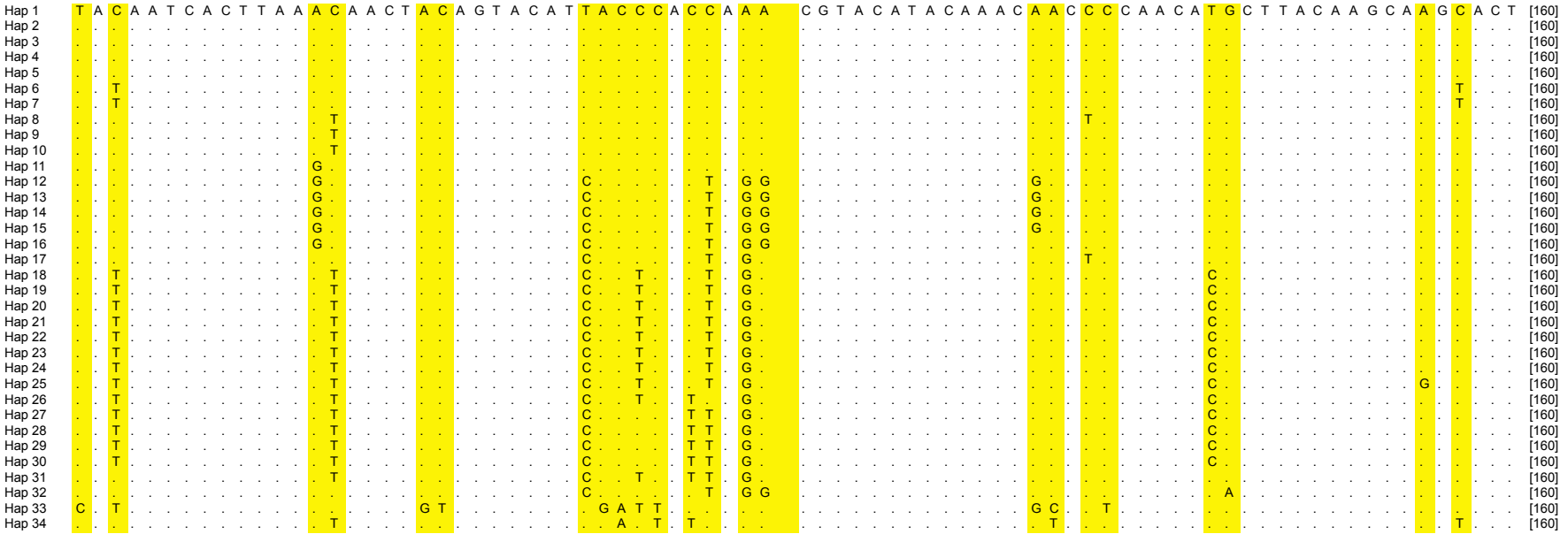
Zhang, Q, Zeng, Z-G, Ji, Y-J, Zhang, D-X & Song, Y-L 2008, 'Microsatellite variation in china's Hainan Eld's deer (*Cervus eldi hainanus*) and implications for their conservation', *Conservation Genetics*, **vol 9**, pp. 507-514.

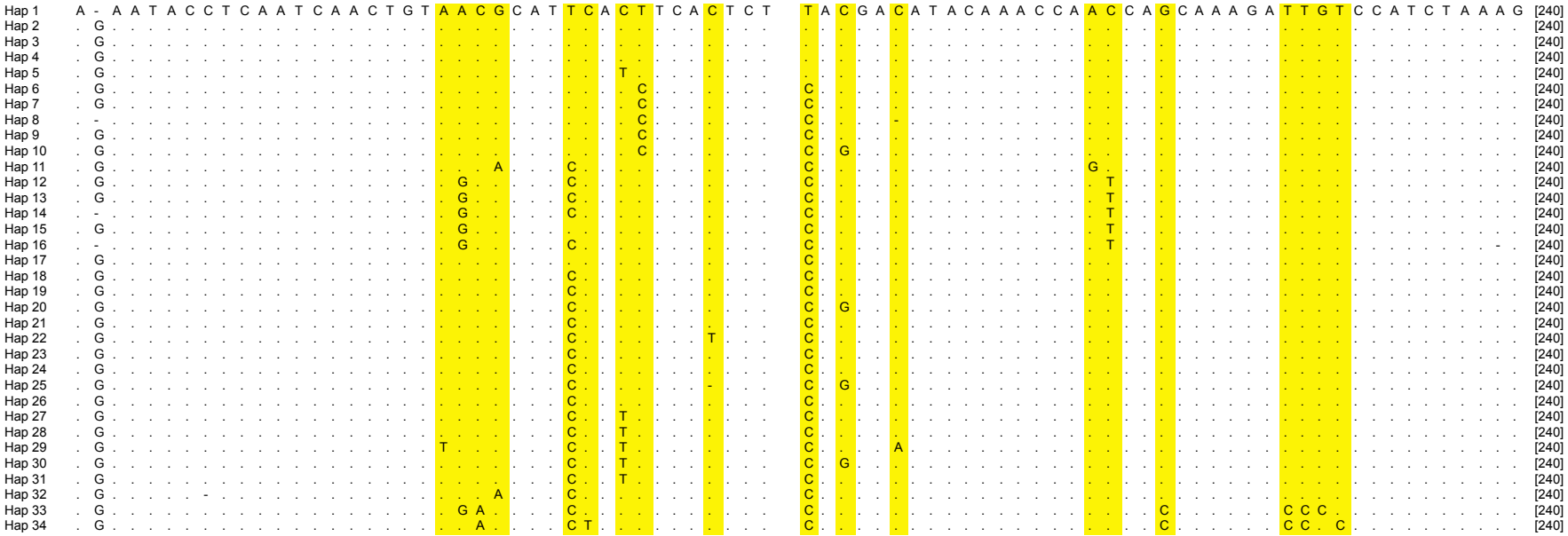
Appendix I - Results summary information per individual

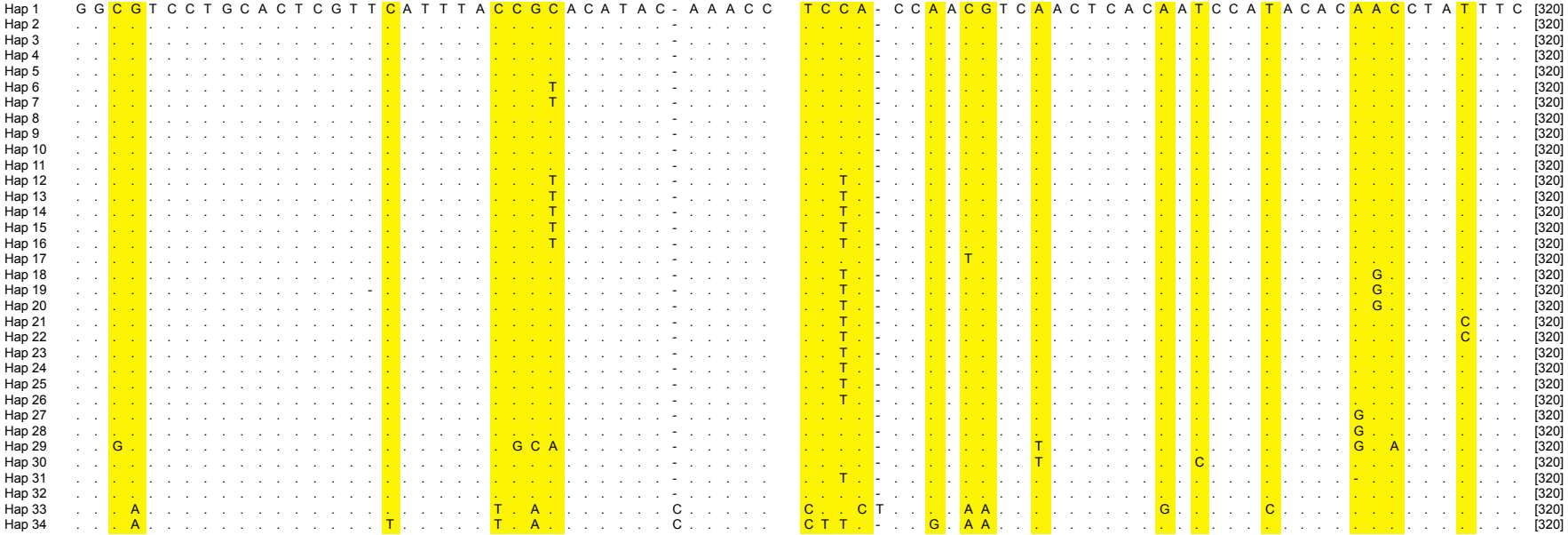
Primate Ref	mtDNA Haplotype	Mean Kinship	Standardised Heterozygosity	MHC Haplotypes	
H.mol1	Hap_12	0.0083	1.195	6	?
H.mol2	Hap_13	0.0875	1.165	3a	5
H.mol3	Hap_18	0.0000	0.698	1	4
H.mol4	Hap_26	0.0000	1.072	-	-
H.mol5	Hap_2	0.0083	0.891	1	3a
H.mol6	Hap_18	0.0000	0.154	1	7
H.mol7	Hap_26	0.0208	1.210	3a	5
H.mol8	Hap_21	-	1.210	7	8
H.mol9	Hap_26	0.1354	1.132	1	2
H.mol10	Hap_26	0.1479	1.163	9	10
H.mol11	Hap_26	0.1271	0.805	3a	5
H.mol12	Hap_26	0.1271	0.926	3a	5
H.mol13	Hap_21	0.1083	0.775	3a	4
H.mol14	Hap_26	0.1271	1.163	5	10
H.mol15	Hap_21	0.0604	1.252	3b	4
H.mol16	Hap_21	0.0688	0.678	4	7
H.mol17	Hap_26	-	1.073	4	7
H.mol18	Hap_26	0.1521	1.073	2	10
H.mol19	Hap_26	0.1167	0.800	2	4
H.mol20	Hap_26	0.1167	1.065	2	7
H.mol21	Hap_26	-	1.028	9	10

Appendix II – Nucleotide sequences for mtDNA haplotypes and DRB exon 2 alleles. Protein sequence for DRB exon 2 allele

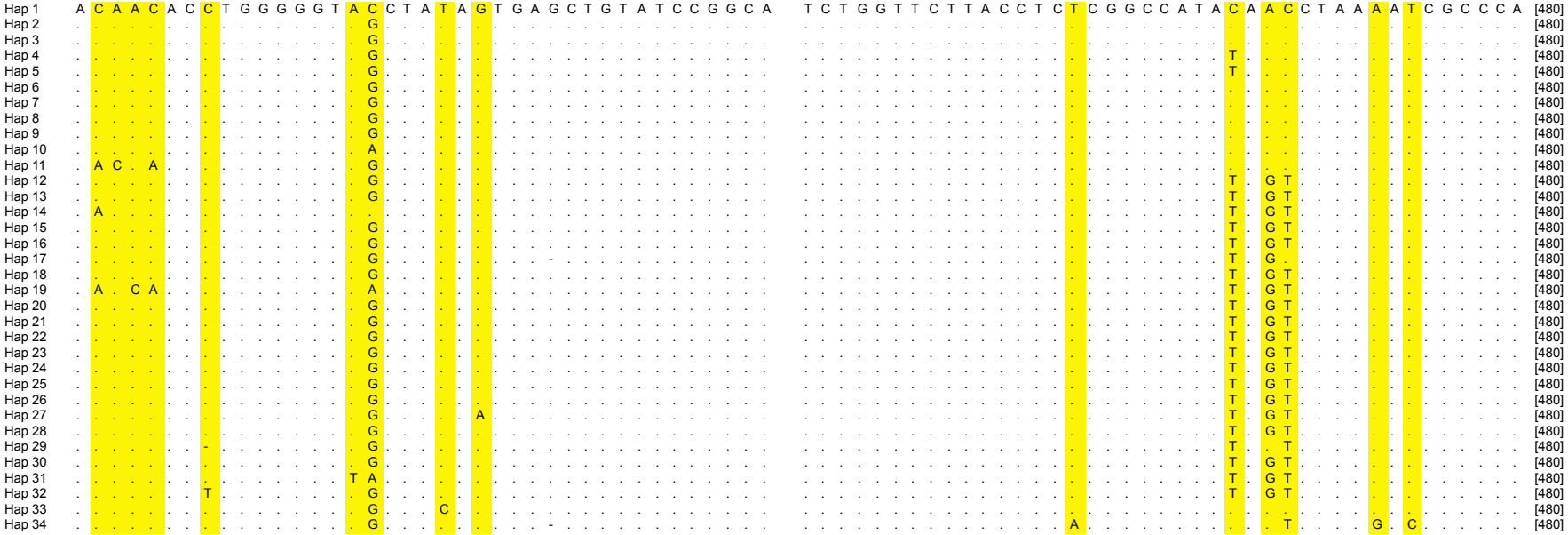
```
Hap 1   C C C C A G T A C T A A C C T G - T T C T T C C A C A A C C C T A T G T A C T   T C G T A C A T T A C T G C C A G T C C C C - A T G A A T A T T G T A - A G T A C [80]
Hap 2   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . . . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 3   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . . . . . . C . . . . . . . . . . . . C . . . . . [80]
Hap 4   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 5   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 6   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 7   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 8   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 9   . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 10  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 11  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 12  . . . . . . . . T . . . . . . . C . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . C . . . C . . . . . [80]
Hap 13  . . . . . . . . T . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . C . . . C . . . . . [80]
Hap 14  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . C . . . C . . . . . [80]
Hap 15  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . C . . . C . . . . . [80]
Hap 16  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . C . . . C . . . . . [80]
Hap 17  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . T . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 18  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 19  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 20  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 21  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 22  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 23  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . - . . . . . . . . . . . . T . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 24  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 25  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 26  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 27  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 28  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 29  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 30  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . C . C . . . . . [80]
Hap 31  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 32  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . G . . . . . . . . C . . . . . [80]
Hap 33  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . T . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
Hap 34  . . . . . . . . . . . . . . . . - . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . . . . C . . . . - . . . . . . . . . . . . C . . . . . [80]
```

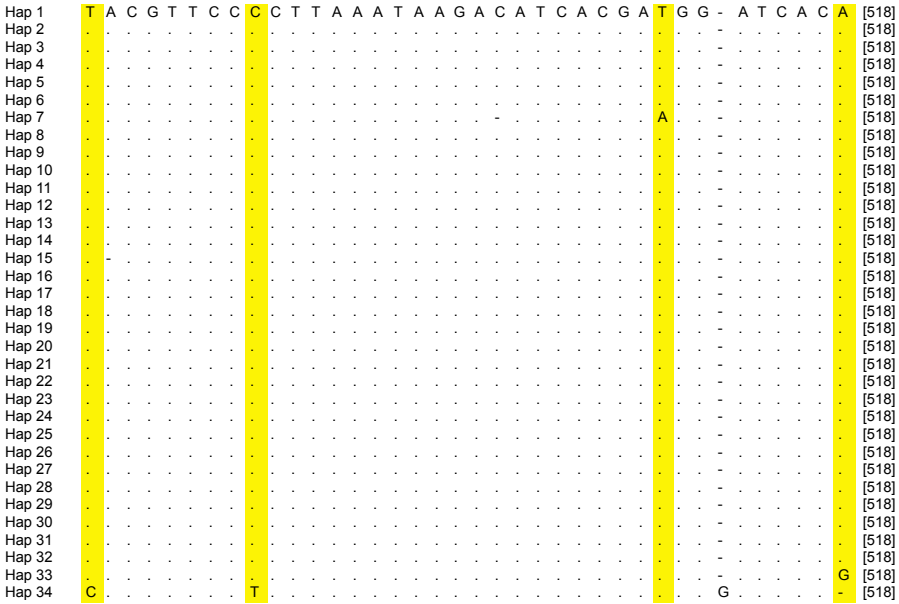

Hap 1 TACAATCACTTAAACAACTACAGTACATTACCCACCAAA CGTACATACAAACAACCCCAAACATGCTTACAAGCAAGCACT [160]
Hap 2 [160]
Hap 3 [160]
Hap 4 [160]
Hap 5 [160]
Hap 6 [160]
Hap 7 [160]
Hap 8 [160]
Hap 9 [160]
Hap 10 [160]
Hap 11 [160]
Hap 12 [160]
Hap 13 [160]
Hap 14 [160]
Hap 15 [160]
Hap 16 [160]
Hap 17 [160]
Hap 18 [160]
Hap 19 [160]
Hap 20 [160]
Hap 21 [160]
Hap 22 [160]
Hap 23 [160]
Hap 24 [160]
Hap 25 [160]
Hap 26 [160]
Hap 27 [160]
Hap 28 [160]
Hap 29 [160]
Hap 30 [160]
Hap 31 [160]
Hap 32 [160]
Hap 33 [160]
Hap 34 [160]


Hap 1 A - A A T A C C T C A A T C A A C T G T A A C G C A T T C A C T T C A C T C T T A C G A C A T A C A A A C C A A C C A G C A A A G A T T G T C C A T C T A A A G [240]
Hap 2 . G [240]
Hap 3 . G [240]
Hap 4 . G [240]
Hap 5 . G [240]
Hap 6 . G [240]
Hap 7 . G [240]
Hap 8 . - [240]
Hap 9 . G [240]
Hap 10 . G [240]
Hap 11 . G [240]
Hap 12 . G [240]
Hap 13 . G [240]
Hap 14 . - [240]
Hap 15 . G [240]
Hap 16 . - [240]
Hap 17 . G [240]
Hap 18 . G [240]
Hap 19 . G [240]
Hap 20 . G [240]
Hap 21 . G [240]
Hap 22 . G [240]
Hap 23 . G [240]
Hap 24 . G [240]
Hap 25 . G [240]
Hap 26 . G [240]
Hap 27 . G [240]
Hap 28 . G [240]
Hap 29 . G [240]
Hap 30 . G [240]
Hap 31 . G [240]
Hap 32 . G [240]
Hap 33 . G [240]
Hap 34 . G [240]


Hap 1 GGCGTCCTGCACTCGTTCATTTTACCCGCACATACAAAACCTCCA-CCAACGTCAACTCACACATCCATACACACAACACCTATTTC [320]
Hap 2 [320]
Hap 3 [320]
Hap 4 [320]
Hap 5 [320]
Hap 6 [320]
Hap 7 [320]
Hap 8 [320]
Hap 9 [320]
Hap 10 [320]
Hap 11 [320]
Hap 12 [320]
Hap 13 [320]
Hap 14 [320]
Hap 15 [320]
Hap 16 [320]
Hap 17 [320]
Hap 18 [320]
Hap 19 [320]
Hap 20 [320]
Hap 21 [320]
Hap 22 [320]
Hap 23 [320]
Hap 24 [320]
Hap 25 [320]
Hap 26 [320]
Hap 27 [320]
Hap 28 [320]
Hap 29 [320]
Hap 30 [320]
Hap 31 [320]
Hap 32 [320]
Hap 33 [320]
Hap 34 [320]

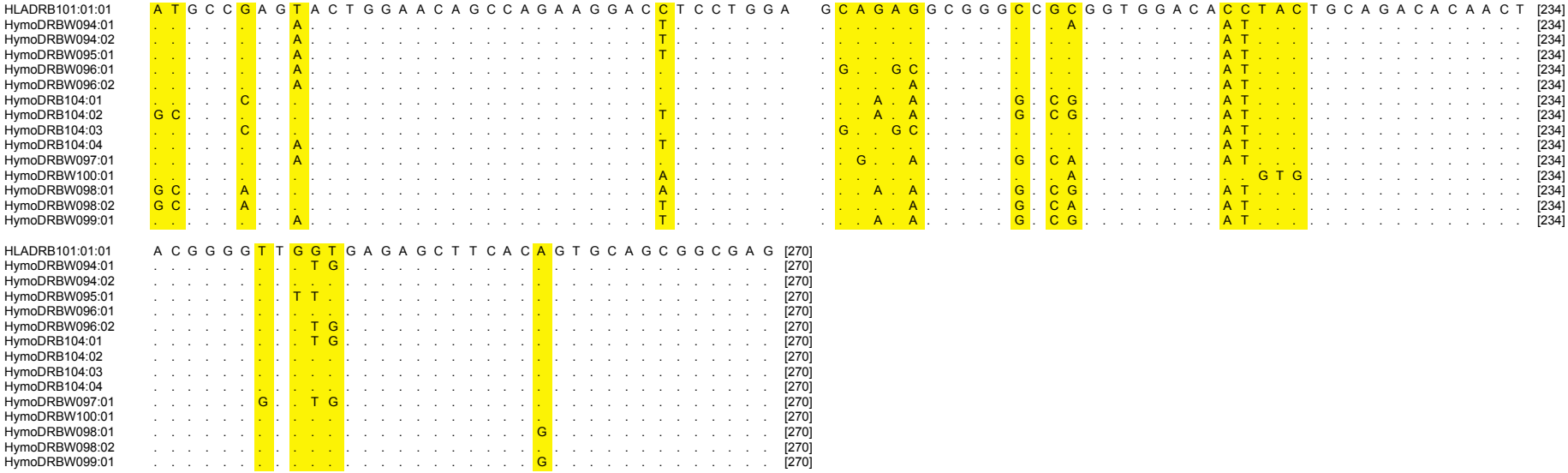
Hap 1	AGATGGGA-TTTCCCGTCCAGCATCCTCCGTGAAATCACCAACCCGCACAA-GAATACTAACTCCCCTCGCTCCGGGCTT	[400]
Hap 2	........G..-..G..........................	[400]
Hap 3	........G..-..G..........................	[400]
Hap 4	.......GG..-..G......T............C.....	[400]
Hap 5	.......GG..-............................	[400]
Hap 6	........G..-..G..........................	[400]
Hap 7	........G..-..G..........................	[400]
Hap 8	........A..-..............TC............	[400]
Hap 9	........-..-..G..........................	[400]
Hap 10	........G..-..G.........T...............	[400]
Hap 11	........G..-..G..A.......................	[400]
Hap 12	....A.A.G..-..G.......................C.	[400]
Hap 13	....A.A.G..-..G.......................C.	[400]
Hap 14	....A.A.G..-..........................C.	[400]
Hap 15	....A.A.G..-..G.......................C.	[400]
Hap 16	....A.A.G..-..G.......................C.	[400]
Hap 17	....A...G..-..G.......................C.	[400]
Hap 18	........G..-..G..........................	[400]
Hap 19	......-.G..-..G..........................	[400]
Hap 20	........G..-..G..........................	[400]
Hap 21	........G..-..G..........................	[400]
Hap 22	........G..-..G..........................	[400]
Hap 23	........G..-..G..........................	[400]
Hap 24	........G..-..G..........................	[400]
Hap 25	........G..-..G..........................	[400]
Hap 26	........G..-..G..........................	[400]
Hap 27	........G..-..............TC............	[400]
Hap 28	........G..-............................	[400]
Hap 29	........G..-..-..........................	[400]
Hap 30	........G..-..G..........................	[400]
Hap 31	....A...-..............................G...............-..G..........................	[400]
Hap 32	.......TG..-..G..........................	[400]
Hap 33	......TGG.CC..T.C.....................A........C...A...............................	[400]
Hap 34	....A.A.G.CC..T.C.....................A........C...A...............................	[400]


Hap 1 A C A A C A C C T G G G G G T A C C T A T A G T G A G C T G T A T C C G G G C A T C T G G T T C T T A C C T C T C G G G C C A T A C A A C C T A A A A T C G C C C A [480]
Hap 2 . . . . . G [480]
Hap 3 . . . . . G [480]
Hap 4 . . . . . G T . . . . . [480]
Hap 5 . . . . . G T . . . . . [480]
Hap 6 . . . . . G [480]
Hap 7 . . . . . G [480]
Hap 8 . . . . . G [480]
Hap 9 . . . . . G [480]
Hap 10 . . . . . A [480]
Hap 11 . A C . A . . . . . G [480]
Hap 12 . . . . . G T G T . . . . . [480]
Hap 13 . . . . . G T G T . . . . . [480]
Hap 14 . A . . . . . G T G T . . . . . [480]
Hap 15 . . . . . G T G T . . . . . [480]
Hap 16 . . . . . G T G T . . . . . [480]
Hap 17 . . . . . G - T G . . . . . [480]
Hap 18 . . . . . G T G T . . . . . [480]
Hap 19 . A . C A . . . . . A T G T . . . . . [480]
Hap 20 . . . . . G T G T . . . . . [480]
Hap 21 . . . . . G T G T . . . . . [480]
Hap 22 . . . . . G T G T . . . . . [480]
Hap 23 . . . . . G T G T . . . . . [480]
Hap 24 . . . . . G T G T . . . . . [480]
Hap 25 . . . . . G T G T . . . . . [480]
Hap 26 . . . . . G . T G T . . . . . [480]
Hap 27 . . . . . G A T G T . . . . . [480]
Hap 28 . . . . . G T G T . . . . . [480]
Hap 29 . . . . . - G T . T . . . . . [480]
Hap 30 . . . . . G T G T . . . . . [480]
Hap 31 . . . . . T A T G T . . . . . [480]
Hap 32 . . . . . T G T G T . . . . . [480]
Hap 33 . . . . . G C . . . . . [480]
Hap 34 . . . . . G - A T G C . . . . . [480]


Hap 1 T A C G T T C C C C T T A A A T A A G A C A T C A C G A T G G - A T C A C A [518]
Hap 2 [518]
Hap 3 [518]
Hap 4 [518]
Hap 5 [518]
Hap 6 [518]
Hap 7 [518]
Hap 8 [518]
Hap 9 [518]
Hap 10 [518]
Hap 11 [518]
Hap 12 [518]
Hap 13 [518]
Hap 14 [518]
Hap 15 [518]
Hap 16 [518]
Hap 17 [518]
Hap 18 [518]
Hap 19 [518]
Hap 20 [518]
Hap 21 [518]
Hap 22 [518]
Hap 23 [518]
Hap 24 [518]
Hap 25 [518]
Hap 26 [518]
Hap 27 [518]
Hap 28 [518]
Hap 29 [518]
Hap 30 [518]
Hap 31 [518]
Hap 32 [518]
Hap 33 [518]
Hap 34 [518]

```
HLADRB101:01:01  CACGTTTCTTGTGGCAGCTTAAGTTTGAATGTCATTT CTTCAATGGGACGGAGCGGGTGCGGTTGCTGGAAAGATGCA [ 78]
HymoDRBW094:01   ...........GA....GC....GC...G........ ......C..................C.C.....G...CA.T [ 78]
HymoDRBW094:02   ...........GA....GC....GC...G........ ......C..................C.C.....G...CA.T [ 78]
HymoDRBW095:01   ...........GA....GC.....C...G........ ......C....................C.....C....A.. [ 78]
HymoDRBW096:01   ...........GA....GG....GC...G........ ......C...................AC.....G....A.. [ 78]
HymoDRBW096:02   ...........GA....GG....GC...G........ ......C...................AC.....G....A.. [ 78]
HymoDRB104:01    ...........GA....G....ACA...G........ ......C....................C.....C....A.T [ 78]
HymoDRB104:02    ...........GA....G....ACA...G........ ......C....................C.....C....A.T [ 78]
HymoDRB104:03    ...........GA....G....ACA...G........ ......C....................C.....C....A.T [ 78]
HymoDRB104:04    ...........GA....G....ACA...G........ ......C....................C.....C....A.T [ 78]
HymoDRBW097:01   ...........GA....G......C...G........ ......C....T...............C.....G....A.. [ 78]
HymoDRBW100:01   ...........GA....G......A...G........ ......C................A.C.......G...C..G [ 78]
HymoDRBW098:01   ...........GA....GC.....C...G........ ...........................C.....G....A.T [ 78]
HymoDRBW098:02   ...........GA.........T.C...G........ ...........................C.....G....A.T [ 78]
HymoDRBW099:01   ...........GA....GC.....C...G........ ...........................C.....G....A.T [ 78]

HLADRB101:01:01  TCTATAACCAAGAGGAGTCCGTGCGCTTCGACAGCGA CGTGGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGGCCTG [156]
HymoDRBW094:01   ..................T.................. ......................................... [156]
HymoDRBW094:02   ..................T.................. ......................................... [156]
HymoDRBW095:01   ..C..............AT.C................ ......................................... [156]
HymoDRBW096:01   ..C...............T.................. ......................................... [156]
HymoDRBW096:02   ..C...............T.................. ......................................... [156]
HymoDRB104:01    ..................A.................. ......................................... [156]
HymoDRB104:02    ..................A.................. ........................................A [156]
HymoDRB104:03    ..................A.................. ......................................... [156]
HymoDRB104:04    ..................A.................. ......................................... [156]
HymoDRBW097:01   ..C...............T.................. ......................................... [156]
HymoDRBW100:01   ..C...............A..C............... ..................A...................... [156]
HymoDRBW098:01   ..........G.......A.................. .............T..........................A [156]
HymoDRBW098:02   ..........G.......A.................. .............T..........................A [156]
HymoDRBW099:01   ..........G.......A.................. ......................................... [156]
```


HLADRB101:01:01 ATGCCGAGTACTGGAACAGCCAGAAGGACCTCCTGGAGCAGAGAGGCGGGGCCGCGGTGGACACCTACTGCAGACACAACT [234]
HymoDRBW094:01 ..A.....T.....A.....AT.....A..... [234]
HymoDRBW094:02 ..A.....T.....A.....AT..... [234]
HymoDRBW095:01 ..A.....T.....A.....AT..... [234]
HymoDRBW096:01 ..A.....G..GC.....AT..... [234]
HymoDRBW096:02 ..A.....A.....AT..... [234]
HymoDRB104:01 ..C.....A.AA.....GCG.....AT..... [234]
HymoDRB104:02 GC.....T.....A.AA.....GCG.....AT..... [234]
HymoDRB104:03 ..C.....G..GC.....AT..... [234]
HymoDRB104:04 ..A.....T.....AT..... [234]
HymoDRBW097:01 ..A.....G..A.....GCA.....AT..... [234]
HymoDRBW100:01 .....A.....A.....GT.. [234]
HymoDRBW098:01 GC...A.....AT.....A.AA.....GCG.....AT..... [234]
HymoDRBW098:02 GC...A.....AT.....A.....GCA.....AT..... [234]
HymoDRBW099:01 ..A.....T.....A.AA.....GCG.....AT..... [234]
HLADRB101:01:01 ACGGGGTTGGTGAGAGCTTCACAGTGCAGCGGCGAG [270]
HymoDRBW094:01 ...T.G..... [270]
HymoDRBW094:02 ..... [270]
HymoDRBW095:01 ..TT..... [270]
HymoDRBW096:01 ..... [270]
HymoDRBW096:02 ...TG..... [270]
HymoDRB104:01 ...T.G..... [270]
HymoDRB104:02 ..... [270]
HymoDRB104:03 ..... [270]
HymoDRB104:04 ..... [270]
HymoDRBW097:01 G..T.G..... [270]
HymoDRBW100:01 ..... [270]
HymoDRBW098:01 .....G..... [270]
HymoDRBW098:02 ..... [270]
HymoDRBW099:01 .....G..... [270]

```
HLADRB101:01:01  H V  S C  G S L  S L  N V     I S S M  G R  S G C G  C W  K D A  S I T K  R S P C  A S T A  T W  G S T G  R *R S W  G G L M  P S T G  T A R R  T S W  S R G G  P R W T  P T A D  T T T G  [80]
HymoDRBW094:01   . . .  W S R .  R .  S .     . . . T  . . . .  . S .  R . T  . . . .  . S .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . . . . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW094:02   . . .  W S R .  R .  S .     . . . T  . . . .  . S .  R . T  . . . .  . S .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . . . . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW095:01   . . .  W S R .  . .  S .     . . . T  . . . .  . S .  T . T  . . . .  R S .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . . . . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW096:01   . . .  W S R V  R .  S .     . . . T  . . . .  . T .  R . T  . . . .  . S .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  R . R . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW096:02   . . .  W S R V  R .  S .     . . . T  . . . .  . T .  R . T  . . . .  . S .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . . S . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRB104:01    . . .  W S R .  N M  S .     . . . T  . . . .  . S .  T . T  . . . .  . T .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . K S . A G . .  . . . . . I . .  . . . . . .  [80]
HymoDRB104:02    . . .  W S R .  N M  S .     . . . T  . . . .  . S .  T . T  . . . .  . T .  . . . . . . . .  . . . . . . . .  . . . . . . . .  A . . . . . . .  . K S . A G . .  . . . . . I . .  . . . . . .  [80]
HymoDRB104:03    . . .  W S R .  N M  S .     . . . T  . . . .  . S .  T . T  . . . .  . T .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . . . . . . . .  R . R . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRB104:04    . . .  W S R .  N M  S .     . . . T  . . . .  . S .  T . T  . . . .  . T .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . . . . . . . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW097:01   . . .  W S R .  . .  S .     . . . T  . W . .  . S .  R . T  . . . .  . S .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . G S . A . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW100:01   . . .  W S R .  . M  S .     . . . T  . . . .  S . .  R . .  . . . .  . T R  . . . . . . . .  Q . . . . . . .  . . . . . . . .  . . . . . . . .  . . . . . . C  . . . . . .  [80]
HymoDRBW098:01   . . .  W S R .  . .  S .     . . .  . . . .  . S .  R . T  . . R .  . T .  . . . . . . . .  . . . . . . . .  A . . . . . . .  . K S . A G . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW098:02   . . .  W S .  . I  S .     . . .  . . . .  . S .  R . T  . . R .  . T .  . . . . . . . .  . . . . . . . .  A . . . . . . .  . . S . A . . .  . . . . . I . .  . . . . . .  [80]
HymoDRBW099:01   . . .  W S R .  . .  S .     . . .  . . . .  . S .  R . T  . . R .  . T .  . . . . . . . .  . . . . . . . .  . . . . . . . .  . R . . . . . .  . K S . A G . .  . . . . . I . .  . . . . . .  [80]

HLADRB101:01:01  L V  R A  S Q  C S  G E  [90]
HymoDRBW094:01   . W  . .  . .  . .  . .  [90]
HymoDRBW094:02   . .  . .  . .  . .  . .  [90]
HymoDRBW095:01   F L  . .  . .  . .  . .  [90]
HymoDRBW096:01   . .  . .  . .  . .  . .  [90]
HymoDRBW096:02   . W  . .  . .  . .  . .  [90]
HymoDRB104:01    . W  . .  . .  . .  . .  [90]
HymoDRB104:02    . .  . .  . .  . .  . .  [90]
HymoDRB104:03    . .  . .  . .  . .  . .  [90]
HymoDRB104:04    . .  . .  . .  . .  . .  [90]
HymoDRBW097:01   V W  . .  . .  . .  . .  [90]
HymoDRBW100:01   . .  . .  . .  . .  . .  [90]
HymoDRBW098:01   . .  . .  R .  . .  . .  [90]
HymoDRBW098:02   . .  . .  . .  . .  . .  [90]
HymoDRBW099:01   . .  . .  R .  . .  . .  [90]
```