# BALANCING SELECTION AT THE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC): SEQUENCE DIVERSITY AND INBREEDING AVOIDANCE

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Amy Louise Sherborne

June, 2008

# Abstract

The major histocompatibility complex (MHC) is a much studied region of the mammalian genome because of its uniquely high level of polymorphism. Both natural and sexual selection have been implicated in the maintenance of MHC diversity. The elevated heterozygosity observed at the MHC could reflect overdominant heterozygote advantage against pathogenic infection. Equally, MHC disassortative mating would lead to an excess of heterozygotes.

Balancing selection acting on the MHC was detected in a natural population of house mice (*Mus musculus domesticus*). Allele frequencies at the MHC deviated from those expected under neutral evolution, whereas allele frequencies of genome-wide markers did not. The genome-wide markers acted as a control, indicating that the deviation from neutrality seen at the MHC was not due to demographic effects in the population. Sequencing of a class I gene (H2-K1) and a class II gene (H2-Ab) and analysis of the rates of synonymous and nonsynonymous substitution revealed that the selective pressure was focused on codons encoding the antigen binding site of MHC molecules.

There was no evidence of an effect of the MHC on mate choice in mice in a number of experiments. Females showed no bias towards either the urine of MHC heterozygous males or that of MHC homozygous males. There was no difference in the reproductive success of males who were heterozygous or homozygous at the MHC. There was no evidence of a role for the MHC in the avoidance of inbreeding in a large enclosure experiment that allowed wild-derived mice free mate choice. Instead, the deficit in the number of matings between full-sibling mice (from expected under random mating) could be fully explained by a reduction in successful matings between mice that shared both major urinary protein (MUP) haplotypes.

MUPs are similar to MHC molecules in that they are highly polymorphic, ligand binding molecules that influence urine odour in rodents. The only known role of MUPs is in scent communication, and they have been implicated as sex-specific markers, indicators of dominance and social status, and more latterly as signals used in individual recognition. The effect of MUPs on sexual selection was compared to that of the MHC. However, there was no influence of heterozygosity at MUP or MHC on reproductive success in males. Instead, genome-wide heterozygosity and weight had significant effects on male reproductive success.

I tested the effects of MHC genotype on the parasite load of wild-derived mice experimentally infected with the gastrointestinal nematode *Trichuris muris*. All mice cleared the worm 35 days post-infection. Levels of parasite-specific immunoglobulin G in the sera of infected individuals was not correlated with heterozygosity at the MHC or with particular alleles. Females showed a significant preference for the urine of males who were sham-infected over those who were infected with the parasite.

Inbred laboratory strains of mice have been used in many experiments investigating resistance to pathogens and mate choice. The near 100% homozygosity and unnatural ancestry of inbred strains may explain why the results in this study which uses wild-derived mice whenever possible, differ from previous studies using inbred strains. The mean pairwise genetic distance at two MHC alleles was calculated for a natural population of mice and compared to the genetic distances between inbred strains used in several studies. Although there was no significant difference in the mean genetic differences, pairwise comparisons between some inbred strains were significantly lower than the mean of a natural population, suggesting that their use in parasite and mate choice trials may not be representative of the MHC in the wild.

# Acknowledgments

First and foremost I would like to express my utmost gratitude to my supervisors Dr. Steve Paterson and Professor Jane Hurst for their advice during my PhD. They have tirelessly provided direction, guidance, support and a kick up the arse when necessary. Micheal Thom, Sarah Cheetham and John Waters at Leahurst also provided invaluable assistance and advice.

Thanks go also to coffee time, with its often much needed caffeine and biscuit supply, but perhaps most importantly its inane chatter. Although the best biscuit for dunking was never agreed on, it was fun trying to decide. So Sarah, Wol, Chris, Suzanne, Marcus, Big Dave, Phill, and Dave, make sure 10.30 is coffee time where ever you are.

My year group of Kath, Jan, Hannah and Steph have been there for the good times and the not so good times and I couldn't have asked for a better group of people to share it with. When I needed dragging kicking and screaming from my desk and to the pub I could always rely on Andy, Becci, Becky, Chris, Emilie, Lawrence, Sam, Tara and Tom. The members of Liverpool University Boat Club became friends as well as team mates, and although my captaincy probably put my hand in date back by about six months, I wouldn't have had it any other way.

Gordon's and Schweppes, I couldn't have done it without you.

Finally, my thanks to my Mum and Dad for their love and support throughout my PhD and always. They have given me every opportunity in life, and I hope I can make them proud by grabbing every single one of them with both hands.

# Contents

# Figures

# Tables

# Abbreviations

Abbreviations used in this thesis are as follows:

| | |
|---|---|
| AIDS | Acquired immunodeficiency syndrome |
| $\chi^2$ | Chi-squared |
| cM | centi Morgan |
| D | Tajima's test statistic |
| $\hat{d}$ | Number of nucleotide substitutions per site |
| $\bar{d}_N$ | The rate of nonsynonymous substitutions per potential nonsynonymous site |
| $\bar{d}_S$ | The rate of synonymous substitutions per potential synonymous site |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxyribonucleotides |
| DVD | Digital versatile disk |
| ELISA | Enzyme-linked immunosorbent assay |
| enc | Enclosure study (Chapter 3) |
| $F_{IS}$ | Within population fixation index |
| $H_E$ | Expected heterozygosity |
| HIV | Human immunodeficiency virus |
| $H_O$ | Observed heterozygosity |
| HWE | Hardy-Weinberg equilibrium |
| inf | Infection study (Chapter 2) |
| LB | Luria-Bertani |
| MHC | Major histocompatibility complex |
| MUP | Major urinary protein |
| OD | Optical density |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| RFID | Radio frequency identification |
| rpm | Revolutions per minute |
| $r_{xy}$ | Pairwise relatedness |
| S.E. | Standard error of the mean |
| S.M.M | Stepwise mutation model |
| SAP | Shrimp alkaline phosphatase |
| SD | Standard deviation |
| sib | sibling |
| UV | Ultraviolet |
| win | Winsford study (Chapter 4) |

# Chapter 1

# INTRODUCTION

## 1.1 THE MAJOR HISTOCOMPATIBILITY COMPLEX

The major histocompatibility complex (MHC) is a cluster of closely linked genes that encode molecules involved in antigen processing and presentation. MHC genes are highly diverse, allowing MHC molecules collectively to bind a vast array of structurally diverse antigenic peptides. Unlike immunoglobulins and T-cell receptors, MHC genes are encoded by conventional genes that do not undergo rearrangement or other somatic processes of structural change (Parham 2000). Instead, diversity at the MHC is inherited. The high level of genetic polymorphism of MHC genes has been well documented (Gotze et al. 1980a; Soper et al. 1988), with some of the more diverse genes (such as H2-D and H2-K in mice) having between 50 and 100 alleles (Duncan et al. 1979). Nucleotide diversity at the MHC is twice the average for the mammalian genome (Garrigan and Hedrick 2003).

The majority of the substitutions that make the MHC so diverse are located within the region that codes for the antigen binding site. MHC molecules present self and foreign antigen to T-cell receptors and natural killer cells. Class I molecules are found on all nucleated cells and bind endogenous peptides, alerting the immune system to cells that contain peptides that are not native to the host, for example, virally encoded proteins (Bjorkman et al. 1987; Garboczi et al. 1996). Class II molecules are found only on B-lymphocytes, macrophages and other antigen presenting cells, and they

bind and present exogenous antigens, such as bacterial peptides (Stern et al. 1994). The presentation of protein fragments is mediated by a specialized groove on the surface of the MHC molecule encoded by approximately 50 amino acids, known as the antigen binding site (Brown et al. 1988; Fremont et al. 1992; Zhu et al. 2003). The antigen binding site is located in the extracellular portion of the molecule, the $\alpha 1$ and $\alpha 2$ domains in class I and $\alpha 1$ and $\beta 1$ domains in class II (Figure 1.1).

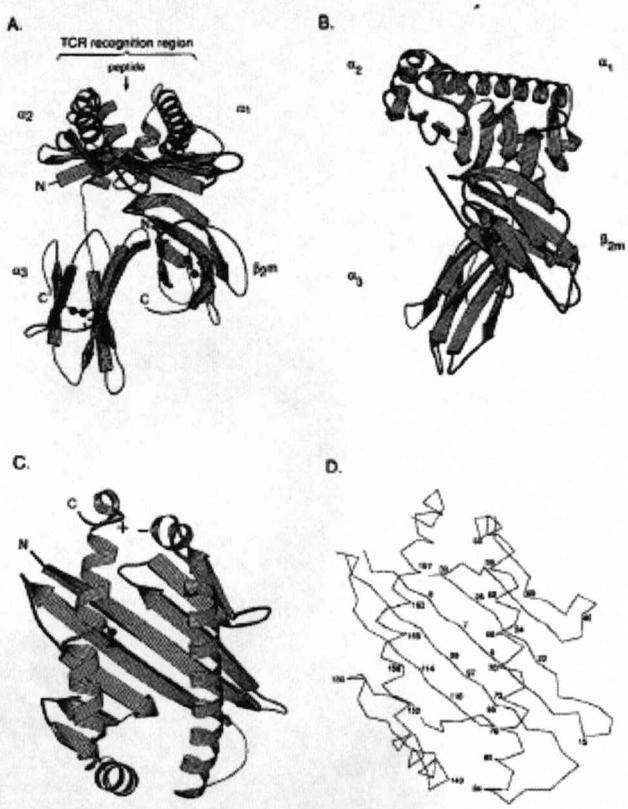



Figure 1.1 A. Structure of a murine MHC class I molecule, H2-A2, complexed with $\beta_2$-microglobulin. B. The molecule rotated 90° from the orientation in A. C. The $\alpha 1$ and $\alpha 2$ domains as viewed from the top of the molecule showing the peptide-binding site. D. Carbon- $\alpha$ atoms of the $\alpha 1$ and $\alpha 2$ domains with residue numbers. Taken from Bjorkman and Parham (1990).

## 1.2 SELECTION AT THE MHC

The selection pressures that maintain the high level of diversity at the MHC have been exhaustively studied, but not fully elucidated (reviewed by Potts and Wakeland 1993; and Apanius et al. 1997a). There is little doubt that the ongoing evolutionary battle against pathogens has generated polymorphism at MHC loci (Potts and Slev 1995a), however, there has also been recent support for the role of mate choice in the generation and maintenance of MHC diversity (Jordan and Bruford 1998). Both natural and sexual selection could drive polymorphism through balancing selection (Doherty and Zinkernagel 1975; Hedrick and Thomson 1983; Hughes and Nei 1988; Hughes and Nei 1989). Balancing selection works to maintain genetic polymorphism within a population and contrasts with directional selection which favours a single allele. Balancing selection at the MHC is characterised by a deficiency in homozygotes (Klein 1986), the maintenance of large numbers of alleles at equal frequencies (Hedrick and Thomson 1983; Klein 1986), and a high degree of similarity in alleles between species compared to within species comparisons (trans-species mode of polymorphism) (Klein 1980; Figueroa et al. 1988; Mcconnell et al. 1988).

## 1.3 PATHOGEN MEDIATED SELECTION

There are several mechanistic processes that can result in balancing selection. The overdominance theory (Doherty and Zinkernagel 1975) states that a heterozygote individual is at a selective advantage over either homozygote. In the case of natural selection at the MHC, heterozygotes express two alleles, giving them the ability to bind a greater repertoire of antigens and better protection against infection. This hypothesis has been

widely documented in the case of the gene coding for sickle cell anaemia (Allison 1964; Aidoo et al. 2002). It is postulated that overdominance is the mechanism that maintains the deleterious sickle cell allele within the population. More recently, there has been interest on the level of heterozygosity at the MHC and the progression through the stages of the viral infection HIV. Carrington et al. (1999) showed that no single human leukocyte antigen (HLA, the human MHC) conferred a beneficial effect, however individuals that were heterozygous at a large number of HLA loci showed a slower progression through the stages of the disease and an increased time from the contraction of the virus to death. Heterozygote advantage has also been documented in fish (Arkush et al. 2002; Wegner et al. 2003) and rodents (Penn et al. 2002).

An alternative to overdominance is frequency dependent selection, or rare allele advantage (Bodmer 1972). This hypothesis requires the co-evolution of host immunity and an endemic pathogen or pathogens. The fitness of an allele is inversely proportional to its frequency within the population; rare alleles are at a selective advantage as pathogens have not had time to adapt to evade detection by the new MHC antigen. This leads to the rare allele increasing in frequency over several generations. However, once the allele becomes dominant in the population, its selective advantage is lost and pathogen mediated selection reduces its frequency. Thus the frequency of a particular allele would be seen to cycle within a population, preventing it from going to fixation or being eliminated. Fitness of MHC alleles could also vary due to the frequency and intensity of pathogens within a population. The fluctuating selection model describes selection favouring different alleles dependent on the temporal and spatial pattern of pathogens. Many alleles could be maintained in a changing selective landscape.

## 1.3 MATE CHOICE

On the molecular level the MHC facilitates the distinction between self and non-self in the immune system. Thomas (1975) suggested that the MHC may play a similar role on the much larger scale of mate choice; the distinction between kin and non-kin. Mate choice is one of the most important factors influencing an individual's lifetime reproductive success. This choice can take different forms; the initial choice of mate; the choice of sperm post-copulation, termed cryptic choice; or the differential investment in zygotes or offspring (Kempenaers 2007a). In most species it is the female that is the choosy sex (Andersson 1994). Studies have shown that female mating decisions can be based on: physical characteristics, such as size (Reynolds and Gross 1992) and ornamentation (Petrie and Halliday 1994); behavioural traits (Hale and St Mary 2007); and the quality of the territory held by the potential mate (Leonard and Picman 1988). These characteristics can either indicate direct fitness benefits (for example a male in control of a good territory is likely to provide superior parental care) or can indicate that the potential mate has "good genes" that are likely to be passed to the progeny. In fact the two may be correlated, as males that possess good genes are likely to out-compete males with substandard genes. Good genes are alleles that increase fitness independently of their homologous allele or the rest of the genome (Moller and Alatalo 1999; Jennions et al. 2001).

The compatibility theory of mate choice states that females take into account their own genotype when choosing a mate (Trivers 1972; Zeh and Zeh 1996; Zeh and Zeh 1997). The compatibility theory predicts that mating with a male with the most dissimilar genotype to the female's will produce the most heterozygous offspring (Mays and Hill 2004). Heterozygosity has been shown to increase fitness independent of the

presence of any associated good genes (Mitton 1993). Such fitness advantages may be due to heterozygosity masking recessive deleterious alleles (dominance), or by heterozygote advantage, where the heterozygote is fitter than either homozygote (overdominance) (Comings and MacMurray 2000). A female mating to promote offspring heterozygosity is thus mating to promote offspring fitness. The most compatible male will be different for each female, so there is no universal "best" male, which could make it difficult to detect this type of mate choice in natural populations.

Female choice for genetic compatibility has been hard to reconcile with female choice for good genes (Mays and Hill 2004; Neff and Pitcher 2005). Extravagant male ornamentation can signal good genes that increase offspring fitness, notably in the case of antler size in deer and MHC genes (Ditchkoff et al. 2005). However, it has also been shown that females choose to mate with MHC dissimilar males to increase offspring heterozygosity and thus immunological fitness in their offspring (Potts et al. 1991). A choice for the most genetically dissimilar male, in most cases, will not be a choice for the most ornamented male. There are few studies that aim to resolve this conflict. Colegrave et al. (2002) suggested that in mating systems that include multiple matings, females may be making a pre-copulatory choice for good genes and post-copulatory mechanisms such as sperm competition result in greater offspring production by genetically dissimilar males. Roberts and Gosling (2003) showed that female mice prefer MHC dissimilar males only when indicators of good genes, in this case scent marking rate, are equivalent. This suggests that the choice for good genes is a stronger driver of mate choice than genetic dissimilarity in this system. Further work needs to be done to try and resolve the interactions between mate choice for good genes and that for genetic compatibility.

The compatibility theory of mate choice predicts that animals should avoid mating with close relatives (Kimura and Crow 1963). Matings between individuals that share a recent common ancestor will increase homozygosity in the progeny and can result in lower fitness of offspring due to dominance effects, and to some extent, a loss of heterosis (Charlesworth and Charlesworth 1999). Such a reduction in the fitness of offspring is termed inbreeding depression and has been well categorised in model organisms and in field studies (reviewed by Charlesworth and Charlesworth 1987; Pusey and Wolf 1996; Hedrick and Kalinowski 2000; Lehmann and Perrin 2003). Inbreeding avoidance is one of the driving forces behind the evolution of different mating systems: self incompatibility systems in plants (Charlesworth and Awadalla 1998), gender dimorphism (Ramsey et al. 2006), polyandry (Stockley et al. 1993), and dispersal (Moore and Ali 1984). Many mammals have complex behavioural traits that discourage inbreeding. Pusey and Wolf (1996) mention dispersal, extra-pair matings, reproductive suppression, and recognition and avoidance of kin as mates. These traits are often specific to the type of social and mating systems, for example in rodents recognition and avoidance of kin as mates is more developed in species that live in stable family groups than those that have male biased dispersal (Dewsbury 1988).

Inbreeding depression has received more attention in the literature than outbreeding depression (Pusey and Wolf 1996), possibly because it is more relevant in the conservation of zoo and natural populations. Consistent with this, there have been few studies investigating outbreeding avoidance. Outbreeding depression is a loss in fitness of the offspring of matings between genetically differentiated populations (Dobzhansky 1948). It is thought to be a result of the break up of coadapted gene complexes (Lynch 1991) or due to a lack of local adaptation in immigrants to a population

(Marr et al. 2002). Bateson (1978) suggested the theory of optimal outbreeding, i.e. a level of relatedness between mates that avoids the negative fitness consequences of both inbreeding depression and outbreeding depression. Hoogland (1992) found that prairie dogs did not mate with their parents, offspring, or full and half-siblings but regularly inbreed with more distant kin such as full and half-cousins. The most advantageous level of outcrossing is likely to vary with mating system and between species.

## 1.4 KIN RECOGNITION

For an individual to avoid both inbreeding and outbreeding, they must first be able to recognise kin. It is logical to think that animals that are reared in family groups would be able to recognise their litter mates (coefficient of relatedness, $r = 0.5$) throughout life. However, they are equally related individuals from the same offspring from a different litter. Also, inbreeding depression can be manifest in matings between more distant kin ($r < 0.25$) (Keller 1998), and there is evidence that some species can recognise unfamiliar kin on first encounter (Greenberg 1979). This suggests that, at least in some species, recognition of kin is not based upon familiarity to siblings, but has a genetic component. The process of recognition can be split into 3 stages: i) production of a signal, ii) perception of the signal and reference matching to an internal template to determine if the signal is favourable or unfavourable, and iii) an action in accordance with the information received (Sherman 1997). In this framework, the genetic component is the signal and the internal template that it is compared to could be a parental allele that was learnt during rearing (Buckle and Greenberg 1981; McGregor and Krebs 1982). Alternatively, and facilitating the recognition of unfamiliar kin, a self-referent template could

be used (Dawkins 1982; Mateo and Johnston 2000). In this case, a female would be able to determine when a potential mate has the same phenotype as herself at a particular marker. As close kin are more likely share alleles by descent than non-kin, a female should avoid mating with males that match their self-reference allele. It should be noted that avoiding mates that share alleles at a marker locus will increase heterozygosity at that marker in the progeny. The converse is also true, mating to maximise heterozygosity in offspring will tend to preclude mating with close kin as they are more likely to share alleles than the rest of the population (Karlin and Feldman 1968).

Few genetic markers have been suggested as kin recognition signals used for mate choice. Murine major urinary proteins are signals used for individual recognition (Cheetham et al. 2007) that may play a role in the inbreeding avoidance, and the t-complex in mice has received some attention (Williams and Lenington 1993), but by far the most investigated candidate is the MHC (Alberts and Ober 1993; Apanius et al. 1997b; Arcaro and Eklund 1998; Edwards and Hedrick 1998; Grob et al. 1998; Jordan and Bruford 1998; Penn and Potts 1999; Tregenza and Wedell 2000; Penn 2002; Zelano and Edwards 2002; Ziegler et al. 2005; Milinski 2006).

## 1.5 MATE CHOICE AND THE MHC

Most studies investigating the role of the MHC in mate choice have used inbred house mouse strains (although see Piertney and Oliver 2006 for a review of MHC studies in non-model species). It has been shown that inbred mice can be trained to distinguish between MHC determined urinary odour types (Yamazaki et al. 1979; Yamaguchi et al. 1981; Bard et

al. 2000) and Penn and Potts (1998b) used habituation-dishabituation tests to show that mice could distinguish between MHC congenic strains without training. However, these studies only show a preference for an odour, or the ability to detect a change in an odour. They do not show that mice use MHC based odour cues to make mating choices. The MHC-mate choice theory gained support when Yamazaki et al. (1976) showed that male mice had MHC determined mating preferences. Of the six inbred strains tested, four preferred to mate with females with MHC haplotypes different to their own, one preferred to mate with MHC similar females and the sixth showed no preference. Further studies showed that cross-fostering new born mice onto a dissimilar strain reversed MHC mating bias (Beauchamp et al. 1988; Yamazaki et al. 1988) suggesting that the mice were learning the MHC type representing close kin during rearing. In these studies, female preference showed less of a trend, which was unexpected as females usually have the stronger preference in a species such as mice where they invest more resources in their young (Trivers 1972). Egid (1989) and Eklund (1991) also tested female preference for males of differing MHC congenic strains and vice versa. They found that females preferred dissimilar males and males showed no preference, in contrast to previous results. When considering all experiments using inbred strains, preferences seem to vary between sexes and between strains and no firm conclusion can be drawn beyond the fact that mice are able to distinguish MHC determined odour cues. These ambiguous results may reflect the use of inbred strains which have been forced to mate with genetically identical mates for thousands of generations, resulting in unnatural behavioural patterns (Van Oortmeerssen 1971) and presumably, mating choices (Levine et al. 1966).

Potts has carried out a number of mate choice trials in large pens using semi-wild mice to try and overcome the problems associated with using

inbred strains (Potts et al. 1991; Potts et al. 1992; Potts et al. 1994). Mice were derived from wild populations but were bred to carry one of four MHC haplotypes from inbred strains. Nine experimental populations were set up and mice were allowed to breed freely until ~150 offspring were born. The progeny were typed at the MHC and the observed number of homozygotes compared to the expected number predicted by random mating. They attributed a significant homozygous deficiency of 27% to MHC disassortative matings. Unlike previous work, Potts' study demonstrates evidence of mate choice rather than odour preference. However, what genetic cue these choices are based upon is not established. In Potts' breeding design, as in natural populations, MHC is correlated with the rest of the genome. Animals that share MHC haplotypes are likely to share many other alleles across the genome. Thus the deficit in MHC homozygotes seen may be attributable to the avoidance of a correlated gene. The mice used in the study were F3 - F6 progeny from a cross between wild mice and four inbred strains. In six of the nine populations the genomes of mice homozygous for the MHC were 50% wild derived and 50% derived from a single inbred strain, whereas the genomes of mice heterozygous at the MHC were 50% wild, 25% one inbred strain and 25% a second inbred strain. If females were making their mating decisions based on genome wide cues of heterozygosity they would tend to mate with males that had genetic contributions from three sources, i.e. those that were heterozygous at the MHC, leading to the MHC homozygous deficiency seen, but for different reasons than those suggested. Care needs to be taken in the design and interpretation of experiments to disentangle the effects of MHC heterozygosity and genome wide heterozygosity.

## 1.6 MAJOR URINARY PROTEINS

Major urinary proteins (MUPs) are an alternative genetic marker involved in kin recognition. MUP genes have much in common with the MHC. They are both highly polymorphic gene clusters (Cavaggioni and Mucignat-Caretta 2000; Beynon et al. 2002), and both gene products bind small molecules, the repertoire of which is dependent on the set of alleles expressed as amino acid substitutions in their binding sites influence the range of peptides bound. Whereas MHC molecules bind immunologically informative antigen, MUPs bind semiochemically informative ligands (Marie et al. 2001; Sharrow et al. 2002) and their only known role is in scent communication. MUPs have been shown to play a role in self/non-self recognition (Hurst et al. 2001), individual recognition (Cheetham et al. 2007) and the recent discoveries that male MUPs can prime ovulation (More 2006) and are up-regulated in females at the beginning of oestrus (Stopka et al. 2007) suggest that they play a central role in murine reproduction.

## 1.7 INBRED STRAINS OF MICE

Inbred mice, rats and other rodents make up 83% of the animals used in experimental procedures in Britain (Statistics of Scientific Procedures on Living Animals, Great Britain 2006). Inbred strains of mice are one of the most widely used model organisms in animal studies. Atchley and Fitch (1991) provide a history of the generation and genealogy of all common inbred strains. Mouse laboratory strains are known to have mixtures of ancestral genomes from different *Mus musculus* sub-species (Yoshiki and Moriwaki 2006), the main contributors being *M. musculus musculus*, *M. musculus domesticus*, and *M. musculus castaneus*. This would suggest

genetic diversity between inbred strains above that found in a natural population. However, most inbred lines are derived from a small number of founder animals, in fact phylogenetic analysis of mitochondrial DNA from inbred strains suggests that all common strains descend from a single *Mus musculus domesticus* female (Dai et al. 2005; Goios et al. 2007). Also, characterisation of MUP phenotypes of inbred strains revealed only two different expression patterns (one associated with mice derived from the same lineage as Balb/c, the other from the C57BL/6 lineage) (Robertson et al. 1996). Thus, although laboratory mice were bred from several different sub-species and there has been some addition of new genetic material and some mixing of existing strains, at least some regions of the genome are likely to be monomorphic between strains, as well as within strains. A more recent study comparing whole genomes of inbred strains describes a mosaic structure, with areas of low diversity where the strains show recent common origin, and areas of high diversity where ancestry can be traced back to the subspecies that generated inbred strains (Wade et al. 2002).

The generations of inbreeding used to create inbred lines means that it is not only the near 100% homozygosity that is distinctive of these animals. Phenotypic traits have been shown to vary widely between strain (Figure 1.2), as have behavioural traits (Van Oortmeerssen 1971), and they certainly do not reflect behaviour of wild mice.

Figure 1.2 Four different inbred strains of mice; showing phenotypic diversity in their coat colours. Taken from Callaway (2007).

The suitability of inbred strains for investigations into MHC determined recognition of pathogens must be questioned. Even more importantly, the use of mice that have been forcibly inbred for hundreds of generations in mate choice trails may not be valid.

## 1.8 AIMS

The work described in this thesis explores the evidence for balancing selection at the MHC, and the pressures maintaining that selection, in a natural population of house mice (*Mus musculus domesticus*). This is achieved through a number of studies;

### 1.8.1 Chapter 2

To assess the influence of the MHC on parasite resistance and thus the influence of parasites on selection at the MHC, wild mice of known MHC genotypes were infected with the parasitic gastrointestional nematode *Trichuris muris*. Their immune response was measured by determining

worms burdens 35 days post-infection and by quantifying levels of parasite-specific immunoglobulin G in the serum. To test the role of health status and susceptibility to pathogenic infection on sexual selection, female preference for uninfected males was tested in scent choice trials using urine collected from infected and sham-infected males. Scent choice trials were also used to test female preference for MHC heterozygous males.

### 1.8.2 Chapter 3

The role of MHC in the avoidance of inbreeding was tested using a large mate choice experiment. Wild-derived mice were allowed to mate freely in semi-natural enclosures. Parentage of offspring was determined to allow the effects of MHC and MUP on the avoidance of full-sibling matings to be observed. The reproductive success of males with differing levels of heterozygosity at MHC, MUP and background genes was tested in a second study. The more controlled set-up of this experiment allowed the effects of heterozygosity on reproductive success to be tested in the absence of other factors that affect mate choice such as dominance relationships and territorial status.

### 1.8.3 Chapter 4

Evidence of balancing selection at the MHC was tested for using a single population of wild house mice (*M. musculus domesticus*). MHC-linked markers allowed allelic frequencies to be determined and Watterson's test was used to investigate deviation from neutrality. Twelve genome-wide markers provided a control for recent demographic effects within the population. A class I gene (H2-K1) and a class II gene (H2-Ab) were sequenced for the same population. Rates of synonymous and nonsynonymous substitution were compared, and Tajima's D test statistic was used to further examine deviations from neutrality and evidence for balancing selection.

### 1.8.4 Chapter 5

The mean pairwise genetic distance between MHC alleles in a natural population of house mice was calculated. This was then compared to the genetic distances between the inbred strains of mice used in a number of studies, and the mean distances between MHC alleles in the enclosure study in Chapter 3. These comparisons allowed conclusions to be drawn regarding the validity of using laboratory mice in studies investigating parasite resistance and mate choice.

# Chapter 2

# MHC DIVERSITY, RESISTANCE TO *TRICHURIS MURIS*, AND FEMALE ODOUR PREFERENCE

## 2.1 INTRODUCTION

Parasite driven balancing selection is one of the hypothesised mechanisms maintaining high levels of allelic diversity at the MHC (Apanius et al. 1997a; Edwards and Hedrick 1998; Hughes and Yeager 1998; Meyer and Thomson 2001; Penn 2002). This type of selection can act through rare allele advantage, or heterozygote advantage. Heterozygosity at the MHC has been linked to disease resistance in several studies. Wegner (2003) showed that stickleback MHC (and not genome wide) heterozygosity could be linked with infection levels in the wild. MHC heterozygous Chinook salmon demonstrate a lower infection rate when challenged with multiple parasites (Arkush et al. 2002), as do mice (Penn et al. 2002), and patients infected with HIV show a slower progression to AIDS if they are heterozygous at the MHC (Carrington et al. 1999). MHC molecules bind foreign antigens and present them to cells of the immune system, activating an immune response. Small changes in the binding site of an MHC molecule can vastly change the pool of antigen that can be bound (Nathenson et al. 1986). Individuals heterozygous for a particular MHC gene can thus bind a greater range of antigenic peptides than homozygous individuals making them more able to recognise and mount an immune response against parasites. A heterozygote having a fitness advantage over

either homozygote is known as the overdominance theory of balancing selection.

MHC-based mate choice is another mechanism that could contribute to the maintenance of high levels of MHC diversity in natural populations (Potts and Wakeland 1993; Brown 1998; Penn 2002; Roberts et al. 2005). MHC disassortative mating (Penn and Potts 1999) facilitates the avoidance of inbreeding, as animals are more likely to share alleles with close kin than unrelated individuals. Females mating with individuals with MHC types different to their own would produce offspring with higher levels of MHC and genome wide heterozygosity, with potentially better immuno-competence and overall fitness. Conversely, females may be using MHC heterozygosity as a genetic reporter of genome wide heterozygosity. Thus mating with an MHC heterozygous male would mean mating with a male possessing fitness benefits associated with genome wide heterozygosity (Mitton 1993; David 1998; Hansson and Westerberg 2002). The difference between mating to avoid inbreeding and increase offspring heterozygosity, and mating with the most heterozygous male is further discussed in chapter 3. In this study I assess female preference for males heterozygous at the MHC to detect use of the MHC as a marker of genome-wide heterozygosity and associated fitness.

The avoidance of parasitised males is another component of female mate choice that should result in offspring having better immune genes (Hamilton and Zuk 1982; Zuk 1992; Ehman and Scott 2002). A direct benefit of avoiding contact with infected cohorts is a reduced risk of getting infected, but avoiding mating with males who are susceptible to disease also reduces the risk of passing on genes associated with susceptibility to offspring. It is unclear whether females choose not to mate with infected males because they are infected per se, or because they

recognise poor quality immune genes in males that are susceptible to parasites (see Penn and Potts 1998a for a review of studies).

The whipworm parasite of mice, *Trichuris muris*, has been used as a model system for gastointestinal nematode of infections in humans (see Bancroft et al. 2001; Cliffe and Grencis 2004; Hayes et al. 2004 for reviews). As a result *T. muris* infections have been well characterised in laboratory strains of mice. Resistance to the parasite has been shown to have a MHC determined component (Else and Wakelin 1988), making it a possible driver of MHC polymorphism in the wild and an appropriate parasite for use in studies investigating the effects of different MHC haplotypes. The ability of a host to expel *T. muris* can be measured by the worm burden 35 days post infection and the level of parasite specific immunoglobulin G (IgG) in the serum. The level of IgG production has been shown to influence the speed of clearance of *T. muris* (Else and Grencis 1991; Koyama et al. 1999) and when transferred to normally susceptible strains IgG alone can confer resistance (Else et al. 1990b).

Mating decisions can be influenced by many conflicting factors and can be hard to measure in field and laboratory situations alike (Wagner 1998). In rodents, the olfactory system plays a major role in social and mating behaviour (Wyatt 2003). Male mice have volatile compounds in their urine which can signal sex and social status (Lin et al. 2005). Mice are able to distinguish MHC determined odour cues in urine (Yamazaki et al. 1979; Yamaguchi et al. 1981; Penn and Potts 1998b; Bard et al. 2000; Zufall and Leinders-Zufall 2007) and can discriminate between the cues of uninfected individuals and individuals infected with the intestinal nematode *Heligmosomoides polygyrus* (Kavaliers and Colwell 1995b; Kavaliers and Colwell 1995a; Kavaliers et al. 1998; Ehman and Scott 2002; Kavaliers et al. 2003; Kavaliers et al. 2005); the protozoan *Eimeria vermiformis* (Kavaliers and Colwell 1993); and mouse mammary tumour virus

(Yamazaki et al. 2002). Male urine can be used in trials testing female preference for infection status and MHC type independent of other factors that may influence female behaviour in trials that allow direct contact. Visual cues such as size and general condition are likely to affect female mating decisions, so using odour alone removes these confounding variables.

## 2.2 AIMS

This study investigated the susceptibility of wild *Mus musculus domesticus* to the parasitic gastrointestinal nematode *Trichuris muris*. The immune response to the parasite was measured by quantifying the levels of immunoglobulin G in the serum and determining worm burdens 35 days post infection. Immune responses between MHC homozygotes and MHC heterozygotes, as well as between specific MHC genotypes were compared. Female preference for uninfected males and MHC heterozygous males was tested in scent choice trials using urine collected from infected and sham-infected males.

## 2.3 MATERIALS AND METHODS

### 2.3.1 Infected males

Subjects were captive bred adult male *Mus musculus domesticus* from a laboratory maintained colony derived from wild ancestors captured from five different populations in the northwest of England, UK. To obtain MHC homozygous individuals we had to use a degree of inbreeding in our breeding design. The sires and dams used to breed subject males had previously been used in a large outdoor enclosure experiment (chapter 3).

They were either paternal half-siblings or full-siblings and were typed at microsatellite markers within and surrounding the MHC. Breeding pairs were set up to generate offspring with three MHC homozygous types and two MHC heterozygous types (termed gg, hh, jj, gh and gj). Sixty-nine males from 12 family groups were used. Males were singly housed in M3 cages (40 × 11.5 × 12.5 cm, North Kent Plastics, UK) from 8 - 10 weeks post partum and were infected when 11 - 19 weeks old. Genetic samples (1 - 2 mm removed from the tip of the tail under general anaesthesia (halothane)) were taken from all males and were typed using MHC linked microsatellite markers.

### 2.3.2 MHC genotyping

I extracted DNA from tail snips using mouse-tail extraction kits (Promega Wizard SV Genomic DNA Purification System). As the MHC haplotypes had previously been characterised using a suite of microsatellite markers within the region (see 3.3.1.2 MHC and MUP genotyping, for details), only two markers were required to distinguish between the three different haplotypes (Appendix Table 3.2). PCR amplifications were conducted in 10 μl reactions containing 10ng of DNA, 0.5μM of each primer and 5.0μl of 2X BioMix Red reaction mix (Bioline, London, UK). Plates were incubated in a thermal cycler (Peltier) at 95°C for 2 min, followed by 30 cycles of 95°C for 30s, 52°C-58°C (depending on primer) for 2 min, then 72°C for 30s, with a final extension at 72°C for 10 min. The PCR reactions were then diluted 50- to 100- fold, multiplexed and 1 μl of the mix was added to 8.8 μl of formamide and 0.2μl GeneScan LIZ500 size standard (Applied Biosystems). Fragment size was determined with an ABI PRISM 3100 DNA analyzer (Applied Biosystems) and GeneMapper v3.0 software designed for microsatellite analysis (Applied Biosystems).

### 2.3.3 *Trichuris muris* infection

Infectious *T. muris* eggs were obtained from Dr Kathryn Else (University of Manchester). Stock infections were maintained in female nude (nu/nu) mice (Harlan-Olac, UK). Mice were culled between day 35 and day 45 post infection and their digestive tracts removed by dissection. Live adult worms were removed from the caecae and colon with fine forceps. Worms were cultured in Petri dishes for 4 hours and overnight in 10 ml RPMI 1640 (Gibco) with 500 μg/ml penicillin and 500 μg/ml streptomycin (Invitrogen) at 37°C. Eggs were separated from the 4 hour and overnight supernatants by centrifuging at 2000 rpm for 10 minutes. Eggs were re-suspended in ultra pure water in 1.5 ml Eppendorf tubes and left in the dark for 6 weeks at room temperature to allow the eggs to embryonate.

When needed for infection, eggs were centrifuged at 2000 rpm for 10 minutes then re-suspended in ultra pure water. The egg suspension was mixed using a whirlimixer and then the number of embryonated eggs in a 50 μl aliquot counted under x40 magnification. The volume of the egg suspension was adjusted to give 100 infective eggs per 100 μl. Male mice were infected by oral gavage with 100 μl of the egg suspension, using a blunt-tip needle. Of the 69 mice, 39 were infected with *T. muris* and 30 were given sham infections with pure water. MHC types were taken into account to ensure there were both infected and uninfected males of every possible genotype. Mice were monitored at regular intervals during the infection period but all were asymptomatic. Urine was collected from all males at day 30 post infection (± 1 day) for use in odour preference trials (section 2.3.5) by holding the individual by the scruff of the neck over a clean 1.5 ml Eppendorf tube. Urine was stored at -18°C until use.

Thirty five days post infection mice were culled. Blood was collected by cardiac puncture and centrifuged at 14000 rpm for 10 minutes to separate

the serum supernatant. Worm burdens were assessed using the method described by Else and Wakelin (1990). At autopsy, the caecum and colon were removed and slit open in a petri dish of water. Gut contents were shaken away from the gut tissue and the mucosa inspected for the presence of second, third and fourth larval stages and adult worms.

### 2.3.4 Serum IgG measurements

The parasite-specific sera IgG1 levels of infected mice were quantified using ELISAs. Samples of 11 uninfected mice were also tested to assess the background level of IgG in mice not mounting an immune response. 96-well plates (Greiner Bio-One, UK) were coated in 100 μl of carbonate bicarbonate buffer with 5 μg/ml parasitic adult worm antigen, obtained by homogenising male and female worms in phosphate buffered saline (PBS), and left overnight at 4°C. Plates were washed with PBS, and then blocked with PBS containing 0.5% bovine serum albumin and 0.05% Tween-20 for 1 hour at room temperature. After washing, sera samples were added in a doubling dilution in PBS Tween-20 of 1 : 8 to 1 : 1012 and incubated at room temperature for 2 hours. After washing, the detection antibody (IgG1 goat anti-mouse horse radish peroxidase conjugate, Autogen Bioclear, UK) was added at a concentration of 3 μg/ml and incubated at room temperature for 2 hours. After washing, *o*-Phenylenediamine dihydrochloride (Sigma-Aldrich, Sigma, St Louis, MO, USA) at 0.4 mg/ml in phosphate citrate buffer with 0.004% hydrogen peroxide was used as the substrate for the detection of peroxidase activity. Plates were incubated for 10 minutes in the dark before stopping the reaction with 3M sulphuric acid. Optical density (OD) was read at 490 nm using a microtiter plate reader. OD values were compared to standard curves generated in each plate using purified subclass specific IgG1 protein (AbD Serotec, UK) instead of parasitic antigen, and coating goat anti-mouse polyclonal IgG1 antibody

(Autogen Bioclear, UK) at doubling dilutions from 0.3125 μg/ml to 0.00244 μg/ml instead of sera samples.

2.3.5 Odour preference

Subjects were captive bred adult female *Mus musculus domesticus* from the same laboratory maintained colony as the males, but were always from separate breeding lines and thus were not closely related. At weaning females were transferred in single-sex family groups in MB1 cages (40 × 23.5 × 12.5 cm, North Kent Plastics, UK) maintained at 19 - 23 °C, relative humidity 50 - 60%, under reverse 12:12 light cycle with the lights off at 0900. Mice had Corn Cob Absorb 10/14 substrate with paper wool nest material and ad libitum access to water and food (Lab Diet 5002 Certified Rodent Diet, Purina Mills, St. Louis, MO, USA). Cardboard tubes were provided periodically for additional environmental enrichment. All tests were carried out during the first 6 hours of the dark phase under dim red light.

Twelve females (at least 16 weeks old: age range 16 - 76 weeks) took part in four different trial types, each no less than 10 days apart. Soiled nest material and substrate from unfamiliar males was placed in the female's home cage three days prior to the preference test to induce oestrus during the trial (Marsden and Bronson 1964; Cheetham et al. 2007). To prevent bias due to the recognition of familiar odours during the trials we ensured the urine used to prime the females was from males unrelated to the test subjects by using bedding and nesting material from a mixture of males from different inbred strains (C57BL/6, BALB/c and F1 crosses of the two). Placing soiled bedding material in the females' cages ensures they have contact with urinary volatiles and the non-volatile fraction of the urine. The vomeronasal organ has been shown to play an important role in priming female attraction (Moncho-Bogani et al. 2004; Moncho-Bogani et

al. 2005; Keller et al. 2006; Pankevich et al. 2006; Martel and Baum 2007; Yoshikage et al. 2007; Lanuza et al. 2008; Ramm 2008). All trials were conducted in the base of a clean MB1 cage, with a clear Perspex lid. Females were transferred into individual test cages and allowed to settle for 30 min before the start of the trial. In the first trial female preference for male scent over female scent was tested to ensure the females were behaving as expected of a sexually mature female. Females were presented with male urine (pooled from C57BL/6 males, $n \geq 5$) at one end of the cage and female urine (pooled from C57BL/6 females, $n \geq 5$) at the other end, approximately 25cm apart and equidistant from the cage walls. The urine was presented as 10µl streaked directly onto the underside of the Perspex lid. The position of male and female urine was randomised, and in both the test and subsequent analysis of behaviour the experimenter was blind to each odour source. In the second trial females were presented with urine collected from a male on day 30 post infection with *T. muris* and urine from a male that had been sham infected, again in random positions. Urine pairs in each trial were matched for both MHC heterozygosity and genotype, but over the 12 trials males from all MHC types were used. In the third trial females were presented with urine from a MHC homozygous male and a MHC heterozygous male. Urine pairs were matched for male infection status ($n = 6$ uninfected, $n = 6$ infected).

Female's response to different urine types was remotely recorded to DVD for a 10 minute period using a ceiling mounted camera above the Plexiglas cage lid. I measured two types of behaviour which indicate "preference" for an odour. Time spent directly investigating the scent by sniffing at the cage lid on which the urine was streaked represents the female's interest in gaining further information from the scent. To assess greater attraction to one scent compared to another, the scent sources were deliberately located away from the side walls, in areas where mice normally spend little time.

Thus time spent under or near (defined as the subject's nose being within a 5 cm diameter circle) the urine source when not sniffing it reflects an attraction to spend time with the scent source. To ensure that the females do not naturally spend time in the area where the stimulus is located (due to the layout or markings of the cage) I carried out a control trial with all subject females using water instead of both stimuli. Time spent in direct or indirect contact with the water was recorded to represent female behaviour in the absence of a stimulus.

Differences in time spent sniffing urine and time near urine while not sniffing were initially analysed using Wilcoxon signed rank tests as the duration of female response was not normally distributed. The data from the trials testing male infectivity status and MHC heterozygosity were then combined in a log ratio test,

$$log\ ratio\ =\ \frac{log\,(t_{m2}-t_{m1})}{t_{m2}-t_{m1}}$$

where $t_{m1}$ = time spent sniffing, or near but not sniffing, the urine of the hypothesised preferred male, i.e., the uninfected or heterozygous male and $t_{m2}$ = time spent sniffing, or near but not sniffing, the urine of the other male. Female subject was coded as a random effect in a mixed model.

## 2.4 RESULTS

### 2.4.1 Parasite burden and serum IgG levels

None of the 39 infected males had any adult worms attached to the gut mucosa or any evidence of the presence of larval stages on day 35 post infection. The level of parasite-specific immunoglobulin G in the sera of infected and uninfected mice was determined by ELISA. There was a clear

difference in levels of IgG in the blood of infected males versus sham-infected males (Figure 2.1, comparison of white and hatched bars and Table 2.1).

Table 2.1 Mean levels of immunoglobulin G in the sera of infected and uninfected mice, given by MHC genotype (for infected, gg n = 7, hh n = 6, jj n = 6, gh n = 9, gj n = 11; for uninfected, gg n = 3, hh n = 1, jj n = 1, gh n = 3, gj n = 3). Standard errors of the mean are included.

| MHC Genotype | Serum IgG (μg/ml) Infected | Uninfected |
|---|---|---|
| **gg** | 2.715 ± 0.450 | 0.136 ± 0.014 |
| **hh** | 2.676 ± 0.742 | 0.280 n.a. |
| **jj** | 3.643 ± 0.643 | 0.196 n.a. |
| **gh** | 3.297 ± 0.521 | 0.242 ± 0.151 |
| **gj** | 2.986 ± 0.398 | 0.263 ± 0.130 |

The mean parasite-specific immunolglobulin G for all infected individuals is significantly higher than that for uninfected individuals (infected IgG = 3.06 ± 1.44 μg/ml, uninfected IgG = 0.218 ± 0.17 μg/ml; $t_{41} = 12.06$, $p < 0.001$). The elevated level of IgG in the sera of the males infected with *T. muris* indicates that they are making, or have made, an immune response to the parasite. This verifies that the infection procedure was successful and infective parasitic worms were transferred to the males. Together with the complete absence of worms in the guts upon dissection, this confirms that the wild mice used in this study make an effective immune response when challenged with *T. muris* and are able to completely expel all worms within a 35 day period. The exact time at which the parasites were cleared is not apparent from these results, as IgG has been shown to persist for at least 25 days after the *T. muris* infection has been cleared (Else et al. 1990b; Blackwell and Else 2001; Blackwell and Else 2002). To determine when

exactly the worms were cleared would involve infecting a greater number of animals and dissecting at different time periods during the course of infection.

There was no difference between the levels of IgG in the sera of homozygous males and heterozygous males for either infected or uninfected individuals (Table 2.1, Figure 2.1, for infected $t_{37} = 0.28$, $p = 0.782$; for uninfected $t_9 = 0.80$, $p = 0.452$).

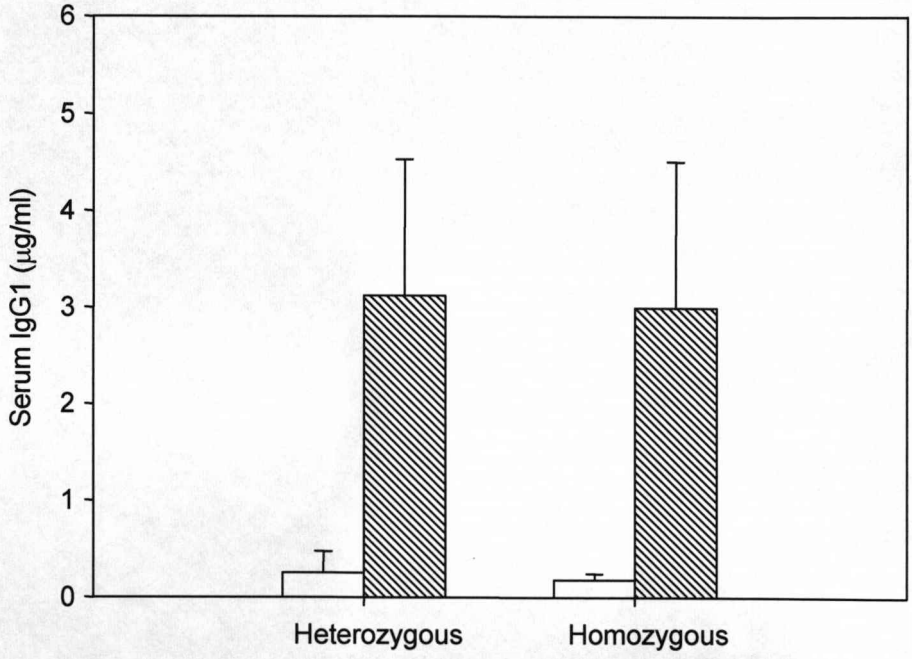



Figure 2.1 *Trichuris muris* specific IgG1 detected in serum collected from uninfected males (white bars) and infected males (hatched bars) at day 35 post infection. Results are mean and S.E. given according to heterozygosity at the MHC.

To further investigate the effects of single MHC haplotypes on the immune response to *T. muris* I calculated the means of IgG production for each genotype for the infected males. I did not include the uninfected males due to the small sample sizes and the obvious base-line level of IgG production in these animals. There is no difference in the levels of IgG produced by

males possessing any of the five different MHC genotypes (Table 2.1 and Figure 2.2, general linear model: $F_{2,34} = 0.57$, $p = 0.570$).

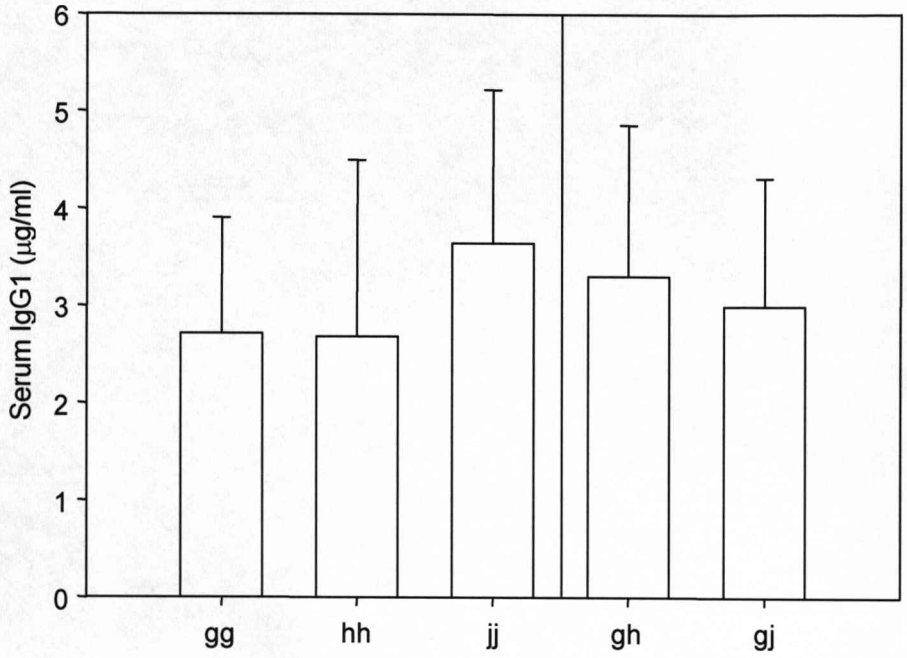



Figure 2.2 *Trichuris muris* specific IgG1 detected in serum collected from infected males at day 35 p.i.. Results are mean and S.E. given according to heterozygosity at the MHC; MHC homozygous in the left panel, MHC heterozygotes in the right; and by MHC haplotype (for gg n = 7, hh n = 6, jj n = 6, gh n = 9 and gj n = 11).

### 2.4.2 Odour preference

Trials using water instead of urine samples revealed that females normally spend very little time in the area of the cage where the stimuli are located (Table 2.2). All three measures used to assess female preference were significantly lower when using water rather than urine stimuli (sniff duration, $F_{3,92} = 5.69$, $p = 0.001$; not sniff duration, $F_{3,92} = 9.47$, $p < 0.001$; total time near stimulus, $F_{3,92} = 7.86$, $p < 0.001$), and all were approaching zero. This supports the observation that mice do not naturally visit the middle part of the cage where the stimuli are located, and thus any time

spent near the urine stimuli during trials is likely be due to the investigation process or a preference for being close to the urine source.

Table 2.2 Mean time spent near stimuli in the tests and the water control, with standard error of the mean (n = 12). Time periods refer to the first stimuli listed under trial type, i.e., time spent with male urine, not female. Given as time spent sniffing, time spent under the stimuli but not sniffing and total time spent near the stimuli.

| Trial | Sniff duration (s) | Not sniff duration (s) | Total time near stimuli (s) |
|---|---|---|---|
| **Water-water** | 0.37 ± 0.17 | 1.16 ± 0.25 | 1.53 ± 0.36 |
| **Male-female** | 3.20 ± 0.59 | 7.78 ± 2.52 | 10.97 ± 2.77 |
| **Uninfected-infected** | 4.13 ± 0.74 | 11.13 ± 1.83 | 15.25 ± 2.17 |
| **MHC heterozygous-MHC homozygous** | 3.49 ± 0.49 | 12.63 ± 2.92 | 16.12 ± 3.20 |

Female preference for male over female odour was confirmed before any other tests were carried out. Female preference for spending time in direct contact with male scent has been shown in previous studies (Moncho-Bogani et al. 2002; Moncho-Bogani et al. 2005; Martinez-Ricos et al. 2007; Ramm 2008) and this test ensured the females were behaving as expected and that the test was capable of determining female preferences. The duration of female response was not normally distributed throughout, so differences in behaviour to test odours were assessed using Wilcoxon signed rank tests for all three test types.

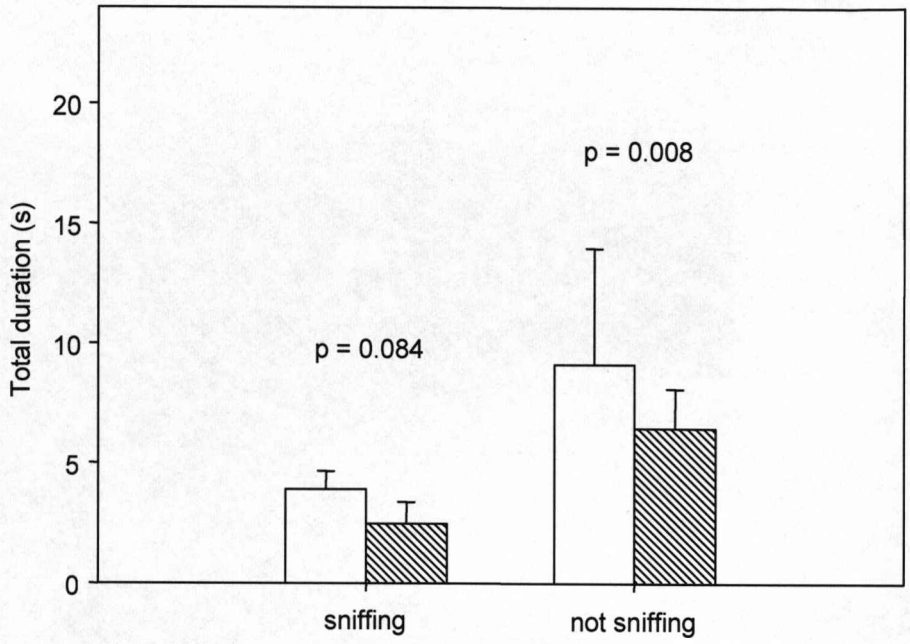



Figure 2.3 Preference for male (white bars) over female (hatched bars) urine. Response is broken down into time spent sniffing the stimulus and time under the stimulus but not sniffing the urine (median and 75th percentile, n = 12). P values indicate Wilcoxon matched-pair tests.

Females consistently spend more time with male urine than female urine (Figure 2.3). Whereas this is not the case when sniffing the stimulus, they spend significantly more time with the male urine not investigating the odour. This shows that although females in this trial do not spend longer investigating the odour of unfamiliar males, they do prefer to spend longer close to the urine stimulus, reflecting a preference for spending time close to the odour source.

I tested female preference for urine from uninfected males over urine from males infected with *Trichuris muris* (Figure 2.4). Females preferred spending non-investigatory time near the urine of uninfected males over infected males. Interestingly, the difference in the amount of time spent near the urine of uninfected compared to infected males (8.02 s difference) is greater than the difference in time spent near the urine of males

compared to females (2.12 s difference) and it is almost as significant a difference ($p = 0.034$ for uninfected-infected, $p = 0.008$ for male-female). Once again there was no significant difference in the time spent sniffing the different urine stimuli. This shows that, although females spend an equal amount of time gathering information from the urine of uninfected and infected males, they subsequently prefer to position themselves in close proximity to the urine of uninfected males rather than the urine of infected ones.

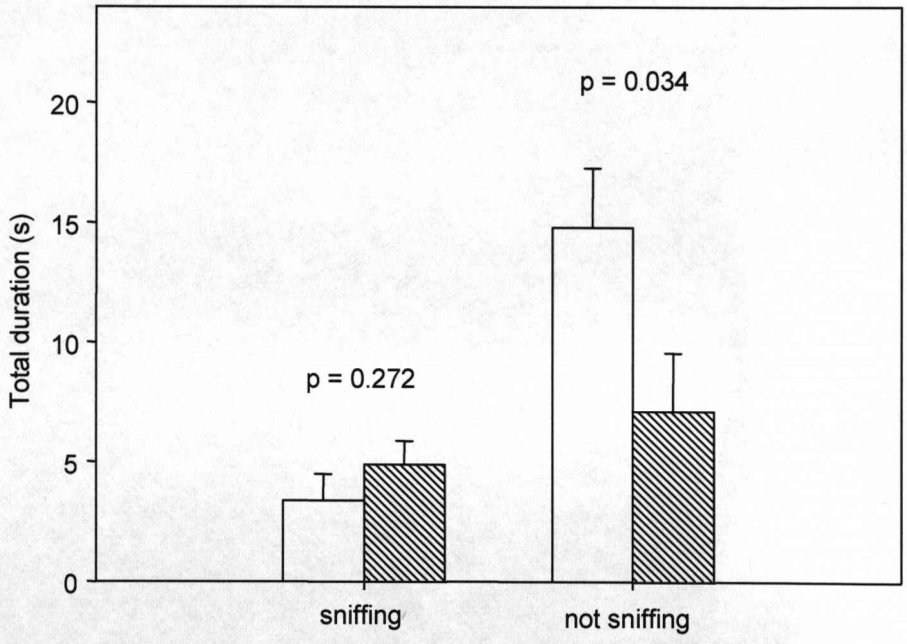



Figure 2.4 Preference for uninfected (white bars) over infected (hatched bars) urine. Response is broken down into time spent sniffing the stimulus and time under the stimulus but not sniffing the urine (median and 75$^{th}$ percentile, n = 12). P values indicate Wilcoxon matched-pair tests.

In the final trial I tested female preference for urine from males that were homozygous or heterozygous at the MHC (Figure 2.5). Infection status was always matched in these trials and an equal number of uninfected and infected pairs were used. There was no significant difference in either the duration of sniffing or time spent near the urine not sniffing. Females did

not appear to show any MHC determined preference. This may be because they are unable to tell the difference between the urine of an MHC homozygote and the urine of an MHC heterozygote without the use of other cues. Alternatively, it could indicate that they do not have an innate attraction to MHC hetero- or homozygosity.

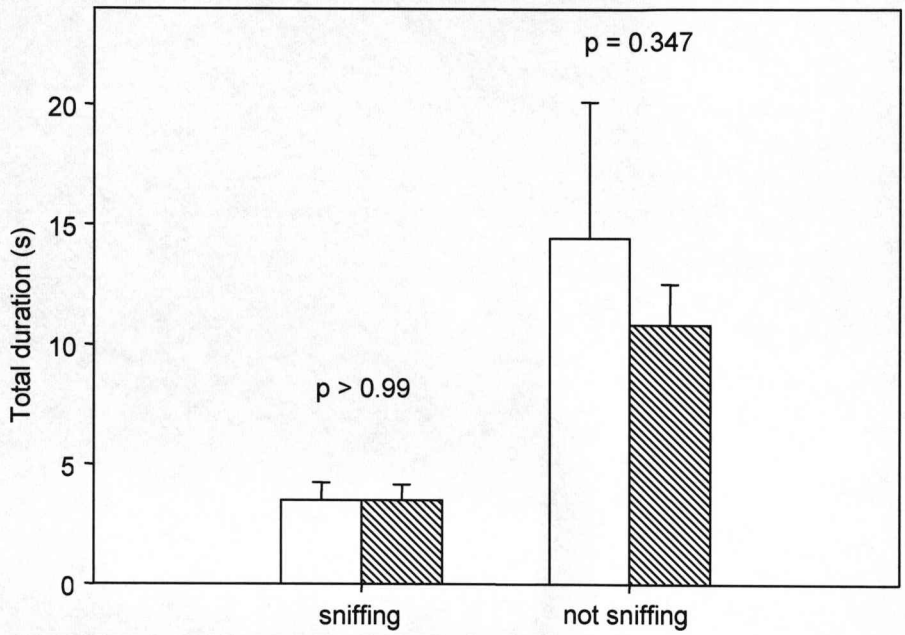



Figure 2.5 Preference for MHC heterozygous (white bars) over MHC homozygous (hatched bars) urine. Response is broken down into time spent sniffing the stimulus, time under the stimulus but not sniffing the urine and total time under the stimulus (median and 75$^{th}$ percentile, n = 12). P values indicate Wilcoxon matched-pair tests.

Interestingly, there was no trend in the mean amount of time spent sniffing or close to the urine correlated with infectivity status in this trial ($t_{19} = 0.18$, $p = 0.857$). The sample size is small, n = 6 for each infection status, so I analysed the time spent investigating the urine of males infected with parasites and time spent investigating the urine of uninfected males across all trial types. This also shows no significant difference ($t_{43} = 0.55$, $p = 0.582$). The cue in the urine that reports the infection status of the male should be completely novel to the subject females. Whereas they have been

housed in the same room as males prior to the trials, thus encountering, however indirectly, the airborne urinary volatiles from male mice, they have never previously had contact with mice experimentally infected with parasites. It would be expected that they should spend longer investigating the novel, infection related odour to gain information from this unfamiliar scent. That they do not, but still avoid spending time with the urine stimulus from infected males, suggests that may be an innate component to the avoidance of infected individuals mediated through scent.

The results of the mixed model using data from the infectivity and heterozygosity scent trails combined corroborate the results found in individual analyses. Females show a non-significant trend to sniff infected urine more than uninfected (log ratio = 0.39, $p = 0.07$). Females spent significantly more non-investigatory time near the urine of uninfected males over infected males (log ratio = -1.24, $p < 0.01$). There was no difference in the time spent with urine from heterozygous males compared to that spent with urine from homozygous males. The odour trials carried out in this study confirm that females prefer to spend time close to the urine of male mice over female mice, and that females have a preference for uninfected males over infected males. They fail to find any significant preference for MHC heterozygosity over homozygosity.

## 2.5 DISCUSSION

Previous studies investigating the pathogenicity of *Trichuris muris* in inbred laboratory strains of mice have resulted in complete expulsion, complete susceptibility or a split response, dependent on strain (Wakelin 1967; Wakelin 1970; Wakelin 1975; Else and Wakelin 1988). Susceptibility to *T. muris* has been linked to both MHC and non-MHC

genes (Else and Wakelin 1988), with B10 genetic background being a poor responder compared to strains with a BALB background, and the $H2^k$ haplotype being completely susceptible to the parasite independent of genetic background. These results suggest that both the initial recognition of pathogenic antigen by the MHC, and the downstream immune processes controlled by non-MHC genes are important in the clearance of *T. muris*. Only one previous study has infected wild *Mus musculus domesticus* with *T. muris* (Behnke and Wakelin 1973). The majority of the infected laboratory-bred wild-derived mice (14 out of 15) had no worms present at 35 days post infection. The results of this study are consistent with Behnke and Wakelin. All mice infected were able to clear the parasite within 35 days post infection. It appears that possessing a susceptible MHC type and susceptible background genes may be a rarity in natural populations and that those inbred strains that are able to mount an effective immune response are more representative of animals in the wild. The immuno-competence of wild mice challenged with a *T. muris* infection is more closely characterised by those inbred strains that expel the parasite over a 35 day period post infection; those with inbred MHC haplotypes b and q (Else et al. 1990a).

As *T. muris* has a prepatent period of approximately 6 weeks, an effective, acute immune response such as was mounted in this study would not allow worms to reach sexual maturity before being expelled from the host. This begs the question of how *T. muris* populations are maintained in the wild. Behnke and Wakelin (1973) also investigated the worm burden of *T. muris* in a natural population and found high levels of mice were infected (88%) with very low numbers of worms (<20). They concluded that *T. muris* could persist in natural populations only at levels low enough to avoid being detected by the immune system.

The early development of high levels of IgG production in response to infection with *T. muris* has been shown to be dependent on MHC type and correlated with quick clearance of the parasite (Else et al. 1990b). The IgG levels of the infected males in this study are all above that seen in sham-infected individuals. Despite a relatively large variance in the levels of IgG found in the sera, there seems to be no association between MHC genotype and the amount of IgG produced. Else et al. (1990b) used SDS-PAGE to show that there existed a H2-linked control of antibody responses; different MHC-congenic strains bound different excretory/secretory (E/S) antigens of *T. muris* and the recognition of these different antigens correlated with the level of IgG produced. If the B-cell-surface immunoglobulins that bind parasitic antigen are determined by MHC type as Else's result suggests, it would be plausible to assume that MHC heterozygotes could have B cells that bind a greater number of differing antigenic peptides than MHC homozygotes, and potentially produce a greater IgG response to the parasite. However, there is no increase in IgG production associated with MHC heterozygosity in this study. With codominant genes the effect of heterozygosity at an allele can be additive, i.e. the heterozygote phenotype is half-way between the two homozygous phenotypes or, as with the hypothesis suggested for the theory of over-dominance at the MHC, the heterozygote can be fitter than either homozygote (heterozygote superiority). There is no evidence for either hypothesis in the IgG levels seen in this study as all genotypes do equally well at expelling the parasite and all produce equivalent levels of IgG. However, this does not mean that the theory of heterozygote advantage does not apply to MHC mediated resistance to *T. muris*. It may be that the functional MHC alleles present in the three haplotypes tested are all equally good at binding and presenting *Trichuris* antigens to cells of the immune system. Thus no additive or over-dominant effects would be seen. Alternatively, a non-MHC linked gene may have a greater affect on the immune response to *T. muris* than MHC

genes. This would mean that once the immune response had been activated, the level of response shown would be independent of the MHC.

The limitations of the current study must be kept in mind when interpreting the result. Firstly, all animals cleared the infection regardless of MHC type. This means that no real conclusion can be drawn regarding MHC heterozygote advantage, or the immuno-competence of individual MHC haplotypes. The use of a parasite with a broader range of infectivity would have allowed a comparison of MHC and non-MHC related genes on susceptibility. Secondly, the three MHC haplotypes compared in this study were chosen due to their abundance in the breeding stock rather than any intrinsic properties. If more genetically diverse haplotypes had been used, there may have been greater differences in the response to the parasite infection. Finally, the theory of heterozygote superiority could not be tested in this experimental set-up. In contrast to heterozygote advantage where the heterozygote is always fitter than either homozygote, the theory of heterozygote superiority states that the heterozygote is only fitter in certain situations e.g., when challenged with multiple parasites, or when exposed to high stress environments. The current study used a single parasitic infection and the animals had food and water ad libitum.

Studies have shown that females prefer the urine of males that are not infected to those that are infected; in mice using the intestinal nematode *Heligmosomoides polygyrus*, (Kavaliers and Colwell 1995b); and the protozoan *Eimeria vermiformis* (Kavaliers and Colwell 1993); female meadow voles discriminated against males infected with the nematode *Trichinella spiralis* (Klein et al. 1999); and female rats preferred the odours of uninfected males over those infected with *Hymenolepis diminuita* (Willis and Poulin 2000). Further studies show female birds and mice avoid mating with males that are infected with parasites (Hamilton and Zuk 1982; Zuk 1992; Ehman and Scott 2002). Females choosing not to mate

with infected males directly reduce their risk of being infected and Able (1996) showed that when male traits indicate parasite load females will choose to mate with males with lower infection levels to that end. Females may also be choosing to mate with un-parasitized males to increase the likelihood of passing parasite resistance associated genes to their offspring (Zuk 1992).

The olfactory pathways involved in the distinction between the odours of parasitized and non-parasitized individuals have been investigated; the hormone estrogen and the neuropeptides oxytocin and vasopressin have been shown to be essential components in the recognition and avoidance of the odour of parasitized males (Kavaliers et al. 2003; Kavaliers et al. 2004; Kavaliers et al. 2005). However the actual cue that conveys information on infection status is unknown. Females could be responding negatively to an odour cue associated with infection, or an active immune response, that males are unable to hide; or they may be responding positively to a signal of health, or the ability to resist infection, in the males that are uninfected. As in previous studies demonstrating female preference for the odour of un-parasitized males (Kavaliers and Colwell 1995a), the male mice in this study showed no clinical symptoms of disease, which makes it unlikely that females were responding to stress or illness related odour cues. It is more probable that females are responding to an odour cue associated with infection, or with an active immune response, rather than the general condition of the males. The reduction in attractiveness of males which have had immune activation by antigenic challenge alone further suggests that it is host generated signs of infection that are being detected by females, rather than an odour cue of parasite origin (Klein and Nelson 1999; Yamazaki et al. 2002; Zala et al. 2004; Litvinova et al. 2005). Host immune responses to parasites can induce changes in production of MHC related volatile and non-volatile odorants (Penn and Potts 1998a). There is

substantial evidence showing that mice can distinguish MHC differences in urinary odour alone (Yamazaki et al. 1976; Yamaguchi et al. 1981; Potts et al. 1991; Manning et al. 1992; Potts et al. 1994; Hurst et al. 2001; Beynon and Hurst 2003; Brennan 2004; Cheetham et al. 2007), so it is possible that parasite induced changes in these odour cues may indicate the infectivity status of potential mates.

There has been conflicting evidence about the influence of MHC determined odours on mate choice. MHC biased mate choice has been found in a number of studies using inbred strains of mice (Yamazaki et al. 1976; Beauchamp et al. 1988; Yamazaki et al. 1988), wild-inbred hybrid mice with inbred MHC haplotypes (Potts et al. 1991; Potts et al. 1994), fish (Reusch et al. 2001; Milinski et al. 2005) and humans (Wedekind et al. 1995), but there are numerous studies that find no evidence of MHC determined mating patterns (Hedrick and Black 1997; Paterson and Pemberton 1997; Landry et al. 2001; Sauermann et al. 2001; Ekblom et al. 2004; Westerdahl 2004; Sherborne et al. 2007). The current study uses female preference for spending non-investigatory time with an odour to represent differential attraction to males. There was no effect of MHC heterozygosity on the time females spent investigating the urine of males or the time they spent near but not investigating it. Females that choose to mate with the fittest males gain benefits from the resources he holds and could pass increased fitness to her offspring. Individuals that are homozygous across the genome have a loss in fitness when compared to individuals that are more heterozygous, due to a loss in heterozygote advantage and the expression of recessive deleterious mutations (Mitton 1993; David 1998; Hansson and Westerberg 2002). If females were using heterozygosity at the MHC to signify genome-wide heterozygosity (Brown 1998; Roberts et al. 2005), we would expect the females in this study to prefer to spend time with the urine of MHC heterozygous males. It appears

that female mice do not find MHC heterozygous males more attractive than MHC homozygotes based on urinary cues alone. In the wild, MHC heterozygosity should correlate with genome-wide heterozygosity which in turn is likely to correlate with overall condition and fitness. Thus females in the wild may be mating with males that are heterozygous at the MHC simply because it is related to other signs of condition. However the MHC may still have an effect on mate choice, as there are different hypothesise about its influence.

Disassortative mating at the MHC has been hypothesised as a means of avoiding inbreeding (Brown and Eklund 1994), thus females may only show an MHC biased preference when they are presented with close kin as mates. The females used in this study were not closely related to the males and were unlikely to share MHC alleles with them. If the only influence that the MHC has on female mate choice was the avoidance of males with whom they share alleles, this study would not be capable of detecting evidence of a bias. The MHC as a potential mechanism for the avoidance of inbreeding is further discussed in chapter 3. What the current study shows, however, is that while females have no preference for heterozygosity or homozygosity at the MHC, they do show a significant preference for uninfected over infected males.

This study shows that for wild mice infection status is a stronger driver of mate choice than genetic indicators of either immune system or genome wide quality. Females are not using a potentially complex assessment of genotype to assess mates, but are simply attracted to males that are healthy over than those that are infected with parasites.

# Chapter 3

# THE ROLES OF MHC AND MUP IN THE AVOIDANCE OF INBREEDING

## 3.1 INTRODUCTION

Soon after the immunological role of MHC molecules was discovered, Yamazaki et al. (1976) observed an influence of H2 genes, the murine MHC, on mate choice in inbred laboratory strains. When males were offered an H2 similar female and an H2 dissimilar female, they showed a clear mating preference for the dissimilar female. Yamazaki et al. (1983) later showed that male mice could discriminate between very small differences in MHC congenic strains; as small as three amino acid changes within the class I region. These studies show that mice are able to recognise differences in the MHC types of individuals, and that these differences can influence mate choice. However, despite 30 years of subsequent study, the exact role of the MHC in mate choice is still not clear (Jordan and Bruford 1998).

MHC influenced mate choice does not seem to be a choice for "good genes", that is, alleles that increase fitness independently of their homologous allele or the rest of the genome. Instead, most research has focused on the degree of MHC similarity between mates (Grob et al. 1998). There has been evidence of MHC disassortative matings in many species (Yamazaki et al. 1976; Beauchamp et al. 1988; Yamazaki et al. 1988; Potts et al. 1991; Potts et al. 1994; Wedekind et al. 1995; Reusch et al. 2001; Milinski et al. 2005). Choosing to mate with MHC dissimilar

individuals would have two benefits. Firstly, it would facilitate the avoidance of inbreeding. MHC genes are highly polymorphic and, in a natural population, a high degree of intraspecific MHC sharing is likely to indicate a close relationship. Thus the MHC has been suggested as a signal of relatedness that is used for kin recognition (Manning et al. 1992; Brown and Eklund 1994) and the avoidance of inbreeding (Penn 2002). Studies using inbred strains of mice have shown both male and female preference for mates with dissimilar MHC types (Trivers 1972; Yamazaki et al. 1976; Beauchamp et al. 1988; Yamazaki et al. 1988; Egid and Brown 1989; Eklund et al. 1991), although preference seems to be strain dependent, with some strains showing no bias and some preferring MHC similar strains. If the MHC were being used as a marker of close kin and facilitating inbreeding avoidance, mice should always avoid mating with MHC identical individuals. The strain specific results may reflect unnatural behaviour patterns (Van Oortmeerssen 1971) and mating choices characteristic of strains that have been experimentally inbred for hundreds of generations (Levine et al. 1966). Secondly, choosing an MHC dissimilar mate would promote heterozygosity in offspring. Heterozygosity at the MHC provides a greater potential for recognising pathogens. MHC molecules possess distinct binding specificities for antigens (Doherty and Zinkernagel 1975), thus individuals heterozygous at an MHC locus could bind up to twice as many pathogenic antigens than homozygotes. Heterozygote advantage, a type of overdominant selection, is frequently cited as a driver for the maintenance of the high levels of polymorphism at the MHC (Hughes and Nei 1988; Hughes and Nei 1989; Hughes and Hughes 1995). Thus mating to produce MHC heterozygous offspring is likely to produce offspring that have high levels of heterozygosity across the rest of the genome. Genome-wide heterozygosity confers fitness benefits due to heterozygote advantage and the lack of expression of recessive deleterious mutations that can be a consequence of inbreeding

and excessive homozygosity (Mitton 1993; David 1998; Hansson and Westerberg 2002).

This chapter addresses the issues surrounding both the use of the MHC as a marker of relatedness, and the fitness benefits associated with MHC and genome-wide heterozygosity. We use wild-derived *Mus musculus domesticus* in two separate studies observing different aspects of potential MHC influence on mate choice.

The first study builds on the work of Potts et al. (1991; 1994), who used crosses between wild caught mice and inbred strains to generate individuals with MHC haplotypes from one of four inbred strains, while the rest of the genome was half wild and half derived from classical inbred strains. These animals were then further crossed to generate MHC heterozygotes. Replicate populations of MHC homozygotes and heterozygotes (F2 - F6) were released into large semi-natural enclosures and allowed to breed undisturbed until the pups were near weaning age when all offspring were trapped and genotyped. The numbers of MHC homozygote offspring produced in the enclosures were significantly less than expected from random mating. Although Potts et al. (1991; 1994) demonstrate actual mating bias (unlike many previous studies that simply show preference for an odour) there are still several problems with extrapolating their result to behaviour in natural populations. The mice in the study still have 50% of their genome derived from inbred strains, which could alter their behaviour from that of completely wild mice. The enclosures used were quite open to allow observational data to be recorded during the trials. This is very unlike normal wild mice habitats, as they tend to settle in areas providing shelter from predators and from other mice during conflicts. Not providing proper shelter may have prevented the formation of dominance relationships between the males, which could influence male reproductive success. Finally and most importantly, the

MHC haplotypes in this study were correlated with the rest of the genome. Potts et al. do not control genome-wide relatedness or determine parentage, thus parental similarity across the genome or at non-MHC marker loci could not be verified and may have influenced mating behaviour.

The MHC has been suggested as a murine signal of relatedness because of its polymorphic nature (Trowsdale 1993) and MHC related odour cues can be differentiated in urine alone (Yamazaki et al. 1979; Yamaguchi et al. 1981; Yamazaki et al. 1983; Yamazaki et al. 1994; Yamazaki et al. 1998; Carroll et al. 2002; Penn 2002; Brennan and Kendrick 2006; Slev et al. 2006). In this study we investigate the role on mate choice of another highly polymorphic and odour influencing group of genes. Major urinary proteins (MUPs) comprise over 99% of the protein component of mouse urine (Hurst and Beynon 2004). Similar to MHC molecules binding pathogenic antigen, MUPs have a binding groove that binds lipophilic molecules, such as male signalling volatiles. These volatiles persist in the environment for a longer period when bound to a MUP molecule than when unbound (Robertson et al. 1993). Chamero et al. (2007) showed that mice can directly detect major urinary proteins, not bound ligand, through receptors in the vomeronasal organ. Thus MUPs themselves can act as informative scent cues. Like the MHC, MUPs have been suggested as odour cues mediating individual recognition (Hurst et al. 2001; Brennan 2004; Cheetham et al. 2007).

In the first study we investigated the roles of MHC and MUP in the avoidance of mating with close kin. We used four replicate semi-natural enclosures, over five times the size of those used by Potts et al. (1991; 1994), and with natural vegetative cover to provide mice with adequate shelter. We used wild-derived mice that were bred in captivity from the ancestors of mice caught from five different locations across the northwest of England. We typed all founders and offspring at the MHC and MUP and

used a suite of genome-wide microsatellite markers to determine parentage so individual matings could be determined.

## 3.2 AIMS

We investigated the roles of MHC and MUP in the avoidance of inbreeding and the reproductive success of *Mus musculus domesticus*. The first experiment used a natural set-up similar to a previous study looking at mate choice and the MHC (Potts et al. 1991; Potts et al. 1994). Uniquely, we used wild-derived mice in our enclosure study and allowed them to breed freely in semi-natural enclosures. We tested for a deficit in matings between animals that share MHC and/or MUP haplotypes. Our breeding design meant that background relatedness was controlled for and we were able to test the effects of MHC and MUP without genetic matching at areas of the genome not closely linked to MHC and MUP confounding results (Sherborne et al. 2007 in appendix).

The second experiment allowed wild female mice to mate with pairs of male laboratory mice that were: i) homozygous or heterozygous at MHC, with MUP and background heterozygosity controlled, ii) homozygous or heterozygous at MUP, with MHC and background heterozygosity controlled, and iii) heterozygous at both MHC and MUP, but with differing levels of background heterozygosity. This design allowed the effects of MHC, MUP and genome-wide heterozygosity to be separated from each other. Assessing the number of offspring sired by heterozygous males in each of the three trials allowed us to determine whether heterozygosity at MHC, MUP or background genes influenced male reproductive success.

## *3.3 Enclosure Study*

### 3.3.1 METHODS

#### 3.3.1.1 Animals

All animals used to breed founders for release into the enclosures were the outbred F1 - F3 offspring of wild *Mus musculus domesticus* caught from 5 different locations across the northwest of England. The breeding plan was designed to imitate a natural population in which a dominant male sires the majority of the offspring in a single generation resulting in a local population of full-sib or paternal half-sib mice. As inbreeding levels must reach a certain threshold before avoidance behaviour becomes apparent (Kokko and Ots 2006), mating patterns at levels of relatedness lower than full-sib ($r = 0.50$) may reflect factors other than inbreeding avoidance. Our breeding design ensures that mice would benefit from using inbreeding avoidance mechanisms when choosing mates.

For each of the four enclosures a single male was mated with two sets of three unrelated females (each for 14 days) in breeding cages (40 x 23.5 x 12.5 cm, North Kent Plastics, UK). For each of the four populations, three to five females produced offspring sired by the same male (appendix table 4). Females were housed singly prior to parturition (cages 40 x 23.5 x 12.5 cm) so that full-sib founders were familiar from birth but half sibs were only encountered after release into the enclosures when all mice were in adulthood. Offspring were separated into single sex and sibship groups at weaning (4 weeks) and released into the enclosures at 48 - 65 days old. Before release, all founders were given subcutaneous radio frequency identification (RFID) tags for individual identification. Genetic samples (1 - 2 mm removed from the tip of the tail under general anaesthesia

(halothane)) were taken from the founders and their parents to allow determination of the MHC and MUP haplotypes of each of the founders and, later to reconstruct the parentage of the offspring captured from the enclosures.

### 3.3.1.2 Population Enclosures

The founders were released into four adjacent outdoor enclosures (25 x 10 m) designed to closely mimic a natural habitat for house mice (figure 3.1).

Figure 3.1. Photo of a single enclosure showing the grass covering, concrete shelters and food and water stations.

Sheet-aluzinc walls prevented escape or contact between populations (1.3 m high walls with concrete foundations to prevent climbing or burrowing). Wire mesh upper walls and roof prevented predation. The large size of the enclosures ensured that all founders had ample room and resources to establish territories. Cover was provided by a natural long grass and 30 nest boxes spread throughout the enclosure, although mice appeared to nest mainly in the grass. Twelve weeks after release, during a period of heavy rain, ten concrete shelters (45 x 45 x 35 cm) were added to provide extra shelter. Ten food and water stations were spread throughout each enclosure

providing mice with food (Lab Diet 5002 Certified rodent diet) and water ad libitum.

Mice were allowed to breed undisturbed for 15 weeks before a 4 week trapping period began. Sex, weight and age class (juvenile, subadult, adult) were recorded for all captures. Urine samples were obtained for further studies on MUP phenotyping. Founder mice were identified from their RFID tags and housed in captivity for use in further studies. All founders were recaptured, proving they were all available as possible mates throughout the study. Non-founders were culled humanely under halothane anaesthetic and tail snips were taken for genotyping. Blood and gut samples were also taken for other studies.

3.3.1.3 MHC and MUP genotyping

All MHC and MUP genotyping was carried out by Francine Jury and the CIGMR group at the University of Manchester. The MHC and MUP haplotypes of the founders for each population were established by using eight microsatellite markers across the MHC region on chromosome 17 and eight microsatellite markers surrounding the MUP region on chromosome 4. The patterns of alleles expressed in the founders were compared to their parents to identify linked alleles in the same haplotype and any crossover events. DNA was extracted from a 1 – 5 mm tail snip with mouse-tail extraction kits (Tepnel life sciences). Twelve markers were originally selected for the MHC region (chromosome 17) with the ENSEMBL mouse genome database (Hubbard et al. 2007). Three markers were deemed outside the MHC region (confirmed by crossover events in test samples) and were removed from the analysis. Three other markers were excluded from analysis because of amplification problems. Eleven markers were identified in the putative MUP 4 locus (chromosome 4). Three of the markers failed to amplify or were found to be monomorphic in

wild mice and were removed from the data set. The 14 chosen loci (6 MHC and 8 MUP) are listed in appendix tables 1 and 2 (labelled 'enc' for enclosure study). The primer sets were designed so that markers would fall into one of three size groups with nonoverlapping allele lengths with the forward primer in each group 5'-end-labeled with a fluorescent phosphoramidite (6-FAM, HEX, or NED). The loci were organized into four multiplex loading groups, containing mixed loci from both the MHC and MUP regions. PCR amplification reactions were performed in a final 10 ml volume containing 10 ng of DNA, 0.1 mM of each primer, 0.2 mM of each dNTP, 0.05 units of Taq DNA polymerase (Hotstar Taq, QIAGEN), and 2.5mM $MgCl_2$ in the supplied reaction buffer. PCR amplification was performed on a thermal cycler (GRI) with a touchdown PCR program at one cycle of 95°C for 15 min to activate the Taq polymerase and then 94°C for 20 s, 30 s at 66°C, -0.5°C per cycle for 14 cycles, and then 60°C for 20 cycles and 72°C for 30 s, followed by 72°C for 10 min. For each individual sample, PCR products from each multiplex group were diluted 1 ml for 6-FAM-labeled products and 2 ml for the other labelled products into 200 ml ddH20. The multiplexed mixture (1.2 ml) was combined with 5 ml deionised formamide and 0.1 ml internal lane standard (ROX400, Applied Biosystems). Samples were electrophoresed on a 3100 Genetic Analyzer capillary electrophoresis system with a 22 cm array (Applied Biosystems). The fluorescence emission for each dye was collected and analyzed for size variation with the Applied Biosystems GENESCAN (version 3.7) and GENOTYPER (version 3.7) DNA-fragment analysis software.

MUP and MHC haplotypes were determined by comparing the patterns of alleles in the founders to their parents and tracing linked alleles. This also allowed us to identify crossover events. All founders were heterozygous for MHC and MUP except four females in population B, which were MUP

homozygous. To genotype the F1 offspring of the founders, we selected three to four of the MHC markers and three to four MUP markers that reliably discriminated each haplotype within a population (different sets of markers were used for each population, see appendix tables 1 and 2). To check that MHC homozygotes were correctly identified and were not due to poor primer amplification, we ran an additional marker for any animals initially classified as MHC homozygotes. This confirmed homozygosity in all cases.

3.3.1.4 Parentage analysis

DNA was extracted from tail snips using the AGOWA mag DNA Isolation Kit following the manufacturer's instructions. For high throughput extraction we used the AGOWA 96-well magnetic Separator in combination with the Hamilton Microlab STAR robot. The sequences of primers for 12 microsatellite markers were chosen from the Mouse Genome Informatics site (MGI 3.51) such that they were not linked (>50cM) to each other or the microsatellites used for typing at the MHC and MUP. These 12 markers were not informative enough to reliably assign more than 25% of offspring to parent pairs, so another 12 markers were used to increase the power of the parentage analysis (appendix table 3). The forward primer of each marker was 5'-fluorescently labelled with 6-FAM, NED, PET or VIC dependent on the size category of the marker (small, medium or large) such that 12 could be pooled into a single run. PCR amplification was conducted in 10 μl reactions contained 10 ng of DNA, 0.5μM of each primer and 5.0μl of 2X BioMix Red reaction mix (Bioline [London, UK]). Plates were incubated in a thermal cycler (Peltier) at 95°C for 2 min, followed by 30 cycles of 95°C for 30s, 52°C – 58°C (depending on the primer) for 2 min, then 72°C for 30s, with a final extension at 72°C for 10 min. The PCR reactions were then diluted 50- to 100- fold, multiplexed and 1 μl of the mix was added to 8.8 μl of

formamide and 0.2μl GeneScan LIZ500 size standard (Applied Biosystems). Size was determined with an ABI PRISM 3100 DNA analyzer (Applied Biosystems) and GeneMapper v3.0 software designed for microsatellite analysis (Applied Biosystems).

The parentage assignment program Cervus 3.0 (Kalinowski et al. 2007) was used to assign individuals to a parent pair using maximum-likelihood methods.

3.3.1.5 Procedure for assigning matings

As parentage could only reliably be assigned to the offspring of founders (the F1 generation), we excluded all animals that could conceivably be F2s on the basis of weight and date of capture as follows; assuming the founders began mating immediately upon introduction into the enclosures F1 individuals could have started breeding nine weeks after founder release (given a three week gestation period and six weeks for F1s to reach sexual maturity). Trapping began 15 weeks after founder release, thus animals three weeks old or younger could have been from the F2 generation. Sex-specific growth curves were constructed using weights collected from the progeny of the founders re-mated in captivity (n = 19 litters weighed at weekly intervals from 3 weeks after birth) and these were used in conjugation with date of capture to exclude any individuals that could be the offspring of F1 females. The parentage assignment detailed above was carried out blind to MUP and MHC types. Using all offspring produced per parent pair in statistical tests would lead to pseudo-replication, so we determined minimum number of matings that must have occurred to explain the offspring captured. We plotted weight against capture date for the offspring of each female and used the sex-specific growth curves and paternity information to assign offspring to litters. Founding females could have had up to 3 litters that were over 3 weeks old at the start of trapping

and the maximum litter size for wild house mice is 9. We took the most conservative approach and only assigned offspring from the same father to different litters if their weights, date of capture and the litter size precluded a single mating event. This resulted in the final assignment of 193 matings across the four enclosures. All females produced at least two litters and 67% of litters were sired by more than one male.

### 3.3.1.6 Statistical Analysis

The statistics package R v2.3.1 (www.r-project.org) was used to fit logistic multinomial models to the data using likelihood methods. Two response variables were analysed (i) the number of successful matings made by a female from each of a set of available males, and (ii) the number of offspring produced from each of a set of observed matings. The log likelihood can be given by,

$$\left.\begin{aligned} \ln L &= \sum_i \sum_j y_{ij} \ln P_{ij} \\ P_{ij} &= \frac{\exp\left(\left(\mathbf{x}_{ij} - \mathbf{x}_{i1}\right)\beta\right)}{\sum_j \exp\left(\left(\mathbf{x}_{ij} - \mathbf{x}_{i1}\right)\beta\right)} \end{aligned}\right\}$$

where $y_{ij}$ is the observed number of matings or offspring from sire *j* and dam *i*, and $x_{ij}$ is a vector of explanatory variables that describes sire *j* with respect to dam *i* (i.e. whether the sire is a half- or full-sib of the dam, and whether the dyad shares one, two or no MHC or MUP haplotypes respectively). β is the vector of coefficients, fitted by maximising the log likelihood using numerical optimization. In the case of two-tailed tests, the significance of explanatory variables was determined by comparison of LR statistics (i.e., twice the log likelihood ratio for a pair of nested models) to a distribution generated by a permutation procedure where the genotype or relatedness of the dam was randomly reordered with respect to the set of

available males. This permutation approach is more conservative than the comparison of LR statistics against a chi-square distribution since it controls for potentially inflated Type I errors that might arise from the repeated preference of a dam to a particular sire, unconnected to the relatedness or genotype of the sire with respect to the dam. In the case of one-tailed tests, significance was determined by comparison of the coefficient fitted to the observed data with the corresponding distribution generated by the permutation procedure.

### 3.3.2 RESULTS

#### 3.3.2.1 Parentage Analysis

81 founders and 483 offspring were typed at 24 genome-wide microsatellite markers (appendix table 3). The proportion of loci successfully typed was 85%. Repeat typing of 62 individuals (9% of the total number of mice typed) across all loci produced only 5 mismatches, giving an error rate of 0.33%. Using full-sibling and half-sibling founders meant that the average heterozygosity of each enclosure was lower than expected from an outbred population. However, the 24 markers gave a high enough polymorphic information content that the critical delta (delta is the highest natural log of the likelihood ratio (LOD score) minus the next best) needed for 95% confident assignment, as calculated from 10,000 simulations of parent-pair assignment, was zero in all four enclosures. Cervus 3.0 allows input of priors, such as expected rate of inbreeding in the population and the relatedness of candidate parents, into its simulation calculation so we were convinced that assignments were robust. Confidence in assignments was further strengthened when we looked at the MHC and MUP haplotypes of the offspring and their assigned parents. A very low proportion (2.8%) of assignments had MHC or MUP haplotypes

in the offspring that could not have been inherited from the assigned parents, and these mismatches came from offspring that were likely to be from the F2 generation based on their weight and date of capture. These were removed from all subsequent analysis. We checked that the parentage assignment was not biased in favour of either half sib or full sib matings by comparing the mean delta values of assignments between half sibs and full sibs, which did not significantly differ from each other (mean for half sibs 1.86, for full sibs 2.27, $t = -1.46$, $p = 0.147$) (Figure 3.2).

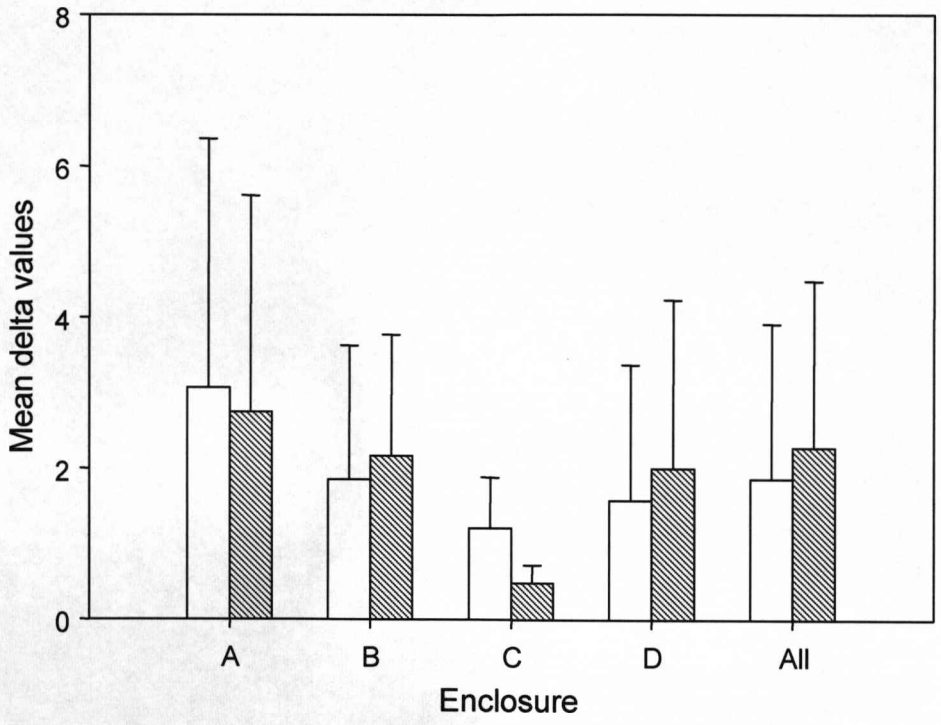



Figure 3.2. The mean delta values (± standard deviation) for offspring assigned to half sib parents (white bars) and full sib parents (hatched bars), by enclosure and all data combined. Error bars represent one standard deviation from the mean.

### 3.3.2.2 Analysis of matings

We found a deficit in matings between full-sibling pairs (Figure 3.3, Tables 3.1 and 3.2: multinomial logistic model 1, comparison of observed values with numbers expected under a model of random mating, $p = 0.08$),

consistent with inbreeding avoidance. This deficit was slightly too small to differ significantly from a model of random mating, however significantly fewer offspring were produced per mating event between full-sibling pairs (Table 3.3: model 1, $p = 0.02$) consistent with the offspring of full-sibling matings having lower viability as predicted by inbreeding depression.

Table 3.1. Summaries of observed and expected frequencies among F1 offspring.

| | | | |
|---|---|---|---|
| **Number of matings** | | | |
| *Relatedness* | *Half sib* | *Full sib* | |
| Observed | 159 | 34 | |
| Expected[a] | 149.2 | 43.8 | |
| *MHC sharing* | *None* | *One haplotype* | *Both haplotypes* |
| Observed | 46 | 130 | 17 |
| Expected[a] | 56.5 | 117.3 | 19.1 |
| *MUP sharing* | *None* | *One haplotype* | *Both haplotypes* |
| Observed | 73 | 114 | 6 |
| Expected[a] | 71.8 | 100.5 | 20.7 |
| **Number of offspring** | | | |
| *Relatedness* | *Half sib* | *Full sib* | |
| Observed | 414 | 69 | |
| Expected[b] | 395.9 | 87.1 | |
| *MHC sharing* | *None* | *One haplotype* | *Both haplotypes* |
| Observed | 135 | 311 | 37 |
| Expected[b] | 124.8 | 310.6 | 47.7 |
| *MUP* | *None* | *One haplotype* | *Both haplotypes* |
| Observed | 159 | 314 | 10 |
| Expected[b] | 164.5 | 304.7 | 13.8 |
| **Offspring genotype** | | | |
| *MHC* | *Homozygous* | *Heterozygous* | |
| Observed | 94 | 389 | |
| Expected[c] | 96.3 | 386.8 | |
| *MUP* | | | |
| Observed | 77 | 406 | |
| Expected[c] | 83.5 | 399.5 | |

[a] Expected based on proportion of dyads of each type available for each female
[b] Expected based on matings of each type per female
[c] Expected based on parental genotypes and number of offspring per mating

We found no evidence of avoidance of matings between mice sharing either one or both MHC haplotypes (Figure 3.1 and Table 3.2: model 2, p =

0.27). In fact, there were more matings between parent pairs sharing one MHC haplotype, rather than fewer as expected with MHC disassortative mating. There was however a deviation from the expected number of matings between parent pairs sharing one or both MUP haplotypes (Figure 3.3 and Table 3.2: multinomial logistic model 3, p = 0.005). This was due to a deficit in successful matings between mice sharing both MUP haplotypes (Figure 3.3 and Table 3.2: multinomial logistic model 4, p = 0.002).

Table 3.2. Multinomial logistic models for frequency of mating.

| | β | Log likelihood | LR statistic[a] | df | P[b] |
|---|---|---|---|---|---|
| **Model 1: full sib avoidance** | | | | | |
| full sib | -0.36 | -491.07 | 3.42 | 1 | 0.08 (1 tailed) |
| **Model 2: MHC sharing** | | | | | |
| one MHC haplotype | 0.36 | -490.63 | 4.31 | 2 | 0.27 |
| both MHC haplotypes | 0.12 | | | | |
| **Model 3: MUP sharing** | | | | | |
| one MUP haplotype | 0.15 | -483.60 | 18.36 | 2 | 0.005 |
| both MUP haplotypes | -1.36 | | | | |
| **Model 4: full MUP sharing** | | | | | |
| both MUP haplotypes | -1.44 | -484.04 | 17.48 | 1 | 0.002 |
| **Model 5: relatedness and full MUP sharing** | | | | | |
| full sib | -0.12 | -483.86 | 17.84 | 2 | 0.65[c] |
| both MUP haplotypes | -1.38 | | | | |
| **Model 6: maternal MHC imprinting and full MUP sharing** | | | | | |
| one MHC haplotype match between male and female's mother | -0.33 | -482.61 | 20.34 | 2 | 0.18[c] |
| both MHC haplotypes | -1.31 | | | | |
| **Model 7: maternal MUP imprinting and full MUP sharing** | | | | | |
| one MHC haplotype match between male and female's mother | -0.11 | -483.76 | 18.04 | 2 | 0.54[c] |
| both MUP haplotypes | -1.35 | | | | |

[a] Compared to null model (log likelihood = -492.78)
[b] Probabilities calculated by random permutation of data (n = 10,000)
[c] Comparison of model 5 to model 4

The avoidance of mates with complete MUP sharing is enough to account for the deficit of matings between full sibs (Table 3.2: model 5 versus

model 4, $p = 0.65$) as full sibs are more likely to share both MUP types than half sibs (appendix Table 3.1).

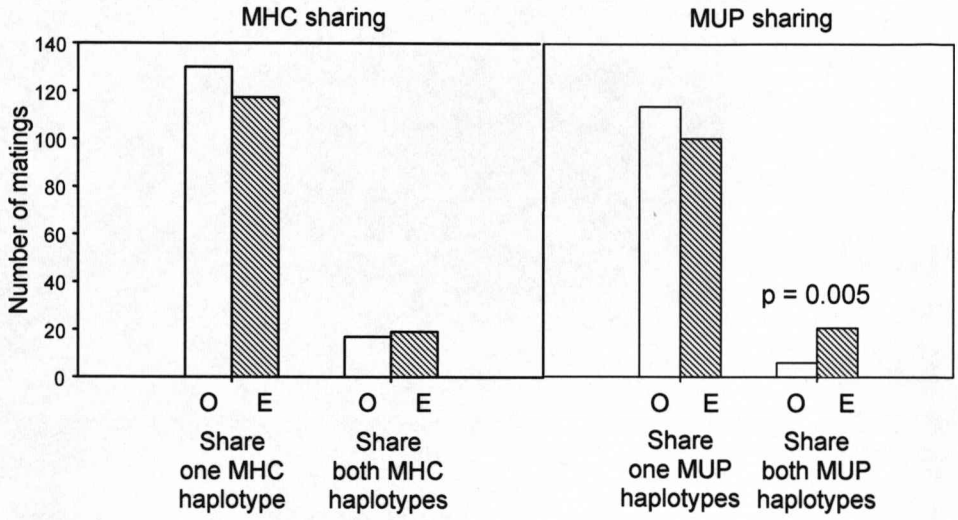



Figure 3.3. Comparison of observed (O) and expected (E) numbers of matings between parent-pairs sharing one or both, MHC or MUP haplotypes. The p value refers to multinomial logistic model 4. All other comparisons are non-significant.

We confirmed that the reduction in matings between full sibs was not due to lower male quality by comparing the difference in the overall mating success of males in full sib dyads (mean ± standard error of the mean matings with any female: $4.4 \pm 0.6$, $n = 23$) to the success of any other male ($3.8 \pm 0.5$ matings per male, $n = 25$; $t_{46} = -0.70$, $p = 0.49$). This shows that there was no difference in inherent male quality between the males in full sib dyads and those in other mating combinations, and that the significant difference in matings between parent pairs sharing both MUP haplotypes must be due to mating avoidance.

Familial imprinting has been hypothesised as a means to avoid mating with close kin i.e., litter mates would "learn" some genetic loci from their mother and avoid mating with animals carrying those alleles. All full sibs carry a maternal haplotype at the MHC, MUP and all other potential genetic markers, whereas paternal half sibs should not. Thus the effects of

maternal imprinting in our study would be manifested in a general avoidance of full sib matings. However, the inclusion of overall relatedness in the model describing full MUP sharing does not explain the matings seen any better than the use of MUPs alone (Table 3.2: model 5 versus model 4, $p = 0.65$). We looked specifically for evidence of imprinting on maternal MHC and MUP haplotypes (Table 3.2: models 6 and 7, $p = 0.18$ and $p = 0.54$), but neither explained the data better than avoidance of mates that match at both MUP haplotypes alone.

### 3.3.2.3 Analysis of the number of offspring per mating

There were fewer offspring per successful full-sib mating than expected by the null model (Table 3.3: model 1, $p = 0.02$). This is consistent with lower viability of offspring from full sibling matings through inbreeding depression or post-copulatory sperm or embryo selection. However, this reduction does not seem to be associated directly with MUP or MHC sharing (Table 3.3: models 2 and 3, $p = 0.25$ and $p = 0.52$). The deficit in the number of offspring produced from matings between animals sharing both MUP and both MHC haplotypes (Table 3.1) is attributable to the fact that most of the dyads in these categories were full sibs. Both MHC and MUP haplotypes showed Mendelian pattern of inheritance (Table 3.1).

Table 3.3. Multinomial logistic models for number of offspring per mating.

| | B | log likelihood | LR statistic[a] | df | P[b] |
|---|---|---|---|---|---|
| **Model 1: full sib avoidance** | | | | | |
| full sib (1 tailed) | -0.44 | -497.9 | 7.69 | 1 | 0.02[c] |
| **Model 2: MHC sharing** | | | | | |
| one MHC haplotype | -0.12 | -499.24 | 5.02 | 2 | 0.25 |
| both MHC haplotypes | -0.49 | | | | |
| **Model 3: MUP sharing** | | | | | |
| one MUP haplotype | 0.11 | -500.59 | 2.33 | 2 | 0.52 |
| both MUP haplotypes | -0.34 | | | | |

[a] Compared to null model (log likelihood = -501.75)
[b] Probabilities calculated by random permutation of data (n = 10,000)
[c] Model 1 would also be significant at $P < 0.05$ in a 2 tailed test

### 3.3.3 DISCUSSION

The inbreeding avoidance seen in this study can be completely explained by a deficit of matings between animals that fully share at MUP. There was no indication of an effect of MHC or background genes on matings. Mice appear to be using full MUP matching as a marker for high relatedness. There is no evidence that they are imprinting on the MUP haplotypes of their mother, so it is likely that they are using a self-referent method of comparing MUP. Sharing both MUP haplotypes is a good indicator of a close relationship in our study. 31% of full sibs and 5% of half sibs share both MUP haplotypes. Considering the percentages of mice sharing one MUP haplotype (45% of full sibs and 55% of half sibs) provides an explanation as to why we do not see a reduction in the numbers of matings between dyads that match at a single haplotype. If females were to avoid males that match at one MUP haplotype as well as two, they would be excluding a high proportion of half siblings without excluding many more full siblings. This situation is likely to be true in a natural population. MUP is highly polymorphic, so in an outbred population it is unlikely that half siblings, or even full siblings, would share both MUP haplotypes. This suggests that MUP based inbreeding avoidance may only have a marked effect in populations that were already relatively inbred and would generally be a weak pressure in normal situations. However, a weak discrimination against males that share both MUP haplotypes could still be enough to encourage male dispersal (Lehmann and Perrin 2003). Dispersal of young males from the natal area will have the effect of reducing matings between close relatives.

A full sibling is twice as likely as a half sibling to share deleterious recessive alleles that could be passed on to progeny, thus we would have expected to observe a significant deficit in the number of matings between full siblings. The reduction in matings seen is pronounced but not quite significant. This may be because the animals used to breed the founders were specifically chosen to be outbred. In this scenario, it may be advantageous for some animals to inbreed to avoid outbreeding depression. A slight level of inbreeding in a totally outbred population would reduce the chance of breaking up of any more coadapted gene complexes. Females may prefer a certain level of inbreeding as kin are likely to share genes, meaning there is a greater chance of maternal genes being passed onto the next generation. We may not see a significant level of inbreeding avoidance because there are two conflicting pressures at work in our experimental populations; the avoidance of inbreeding and the avoidance of outbreeding. Avoidance of inbreeding is obviously the stronger of the two pressures, probably because the least related individual available as a mate is a parental half sibling ($r = 0.25$), giving each population an inbreeding coefficient of at least $F = 0.125$, which is high enough to lead to effects of inbreeding depression (Durel et al. 1996).

My results were contrary to those from other studies examining the role of the MHC in mate choice. I find no correlation between MHC sharing between mates and inbreeding avoidance. Studies that test the ability of mice to recognise MHC encoded odour differences are not directly testing whether they use these differences for kin recognition, individual recognition or mate choice. It has been previously shown that repeated exposure to an odour can induce olfactory receptors that do not usually function (Wang et al. 1993), thus just because mice are capable of determining MHC encoded odour differences after training or repeated exposure, it does not mean that they are able to do so in the wild. Even if

they are, these studies do not show evidence of mate choice, simply preference. In a previous study that directly tested mate choice (Potts et al. 1991; Potts et al. 1994) MHC type was closely correlated with the rest of the genome. Thus the effects seen could be due to the use of another gene that is correlated to the MHC being used for inbreeding avoidance. Also, MHC heterozygosity correlates with genome-wide heterozygosity in these studies. Another interpretation of the results of the Potts' study (Potts et al. 1991) and others is that the females were choosing mates based on non-MHC markers of heterozygosity to promote the heterozygosity of their offspring.

We have shown that wild mice in naturalistic enclosures do not use MHC derived cues to avoid inbreeding. Instead, in our study, they are avoiding mating with potential mates that fully match at MUP. MUP expression appears to be species-specific (Robertson et al. 1996), and can even vary between *Mus musculus* sub-species (Robertson et al. 2007; Stopkova et al. 2007). This implies that there may be taxa-specific mechanisms for the avoidance of inbreeding, not just a vertebrate-wide marker such as MHC or MUP. We suggest that further work needs to be done in model systems and in the wild to further investigate the roles of the MHC, MUP and other potential signals of individual recognition.

## *3.4 Heterozygosity Study*

### 3.4.1 INTRODUCTION

In the second study I investigated the affects of MHC and MUP heterozygosity on the reproductive success of males. If females are

choosing MHC dissimilar males, offspring would be more heterozygous whether the driver is the promotion of heterozygosity per se or the avoidance of inbreeding. Heterozygosity itself is not heritable, but do the fitness gains associated with increased heterozygosity confer a greater reproductive success to individuals? Heterozygosity has been shown to increase many different fitness traits; survival (Kempenaers et al. 1996; Cordero et al. 2004), health (Acevedo-Whitehouse et al. 2003; Hawley et al. 2005), body size (Coltman et al. 1998; Merila et al. 2003), dominance relationships (Hoffman et al. 2004; Lieutenant-Gosselin and Bernatchez 2006) and reproductive success (Keller 1998; Sauermann et al. 2001). Increased reproductive success for heterozygotes may be correlated with other gains in fitness associated with heterozygosity, for example, if more heterozygous males hold superior territories, they are likely to be more attractive as potential mates. Meagher et al. (2000) showed that more heterozygous males mice sired significantly more offspring in semi-natural enclosures, largely as a result of their superior competitive ability. However, heterozygosity associated gains in reproductive success could also be due to mate choice favouring heterozygosity in a marker locus. Thom et al. (2008) showed that when intrasexual competition is minimised, female mice prefer to associate with MUP heterozygous over MUP homozygous males. To directly test the effects of heterozygosity on reproductive success we eliminated any correlations with other factors that influence female choice, such as established dominance relationships or the ability to defend a territory. This required a more controlled approach than that used in the enclosure study where males were allowed to interact freely. We ensured that any increased reproductive success seen in the second study was independent of competitive ability. We allowed wild females to mate with males who were either heterozygous or homozygous at the two signalling systems, MHC and MUP, or who were either heterozygous or homozygous across the rest of the genome. In each trial

type the other two factors were kept constant so individual effects of MHC, MUP and genome-wide heterozygosity on reproductive success could be differentiated.

## 3.4.2 MATERIALS AND METHODS

### 3.4.2.1 Animals

Females were from population of wild-derived house mice *Mus musculus domesticus* maintained in the laboratory. Wild-mice were caught from five locations throughout the northwest of England and populations maintained through laboratory breeding for up to five generations. Females were weaned into single-sex groups at 21 - 28 days. All females were sexually mature at the time of testing and were aged between 7 and 21 months (mean 15 ± 0.6 months). Some females had been previously used to test female preference for male heterozygosity. Females were allowed visual and olfactory contact with males but a barrier prevented physical contact and no mating was permitted. Three females were used twice in this study. In all three cases we ensured that the second trial was at least 8 weeks after the first and was tested a different preference using different males to avoid any influence of previous experience.

Males were from the F1 or F2 generation of a cross between two inbred strains (Balb/c and C57BL/6, Harlan-Olac, UK). These laboratory strains have been well characterised and are known to differ across the MHC (Balb/c is haplotype d, C75BL/6 haplotype b) and MUP regions (Balb/c is phenotype A, C75BL/6 is B (as defined by Snell (1958a; 1958b) and Fischer Lindahl (1997) for MHC and Robertson et al. (1996) for MUP). Thus the F1 offspring are heterozygous for both MHC (bd), and for MUP (AB) (Figure 3.4). We typed all F2s at both MUP and MHC. All MHC and

MUP genotyping was carried out by Francine Jury and the CIGMR group at the University of Manchester (see 3.3.1.3). We used pairs of males that were homozygous and heterozygous for MHC, while heterozygosity at MUP was held constant and vice versa. In these trials the level of heterozygosity across the rest of the genome was held constant. Thus male reproductive success would be correlated with MHC or MUP heterozygosity alone. We tested for effects of background heterozygosity by pairing MHC/MUP heterozygous F2 males with MHC/MUP F1 males which have double the genome-wide heterozygosity on average. This test mimics a relatively inbred and relatively outbred male, F2 and F1 respectively. However, two hypothesised signalling systems for genomic heterozygosity are identical in both.

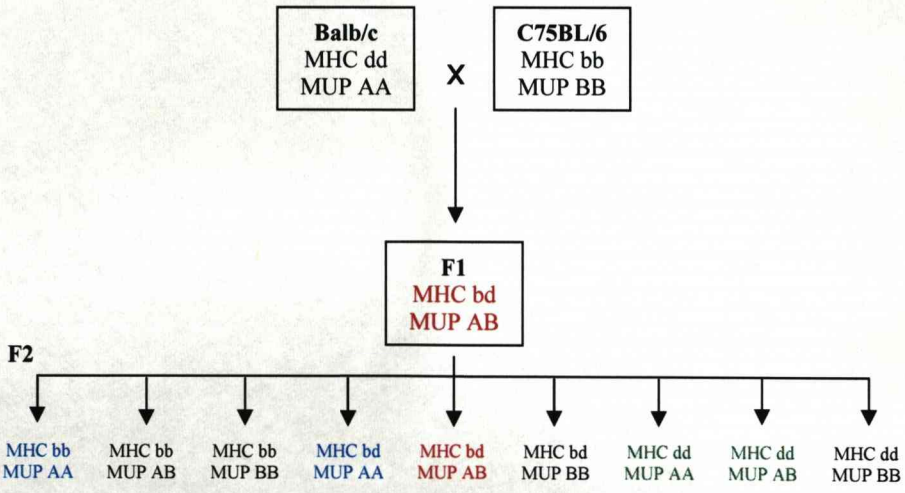



Figure 3.4. Breeding program of males used showing MHC and MUP types. The pair in blue show differing heterozygosity at MHC while MUP and background heterozygosity is controlled. The pair in green show the same but with MUP as the test region. Note these two types of trial can also be carried out with the region that is not being tested kept heterozygous in both. The pair in red show constant heterozygosity at MHC and MUP but differing levels of heterozygosity at the rest of the genome.

Males were singly housed at weaning (21 - 28 days) to control for social experience. At the time of testing males were aged between 7 and 20 months (mean 14 ± 0.5 months), and they were aged matched within each

pair as closely as possible (mean 5.1 ± 1day, maximum 25 days difference). Many of the males used had previously been used in an experiment examining the role of heterozygosity on female association with males according to male heterozygosity (Thom et al. 2008). Males were kept in their original pairings and the original females used where possible. Some males were used twice, but always in a different test and with males of equal experience.

#### 3.4.2.2 Mating trials

All trials were conducted by Michael Thom at Leahurst Veterinary Facility. The experimental arena consisted of three adjacent cages (each with internal area 960 $cm^2$ x 13 cm height, external dimensions 45 cm x 28 cm x 13 cm) with openings between the end cages and the middle cage that were too narrow to allow males to pass through but allowed the smaller female free movement throughout the three cages. Each male was placed in one of the outside cages and the female in the central cage. Thus the female could choose to pass into either of the male's cages or remain in the central cage that only she could access. This set up did not allow male-male competition so no dominance relationship could be established between the males. Each cage had nesting material inside and a wire lid with access to food (Lab Diet 5002 Certified rodent diet) and water. Mice were left undisturbed for 14 - 21 days, and then separated and pregnant females allowed to carry to parturition. Genetic samples were taken from the female, both males and the offspring and used for paternity assignments.

#### 3.4.2.3 Paternity Assignment

I extracted DNA from tail snips using mouse-tail extraction kits (Promega Wizard SV Genomic DNA Purification System). I used twelve of the primers previously used for parentage assignment in the enclosure study (see 3.3.1.4 Parentage analysis and Appendix Table 3). Amplification and

fragment size analysis was carried out according to 3.3.1.4 Parentage analysis. The parentage assignment program Cervus 3.0 (Kalinowski et al. 2007) was used to assign individuals to a parent pair using maximum-likelihood methods.

### 3.4.2.4 Scent marking and dominance testing

This study tests the reproductive success of male mice which have differing levels of heterozygosity at marker genes or across the genome. Previous studies have shown a link between reproductive success and male scent marking (Gerlinskaya et al. 2007) and dominance relationships (Defries and McClearn 1970). We checked for correlations between any reproductive skew seen and dominance tendencies and scent marking. Males were tested for scent marking in their experimental parings. They were placed either side of a mesh divide in a cage (internal area 960 $cm^2$ x 13 cm height, external dimensions 45 cm x 28 cm x 13 cm) giving them visual and olfactory, but not physical contact. The bottom of the cage was lined with Benchkote. Males were allowed to interact for one hour, after which urine marks on the Benchkote were measured under UV light using image analysis software (Scion Image or ImageJ). Both the number of marks and the area of marks were measured. Number of marks and area of marking is roughly correlated, however frequency is a more reliable measure as a subordinate male can leave a large area of urine in a non-aggressive single marking; giving a high area score, but a frequency score of one.

Immediately before introduction into the experimental arena the pair of males was briefly tested for aggressive behaviour. Males were placed in a handling arena for 15 minutes and observed. The male that showed the more aggressive behaviour was labelled as dominant. In many cases (n = 16 from a total of 44 trials) there was no interaction between the males and

neither male was recorded as dominant. It should be noted that this test is a quick measure of aggressive behaviour and does not allow enough of an interaction to allow a proper dominant-subordinate relationship to be formed.

### 3.4.3 RESULTS

#### 3.4.3.1 Parentage assignment

Out of 84 trials, 50 resulted in surviving litters (59.5%). This was consistent across trial types (MHC: 62.0%; MUP: 58.5%; background: 57.5%). Three litters had to be excluded from the dataset because no useable DNA remained after storage. A further litter was excluded from the MHC trials as a male passed through the tunnel into the female cage during the trial. Litter size was fairly constant across trials (MHC: 5.3 ± 0.4 (range 3 - 8); MUP: 4.5 ± 0.4 (range 2 - 7); background: 4.5 ± 0.4 (range 1 - 7)).

The total number of offspring sired by males was independent of their heterozygosity at both MUP ($\chi^2 = 0.143$, $p = 0.705$) and MHC ($\chi^2 = 0.316$, $p = 0.574$). In contrast, genome-wide heterozygosity appeared to have an effect on reproductive success, with the more heterozygous males siring twice as many offspring as more homozygous males ($\chi^2 = 8.471$, $p = 0.004$) (Figure 3.5 and Table 3.4).

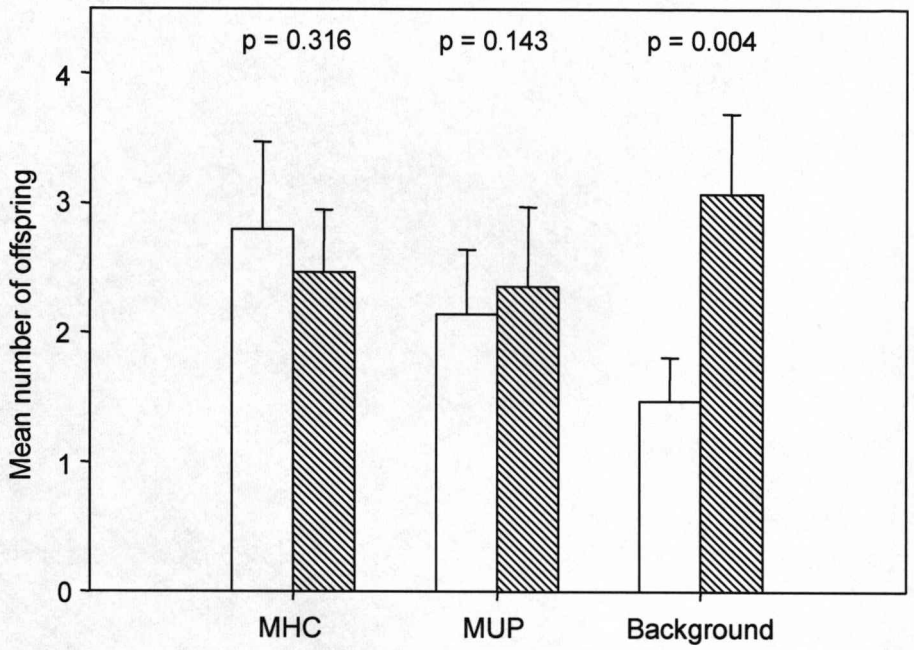



Figure 3.5 Number of offspring sired by homozygous (white bars) and heterozygous males (hatched bars) across the three different trial types. Results are mean and S.E.. P values indicate significance in $\chi^2$ tests.

Table 3.4 Total number of offspring sired across all litters (for MHC n = 15 litters, for MUP n = 14 litters and for background n = 15 litters). Split according to heterozygosity, weight, scent marking and dominance. Shown according to trial type (MHC and MUP are combined for weight, dominance and scent marking as there is no influence of the two marker systems on reproductive success, thus they can be considered equivalent).

| **Trial** | **Total number of offspring sired** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Heterozygosity** | | **Weight** | | **Dominance** | | **Marking** | |
| | **Hetero** | **Homo** | **Heavier** | **Lighter** | **Dom** | **Sub** | **High** | **Low** |
| **MHC** | 37 | 42 | | | | | | |
| **MUP** | 33 | 30 | | | | | | |
| **Background** | 46 | 22 | 44 | 24 | 23 | 19 | 34 | 34 |
| **MHC+MUP** | | | 87 | 13 | 69 | 41 | 46 | 80 |

Differences in the total number of offspring produced will be affected by the reproductive success of the females. Thus if a highly fecund female chooses to mate with a more heterozygous male and a less fecund female chooses to mate with a more homozygous male, the total number of

offspring produced would favour the more heterozygous male even though there is no overall mating bias. To test male affects alone we compared the proportion of offspring sired by each male per female. The proportion of offspring produced across all trials is independent of MHC ($\chi^2 = 0.021$, $p = 0.885$), MUP ($\chi^2 = 0.010$, $p = 0.920$) and background heterozygosity ($\chi^2 = 0.565$, $p = 0.452$). Again the trend is towards the more overall heterozygous male siring the most offspring but the difference is not big enough to see a significant effect.

As we found no effect of MHC or MUP heterozygosity on total number of offspring sired, we combined these data to look for other factors affecting offspring production. Heavier males sired more offspring than lighter males ($\chi^2 = 5.88$, $p = 0.015$ for background heterozygosity trials, $\chi^2 = 10.618$, $p = 0.001$ for MHC and MUP data combined). Mass was independent of genome-wide heterozygosity (heterozygous male heavier in 10 out of 15 pairs; binomial test: $p = 0.3$) (Table 3.4). It appears that reproductive success is affected by male size, as well as heterozygosity at background loci. Dominance behaviour and frequency of scent marking had no effect on offspring production in the background heterozygosity trials ($\chi^2 = 0.381$, $p = 0.537$ for dominance, the number of offspring produced by high frequency markers and low frequency markers was identical). However, both dominance and frequency of scent marking had an effect on offspring production in the MHC and MUP trials ($\chi^2 = 7.127$, $p = 0.008$ for dominance; $\chi^2 = 10.452$, $p = 0.001$ for frequency of scent marking) (Table 3.4). Thus, when background heterozygosity is controlled for, the effects of other factors affecting reproductive success are significant. This suggests that overall genetic heterozygosity is a greater determinant of reproductive success than dominance or scent marking.

### 3.4.4 DISCUSSION

When the two signalling systems, MHC and MUP, were kept constant we found an effect of genome-wide heterozygosity on offspring production, with the more heterozygous males producing significantly more offspring. When genome-wide heterozygosity was controlled for there was no effect of heterozygosity at MHC and MUP on offspring production. This suggests that reproductive success was determined by heterozygosity at a locus, or loci, other than the MHC and MUP.

Weight was a significant factor in reproductive success of males in all trials. On average, the heavier male sired significantly more offspring than the lighter. This supports previous findings in rodents (Defries and McClearn 1970) and suggests that male size is an honest signal of genetic quality and potential reproductive success.

Other factors tested had differing influences in different trial types. In the genome-wide heterozygosity trial, where heterozygosity affected offspring production, dominance and scent marking frequency had no affect on the number of offspring sired. In the MHC and MUP trials, where heterozygosity had no affect, the more dominant male and the male who scent marked with the lower frequency sired significantly more offspring. This suggests that in the absence of genome-wide heterozygosity the weaker influences of dominance and scent marking have an effect on reproductive success. Male dominance has been shown to have a positive affect on reproductive success in rodents and other species (Franks and Scovell 1983; Lamprecht 1986; Huck et al. 1988; Pusey et al. 1997) and our study adds to that data. However, our finding that low scent marking males sired more offspring than high scent marking males conflicts with previous work. Urinary scent marks differ qualitatively between dominant and subordinate males and are produced at higher frequencies in dominants

(Desjardins et al. 1973). Scent marking is costly, and marking at high frequency by dominant males may thus be seen as 'honest' signalling of quality. It has been shown that females prefer the odour of males whose marking patterns suggest they are of high quality (Rich and Hurst 1998). Our use of inbred strains in the generation of the males in these tests may account for our novel result. Nevison et al. (2000) compared differing scent marking behaviours of inbred strains, and presumably these behaviours would also be different to those found in the wild. This suggests that our finding that low scent marking males sire more offspring than high scent marking males (when heterozygosity across the genome is controlled for) may not be relevant in a natural population of house mice.

Reproductive success can be affected by both pre- and post-copulatory mechanisms (Simmons 2005). Our study is not capable of separating the two. We are unable to say whether genome-wide heterozygosity acts as a mate choice cue or increases sperm quality. Indeed, pre- and post-copulatory choices may differ in species such as *M. musculus* that engage in extra-pair copulation (Firman and Simmons 2008); females may choose to nest with males that offer superior territories and resources, but may put more investment in sperm from males with more compatible genes (Bretman et al. 2004). This supports the finding of Rolland et al. (2003) that, when given free access to males, female mating preference was not reflected in any other behaviour (such as time spent with the males when not in oestrus). A previous study using some of the same male pairings found that females preferred to associate with MUP heterozygous males, but not with males with greater background heterozygosity (Thom et al. 2008). This preference was not translated into offspring production in the current study, suggesting there may be differences in pre- and post-copulatory mating choices in our system. Genome-wide heterozygosity has been shown to affect sperm quality in rabbits (Gage et al. 2006), with the

more heterozygous males producing better quality sperm and this could be the case in the mice in this experiment.

The enclosure study showed that mice avoid mating with potential mates that fully match at MUP, and that the MHC has no influence on mate choice. In the second study we found no influence of heterozygosity at MHC or MUP on the reproductive success of males, however, heterozygosity at background genes increased reproductive success. Combining the results of both studies indicates that; i) the MHC has no affect on the avoidance of inbreeding or on reproductive success in our experiments, ii) MUP heterozygosity has no affect on offspring production when male-male competition is prevented and iii) intraspecific MUP sharing influences mating. It is unclear whether the avoidance of inbreeding facilitated by full MUP sharing is a bigger driver of reproductive success than the affects of genome-wide heterozygosity, and it would take a complex experimental set-up to test that theory.

# Chapter 4

# EVIDENCE OF BALANCING SELECTION AT THE MHC IN A NATURAL POPULATION

## 4.1 INTRODUCTION

Selection at the MHC has been widely studied (Hughes and Nei 1988; Hughes and Nei 1989; Nevo and Beiles 1992; Takahata et al. 1992) as MHC genes have one of the highest rates of polymorphisms in the mammalian genome (Klein 1986). Hypotheses proposed for maintenance of this polymorphism include; an unusually high mutation rate (Bailey and Kohn 1965); gene conversion (Widera and Flavell 1984); disassortative mating (Reusch et al. 2001) and, with by far the most support, balancing selection (Doherty and Zinkernagel 1975; Klein 1986; Takahata et al. 1992; Paterson 1998).

Balancing selection is a form of natural selection that maintains multiple alleles within a population. The two main mechanisms that lead to this form of balanced polymorphism are overdominance and rare allele advantage, both of which could be at work in the MHC. The overdominance theory, as applied to the MHC, states that individuals that are heterozygous at MHC loci will have a greater chance of recognising pathogenic antigens than individuals that are homozygous and thus MHC heterozygous individuals will incur less of a reduction in fitness due to infections than homozygotes and will persist in the population (Doherty and Zinkernagel 1975; Maruyama and Nei 1981). The trans-species hypothesis of the MHC (first proposed by Klein (1980)) which explains the

persistence of certain MHC alleles through speciation events, also suggests that alleles have been maintained by overdominant selection. Rare allele advantage, or frequency dependent selection, states that individuals whose phenotype is rare within the population will be at a selective advantage. In the case of the MHC and host-pathogen co-evolution, a rare or new allele that a pathogen has not encountered or evolved to evade, will give a selective advantage over common alleles in the population.

Many studies support the hypothesis that pathogen-driven balancing selection acts to maintain allelic diversity at the MHC (Potts et al. 1994; Potts and Slev 1995b; Apanius et al. 1997a; Wegner et al. 2003; Wegner et al. 2004; Meyer-Lucht and Sommer 2005). However, the effects of mate choice at the MHC could also result in balancing selection. Hedrick (1992) showed that female choice for males which differ at a particular locus, termed disassortative mating, will lead to a reduction in the numbers of homozygotes in a population and can lead to alleles being maintained at equal frequencies.

House mice, *Mus musculus domesticus*, have been well characterised at the MHC through the use of inbred strains (Arnold et al. 1984; Watts et al. 1987; Saha 1996; Melvold et al. 1997; Kellenberger et al. 2005), but surprisingly, little has been done to characterise diversity in natural populations. Duncan et al. (1979) used serological data to investigate heterozygosity in a populations in Texas, USA. They found remarkably high levels of MHC heterozygosity (the mean frequency of heterozygotes across the five MHC genes tested was 96%), compared to much lower levels at all other polymorphic loci (6%). Duncan et al. suggest that the near 100% H2 heterozygosity is caused by heterozygote advantage at the H2 loci. I aim to further this work by characterising sequences of two MHC loci from house mice from a wild population, and by looking for signatures of balancing selection acting at the MHC within the population.

## 4.2 AIMS

I tested for evidence of balancing selection at the MHC in a natural population of *Mus musculus domesticus*. Contemporary selection can be detected by deviations from Hardy-Weinberg equilibrium. Population demographics such as recent bottlenecks, or sampling of meta-populations can also cause deviations from allele frequencies expected under Hardy-Weinberg, thus I used 12 genome-wide markers to investigate recent demographic effects within the population. I used MHC-linked markers to measure allelic frequencies and Watterson's statistic test to investigate deviations from neutrality. I sequenced a class I gene (H2-K1) and a class II gene (H2-Ab) to measure the rates of synonymous and nonsynonymous substitution at both the antigen binding sites and structural residues. Comparison of the rate of the two types of substitution was used to detect evidence of selection, and Tajima's D statistic was calculated to test for deviation from neutral evolution using allele frequencies and rates of mutation simultaneously.

## 4.3 MATERIALS AND METHODS

### 4.3.1 Animals

Winsford pig farm is a 35 acre farm in Cheshire, North West England, consisting of three large pens and a dozen feed buildings surrounded by woodland. 62 wild house mice (*Mus musculus domesticus*) were trapped in and around the buildings in seven trapping sessions between September 2003 and January 2005. Seven of the females were pregnant at the time of capture and gave birth in captivity, meaning genetic samples were available for 75 individuals in total. Traps were laid and collected by the

owner of the farm, thus no data about location and exact date of capture are known.

### 4.3.2 Genotyping

DNA was extracted from 1 - 5 mm tail snips with mouse-tail extraction kits (Promega Wizard SV Genomic DNA Purification System). Five markers within the MHC region (chromosome 17) and one marker on either side had previously been selected using the ENSEMBL mouse genome database (Hubbard et al. 2007) (see 3.3.1.2 MHC and MUP genotyping, Appendix Table 1 and Appendix Figure 1 for details). The sequences of primers for 12 genome-wide microsatellite markers had previously been chosen from the Mouse Genome Informatics site (MGI 3.51) (see 3.3.1.4 Parentage analysis and Appendix Table 3 for details). PCR amplifications for the MHC and genome-wide markers were carried out in 10 µl reactions according to the conditions outlined in 2.3.2 MHC genotyping and 3.3.1.4 Parentage analysis respectively. Fragment size was determined with an ABI PRISM 3100 DNA analyzer (Applied Biosystems) and GeneMapper v3.0 software designed for microsatellite analysis (Applied Biosystems).

Genepop 4.0 (Raymond and Rousset 1995) was used to calculate observed and expected heterozygosity ($H_O$ and $H_E$) at genome-wide and MHC microsatellites. In this case expected heterozygosity is calculated under Hardy-Weinberg equilibrium; $H_E = 1 - \sum_{i=1}^{k} p_i^2$ where $p_i$ is the frequency of the *ith* of *k* alleles. Deviations from Hardy-Weinberg equilibrium were investigated using the Markov chain method to evaluate exact tests (Guo and Thompson 1992). Genepop was also used to estimate linkage disequilibrium (LD) for markers within and surrounding the MHC region, and as a control, for the genome-wide markers which were not expected to be in linkage disequilibrium. The null hypothesis is that genotypes at one

locus are independent from genotypes at a second locus. Genepop computes unbiased estimates by randomisation (5,000,000 iterations) and by the Markov-chain method for the exact probabilities of random association in contingency tables corresponding to all possible pairs of loci.

### 4.3.3 Population structure

The 12 genome-wide microsatellites were used to investigate the demography of the Winsford population. I used Structure 2.2 (Pritchard et al. 2000) to test whether the 75 individuals captured belonged to a single panmictic population, or a number of metapopulations. Structure uses a model-based clustering method to infer population structure and assign individuals to populations using multilocus genotype data. $10^6$ Markov chain Monte Carlo iterations were run after a burn-in period of 100,000 iterations for population numbers (termed K) from 1 to 5.

Tests considering population structure and evidence of selection often use levels of heterozygosity at selectively neutral loci, such as the genome-wide microsatellite markers used in this study. A recent reduction of effective population size, termed a 'bottleneck', would result in heterozygosity excess at selectively neutral loci, biasing results. The program Bottleneck 1.2.02 was used (Cornuet and Luikart 1996) to test for evidence of a recent bottleneck in the Winsford population. Observed heterozygosity $H_O$ was compared to expected heterozygosity $H_{E\text{-}S.M.M.}$ as calculated from simulations of gene genealogies, with sample size and allele number identical to those in the data, under the stepwise mutation model (*S.M.M.*). The S.M.M assumes that mutations produce a directional change in length, with an increase (or decrease) in microsatellite length each mutation. The test was also done using the expected heterozygosity calculated under the infinite alleles model ($H_{E\text{-}I.A.M.}$). The infinite alleles model assumes that each mutation that arises is unique. The standard

deviate ($[H_O-H_{E\text{-}S.M.M.}]/SD$) is used to test for significantly inflated values of $H_O$ that suggest rapid population expansion following a bottleneck.

The parentage assignment program Cervus 3.0 (Kalinowski et al. 2007) was used in an attempt to assign individuals to a parent pair using maximum-likelihood methods. As the extent of successful parentage assignments was limited, I used Identix 1.0 (Belkhir et al. 2002) to estimate mean pairwise relatedness, $r_{xy}$. I used the estimator derived by Queller and Goodnight (1989), where $x$ is a reference individual, with alleles $a$ and $b$ at a single locus, and $y$ is the individual to be compared, with alleles $c$ and $d$. If the reference individual is homozygous $S_{ab} = 1$, if heterozygous $S_{ab} = 0$. Similarly, if allele $a$ from the reference is the same as allele $d$ from the individual being compared $S_{ad} = 1$, if the alleles are different $S_{ad} = 0$. Letting $p_a$ and $p_b$ be the frequencies of the alleles a and b in the population;

$$r_{xy} = \frac{0.5(S_{ab} + S_{ad} + S_{bc} + S_{bd}) - p_a - p_b}{1 + S_{ab} - p_a - p_b}$$

To calculate the null distribution of no relatedness Identix uses conventional Monte Carlo resampling (Guo and Thompson 1992), randomly selecting either 2 N alleles without replacement, independently for each locus, assigning them to N individuals, then recalculating the statistic.

Observed and expected heterozygosities ($H_O$ and $H_E$) were used to calculate the within population fixation index, $F_{IS} = (H_E - H_O) / H_E$. Any difference between values of $F_{IS}$ for the MHC-linked and the non-linked markers would represent putative evidence of contemporary selection. I used Watterson's homozygosity test of neutrality (Watterson 1978) to examine evidence of balancing selection at genome-wide and MHC-linked microsatellites. The test statistic $F$ is the sum of the frequencies of

homozygotes under Hardy-Weinberg equilibrium, $F = \sum_{i=1}^{n} p_i^2$, where $i$ represents each allele. Significance is tested against values of $F$ expected for populations under neutrality given a number of alleles (K) and population size (2n) such that;

$$F\ deviate = \frac{F_{(obs)} - F_{(\exp)}}{\sqrt{\mathrm{var}(F_{(\exp)})}}$$

where $var(F_{(exp)})$ = the variance of the expected frequency of homozygotes (Ewens 1972). A positive value of *F deviate* indicates positive selection resulting in a shortage of heterozygotes, a negative value of *F deviate* indicates alleles maintained at equal frequencies within a population, a result of balancing selection. Watterson's test assumes the infinite alleles model of mutation, from which microsatellites are thought to deviate. The genome-wide markers act as a control as they are assumed to be under no selection pressure, so the relative comparison between the MHC and genome-wide markers will provide evidence for selection pressures acting, regardless of whether microsatellites follow the pattern of mutation assumed by the test or not.

### 4.3.4 Sequencing

DNA was extracted from 1 - 5 mm tail snips with mouse-tail extraction kits (Promega Wizard SV Genomic DNA Purification System). Sequences for the MHC class I gene H2-K1 (equivalent to HLA-A in humans) and the MHC class II gene H2-Ab (equivalent to the widely studied human DQB gene) were downloaded from NCBI for several different inbred mouse strains. Alignments of these sequences were used to design oligonucleotide primers (MWG-Biotech [London, UK]) to amplify exons two and three of H2-K1 (H2-K1 exon 2; fwd 5'- TGA GGT GGT CCG GGT CTC ACC-3'; rev 5'-GTG GAG GGG TCG TGA CCT CCG A-3', H2-K1 exon 3; fwd

5'- TGA GCG GGG CTG ACC GCG-3'; rev 5'- CAG GAC TGA GCC CCA GCC-3') and exon two of H2-Ab (H2-Ab exon 2; fwd 5'- GAG CCG TGG TCT GCG TGG TC-3'; rev 5'- GGA GGA AGG AGG CCA TGC AG-3'). These exons were chosen because they code for the antigen binding sites of the MHC molecules.

PCR amplifications were carried out separately for each exon using a random sample of 46 individuals from the Winsford population. PCRs were conducted in 30 μl reactions containing 30ng of DNA, 1.5μM of each primer and 15.0 μl of 2X BioMix Red reaction mix (Bioline [London, UK]). Plates were incubated in a thermal cycler (Peltier) on a touchdown PCR program beginning with 95°C for 2 min, followed by 95°C for 20s, 60°C for 30s, minus 0.25 °C per cycle for 16 cycles, and then 56°C for 20 cycles and 72°C for 30s, with a final extension at 72°C for 10 min. PCRs were purified using Qiagen MinElute PCR kits and pooled into eight groups (six groups of six and two groups of five) and each group was ligated into T-vectors and grown in *Escherichia coli* competent cells using Promega's pGEM-T Easy Vector System. Twelve white colonies were selected from each plate of cells and incubated in 3 ml of LB media overnight (16 - 24 hours). Plasmids were purified from *E. coli* cells using Promega's Wizard *Plus* SV Minipreps DNA Purification System and then cleaned using 100 μl Shrimp Alkaline Phosphatase (SAP, Amersham), 142.5 μl SAP dilution buffer and 7.5 μl Exonuclease I (New England Biolabs) per 96-well plate, heated in a thermal cycler for 40 min at 37 °C, followed by 15 min at 80 °C. Plasmid concentration was estimated using a NanoDrop ND-1000 Spectrophotometer and diluted to give a concentration of 100 ng/μl and 10 μl was sent to Macrogen (Seoul, Korea) for sequencing. Sequences were checked and aligned using BioEdit (Hall 1999). This technique generated a suite of allele sequences from the Winsford population. To assign alleles to individuals I typed the mice at

seven microsatellite markers located in and around the MHC region (Appendix Table 1). MHC haplotypes were determined using the seven known family groups to trace inheritance of these markers and direct sequencing of MHC genes confirmed correct assignments. Direct sequencing involved PCR amplification as previously outlined. PCR products were then purified using Qiagen MinElute PCR kits, DNA concentration estimated and diluted to 16 ng/μl and sent to MWG-Biotech (London, UK), with associated primer at a concentration of 2 pmol/μl, for sequencing. Once haplotypes had been determined, further direct sequencing allowed alleles to be assigned to haplotypes and thus to individuals.

### 4.3.5 Selection tests

The rate of synonymous substitutions per potential synonymous site ($\bar{d}_S$) and the rate of nonsynonymous substitutions per potential nonsynonymous site ($\bar{d}_N$) was analysed in MEGA 4.0 (Tamura et al. 2007). The modified Nei-Gojobori method was used (Nei and Gojobori 1986; Zhang et al. 1998). This assumes the rate of transitional substitutions is different to the rate of transversional substitutions and weights potential substitution sites accordingly (assumed transition/transversion bias = 2). A Jukes-Cantor correction (Jukes 1969) was used to correct for multiple hits. I further investigated the rates of substitution by defining the nucleotide sites coding for the amino acids involved in antigen binding and those coding for structural residues. The amino acids involved in antigen binding were chosen based on the three-dimensional crystal structure of the MHC molecule (Fremont et al. 1992 for H2-K1; Zhu et al. 2003 for H2-Ab1).

The Tajima test statistic, *D* (Tajima 1989), was estimated using MEGA.

$$D = \frac{\pi - \theta / a_1}{[V(\pi - \theta / a_1)]^{1/2}}$$

where $\pi$ is the average number of pairwise nucleotide differences between sequences, $\theta$ is the observed number of polymorphic sites, and $a_1 = 1 + 2^{-1} + 3^{-1} +...+ (m - 1)^{-1}$ where $m$ = the number of sequences. This method of detecting deviation from neutrality compares the number of segregating sites and the observed number of nucleotide differences. As deleterious mutants are maintained in a population at low frequency, they are likely to have a negligible affect on the number of nucleotide differences. However, deleterious mutants would act to inflate the number of polymorphic sites, as this ignores the frequency of mutants. This would lead to a decrease in Tajima's test statistic $D$ ($D < 0$) from that expected in a selectively neutral population in mutation-drift equilibrium ($D = 0$). In contrast, balancing selection maintains alleles at high frequencies, thus the pairwise nucleotide diversity would be enhanced, but the number of polymorphic sites per sequence would not, leading to an increase in $D$ ($D > 0$). The mean and variance of Tajima's $D$ are zero and one, but Tajima (1989) showed that $D$ does not follow the normal distribution and can more accurately be described by the beta distribution. Thus the significance of Tajima's $D$ is computed using the beta distribution. Tajima's test of neutrality depends on the assumption that the population has been in mutation-drift balance for a long evolutionary time period. This is unlikely to be true of most natural populations; care must be taken when testing for significance as $D$ can be significantly negative or significantly positive due to the population history.

## 4.4 RESULTS

### 4.4.1 Genotyping

43 females and 32 males were typed at 12 genome-wide and 7 MHC-linked microsatellite markers. The percentage of typed genotypes was 81% for the genome-wide markers and 88% for the MHC markers. Repeat typing of 17 individuals (22% of individuals typed) across all loci produced 4 mismatches, an error rate of 0.3% per genotype.

Genepop was used to investigate linkage between markers within and around the MHC region. There was no evidence of linkage disequilibrium in any of the pairwise comparisons between genome-wide markers. All pairs of markers within the MHC region showed significant linkage disequilibrium. A significant p value was returned between a marker outside the MHC region (D17Mit192) and the marker nearest to it within the MHC (D17Mit230). The distance between these two markers, 0.10 cM, is relatively small explaining the result. The marker on the other side of the region was 15.08 cM from the nearest marker within the MHC, and showed no evidence of linkage disequilibria.

Table 4.1 Linkage disequilibrium for pairwise comparisons of MHC-linked markers. Significant linkage is shown in bold.

| Loci compared | | Distance between loci (cM) | P value | S.E. |
|---|---|---|---|---|
| **D17Mit192** | **D17Mit230** | 0.10 | **0.040** | **0.010** |
| **D17Mit230** | **D17Mit21** | 0.45 | **0.019** | **0.006** |
| **D17Mit21** | **D17Mit231** | 0.17 | **0.020** | **0.006** |
| **D17Mit231** | **D17Mit13** | 0.23 | **0.029** | **0.005** |
| **D17Mit13** | **D17Nds3** | 0.03 | **< 0.001** | **< 0.001** |
| D17Nds3 | D17Mit20 | 15.08 | 0.806 | 0.022 |

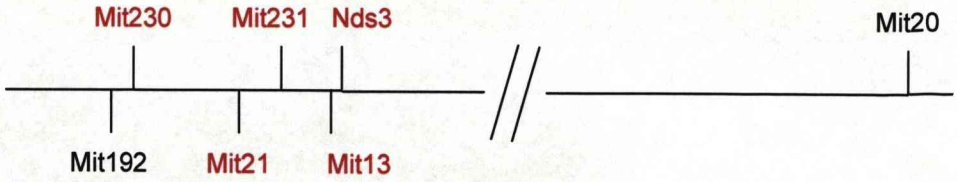



Figure 4.1 Scale map positions of microsatellite markers within the MHC region (in red) and outside the region.

### 4.4.2 Population structure

Structure 2.2 was run using the genome-wide microsatellite data to determine if we had sampled a single population or several smaller metapopulations. The log likelihood values and posterior probabilities (Table 4.2) indicate that the model in which K = 1 fits the data much better than any higher values of K. Thus, data from the animals sampled in the Winsford farm are consistent with that from a single panmictic population.

Table 4.2 Log likelihood values and probabilities for population numbers from K = 1 to K = 5.

| K | log P(X\|K) | P(K\|X) |
|---|---|---|
| 1 | -1493 | 0.999 |
| 2 | -1513 | 0.00021 |
| 3 | -1554 | 0.00003 |
| 4 | -1703 | ~ 0 |
| 5 | -1735 | ~ 0 |

There was no evidence of a recent population expansion characteristic of the period immediately following a bottleneck; in a two-tailed Wilcoxon test with the null hypothesis of a population in mutation-drift equilibrium using the S.M.M. p = 0.577, using the I.A.M. p = 0.123.

The 12 genome-wide microsatellite markers were used to assign parentage using the program Cervus 3.0. No information was known about the age of

mice upon capture so they could not be assigned to generations, meaning few parent pairs could reliably be assigned to offspring. The paternity of three of the seven litters born in captivity could be assigned at the 95% confidence level. Of the remaining offspring with no prior parentage information, one could be assigned to both parents at the 95% confidence level, five at the 80% level. One of the offspring assigned parents in this manner was a female that was pregnant upon capture, meaning we can verify that we have genetic samples from at least 3 generations of the population.

Relatedness between individuals is normally calculated from pedigree data. As there is no pedigree available for the Winsford population, I used the program Identix 1.0 to estimate the mean pairwise relatedness, $r_{xy}$. The distribution of $r_{xy}$ is shown in Figure 4.2. The majority of pairwise comparisons show relatedness equivalent to that of half siblings ($r_{xy} \sim 0.25$), however there is a left skew, showing that a few of the individuals caught were unrelated. Identix found a significant deviation from the null hypothesis of no relatedness (mean $r_{xy} = 0.249$, standard deviation = 0.093, $p = 0.023$). The maximum relatedness value was 0.589, greater than the $r_{xy} = 0.5$ value expected for a parent-offspring or full-sibling relationship. This suggests that a proportion of the animals we have sampled belong to family groups that may have a level of inbreeding in their ancestory.

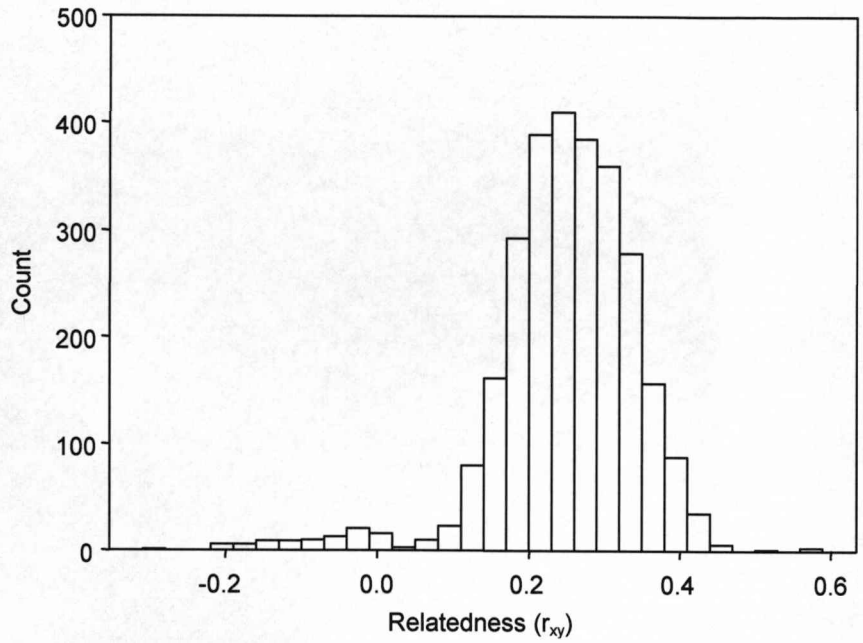



Figure 4.2 Distribution of pairwise $r_{xy}$ values among 75 wild-caught individuals.

### 4.4.3 Heterozygosity tests of selection

A summary of the microsatellite data can be found in Table 4.3. One of the genome-wide microsatellites, D3Mit1, was monomorphic for allele size 120, making it uninformative for further tests. Three genome-wide and two MHC markers show allele sizes just outside those characterised in inbred strains, but only by one or two di-nucleotide repeats (MGI 3.51). The mean number of alleles was lower for genome-wide markers (mean K = 4.33) than MHC-linked markers (mean K = 7.86), although this difference was non-significant ($\chi^2 = 1.01$, $p = 0.313$). Mean heterozygosity was also non-significantly lower in genome-wide markers (mean $H_O = 0.518$) than in MHC-linked markers (mean $H_O = 0.810$) ($\chi^2 = 0.06$, $p = 0.800$). No markers show any significant deviation from levels of heterozygosity expected under Hardy-Weinberg.

Table 4.3. Population data for the 12 genome-wide and 7 MHC liked microsatellite markers. Hardy-Weinberg probabilities (as calculated by exact test in Genepop) are all non-significant. $F_{IS}$ values that are significantly different from zero at the 95% level are shown in bold.

| Locus | N | Allele size range | K | $H_O$ | $H_E$ | HWE prob. | $F_{IS}$ |
|---|---|---|---|---|---|---|---|
| | | | Genome-wide markers | | | | |
| D1Mit58 | 72 | 258 - 280 | 5 | 0.486 | 0.460 | 0.817 | -0.057 |
| D2Mit405 | 75 | 78 - 86 | 3 | 0.453 | 0.516 | 0.064 | 0.123 |
| D3Mit1 | 73 | 120 | 1 | - | - | - | - |
| D4Mit171 | 74 | 117 - 127 | 4 | 0.541 | 0.613 | 0.359 | 0.119 |
| D5Mit25 | 70 | 204 - 240 | 6 | 0.471 | 0.502 | 0.666 | 0.062 |
| D10Mit80 | 74 | 136 - 148 | 6 | 0.230 | 0.238 | 0.123 | 0.033 |
| D10Mit98 | 73 | 152 - 170 | 3 | 0.479 | 0.507 | 0.359 | 0.054 |
| D11Mit63 | 73 | 130 - 146 | 4 | 0.671 | 0.703 | 0.634 | 0.045 |
| D11Mit300 | 74 | 138 - 152 | 6 | 0.473 | 0.579 | 0.058 | **0.183** |
| D12Mit4 | 72 | 198 - 212 | 6 | 0.694 | 0.767 | 0.114 | 0.095 |
| D15Mit161 | 34 | 84 - 114 | 4 | 0.735 | 0.676 | 0.119 | -0.068 |
| D19Mit70 | 30 | 215 - 229 | 4 | 0.467 | 0.578 | 0.281 | **0.195** |
| | | | MHC markers | | | | |
| D17Mit192 | 70 | 286 - 300 | 5 | 0.671 | 0.710 | 0.526 | 0.055 |
| D17Mit230 | 70 | 278 - 308 | 10 | 0.900 | 0.881 | 0.989 | -0.021 |
| D17Mit21 | 43 | 155 - 179 | 6 | 0.837 | 0.770 | 0.917 | -0.089 |
| D17Mit231 | 68 | 285 - 311 | 7 | 0.882 | 0.728 | 0.329 | **-0.215** |
| D17Mit13 | 68 | 212 - 242 | 10 | 0.912 | 0.847 | 0.764 | -0.069 |
| D17Nds3 | 73 | 206 - 246 | 10 | 0.918 | 0.788 | 0.298 | -0.165 |
| D17Mit20 | 71 | 155 - 189 | 7 | 0.549 | 0.5620 | 0.100 | 0.023 |

The majority of the $F_{IS}$ values for the genome-wide markers are positive (Table 4.1 and Figure 4.3), with two markers showing significant deviations from zero (D11Mit300, $F_{IS} = 0.183$, $Z = 2.16$, $p = 0.031$; D19Mit70, $F_{IS} = 0.195$, $Z = 2.30$, $p = 0.021$). Positive $F_{IS}$ values indicate a heterozygote deficiency from that expected under random mating, which can be a result of a level of inbreeding in a population. The two MHC-linked markers either side of the MHC region (D17Mit19 and D17Mit2) also have positive $F_{IS}$ values but the markers located within the MHC coding region all have negative $F_{IS}$ values, with one marker significantly

lower than zero (D17Mit231, $F_{IS}$ = -0.215, Z = -2.20, p = 0.028). The markers in the middle of the MHC region are more negative than those at the edges. The mean $F_{IS}$ is significantly lower for MHC-linked markers ($F_{IS}$ = -0.069) than genome-wide markers ($F_{IS}$ = 0.071) ($\chi^2$ = 7.66, p = 0.006). The difference in $F_{IS}$ values between MHC-linked markers and the control genome-wide markers is an indication of contemporary selection acting at the MHC. However, care must be taken when interpreting results from the MHC markers in this study, as linked markers are not independent. Negative $F_{IS}$ values indicate an excess of heterozygotes from expected under neutral evolution, which can be an indication of balancing selection. I used Watterson's homozygosity test of neutrality to test for evidence of balancing selection acting on the MHC-linked markers.

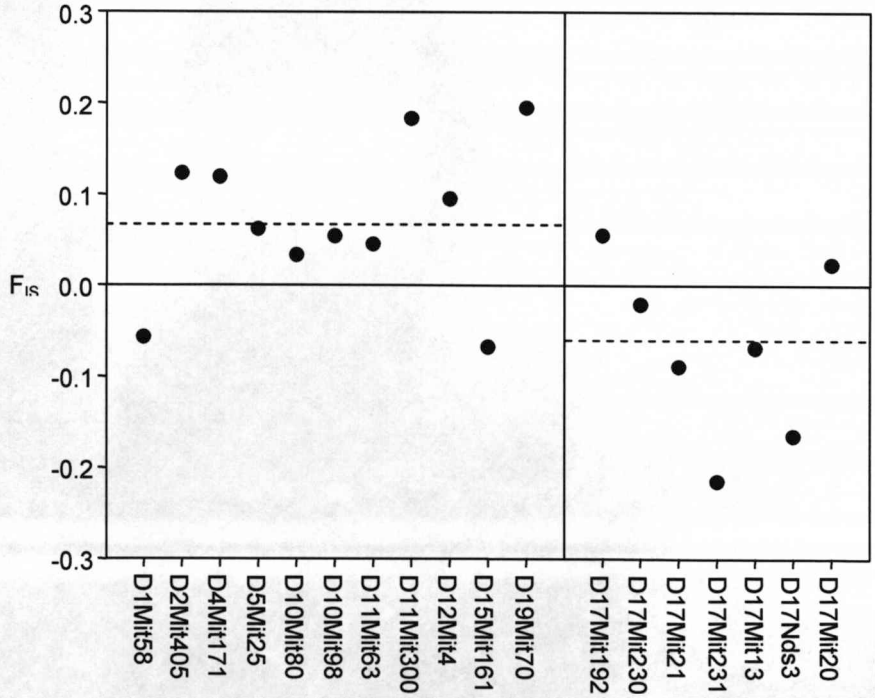



Figure 4.3 $F_{IS}$ values plotted by marker. The mean for genome-wide markers is represented by the dashed line in the left panel; mean values for MHC-linked markers by the dashed line in the right panel. The connecting line shows the trend across the MHC region.

Table 4.4. Summary of the observed and expected values of Watterson's test statistic F. The deviate of F $(F_{(obs)}-F_{(exp)}/(var(F_{(exp)})^{1/2}$ and significance are shown. Significant values ($p < 0.05$) are shown in bold.

| Locus | K | 2N | $F_{(Obs)}$ | $F_{(Exp)}$ | Deviate | Probability |
|---|---|---|---|---|---|---|
| | | | Genome-wide markers | | | |
| D1Mit58 | 5 | 144 | 0.547 | 0.608 | -0.349 | 0.364 |
| D2Mit405 | 3 | 150 | 0.329 | 0.602 | -1.564 | 0.059 |
| D4Mit171 | 4 | 148 | 0.533 | 0.542 | -0.054 | 0.478 |
| D5Mit25 | 6 | 140 | 0.265 | 0.456 | -1.273 | 0.102 |
| D10Mit80 | 6 | 148 | 0.770 | 0.453 | **2.059** | **0.020** |
| D10Mit98 | 3 | 146 | 0.529 | 0.450 | 0.515 | 0.303 |
| D11Mit63 | 4 | 146 | 0.521 | 0.688 | -0.915 | 0.180 |
| D11Mit300 | 6 | 148 | 0.514 | 0.513 | 0.005 | 0.498 |
| D12Mit4 | 6 | 144 | 0.459 | 0.602 | -0.816 | 0.207 |
| D15Mit161 | 4 | 68 | 0.527 | 0.396 | 0.928 | 0.177 |
| D19Mit70 | 4 | 60 | 0.306 | 0.395 | -0.637 | 0.262 |
| | | | MHC markers | | | |
| D17Mit192 | 5 | 140 | 0.329 | 0.450 | -0.797 | 0.213 |
| D17Mit230 | 10 | 140 | 0.100 | 0.268 | **-1.732** | **0.042** |
| D17Mit21 | 6 | 86 | 0.163 | 0.412 | **-1.795** | **0.036** |
| D17Mit231 | 7 | 136 | 0.118 | 0.392 | **-1.973** | **0.024** |
| D17Mit13 | 10 | 136 | 0.088 | 0.228 | **-1.685** | **0.046** |
| D17Nds3 | 10 | 146 | 0.082 | 0.271 | **-1.947** | **0.026** |
| D17Mit20 | 7 | 142 | 0.592 | 0.394 | 1.413 | 0.079 |

The F deviate values for the genome-wide markers are spread around zero (mean = -0.037), but take both negative and positive values (Table 4.4 and Figure 4.4). One marker (D10Mit80) is significantly greater than zero. All the MHC-linked markers have negative values of F deviate, apart from D17Mit20 which is outside the MHC coding region. The five markers in the coding region have values that are significantly lower than zero, indicating a deficit in homozygous genotypes from the expected. This increased heterozygosity is a result of more equal allele frequency distributions than expected under a model of neutrality, which can be a consequence of balancing selection.

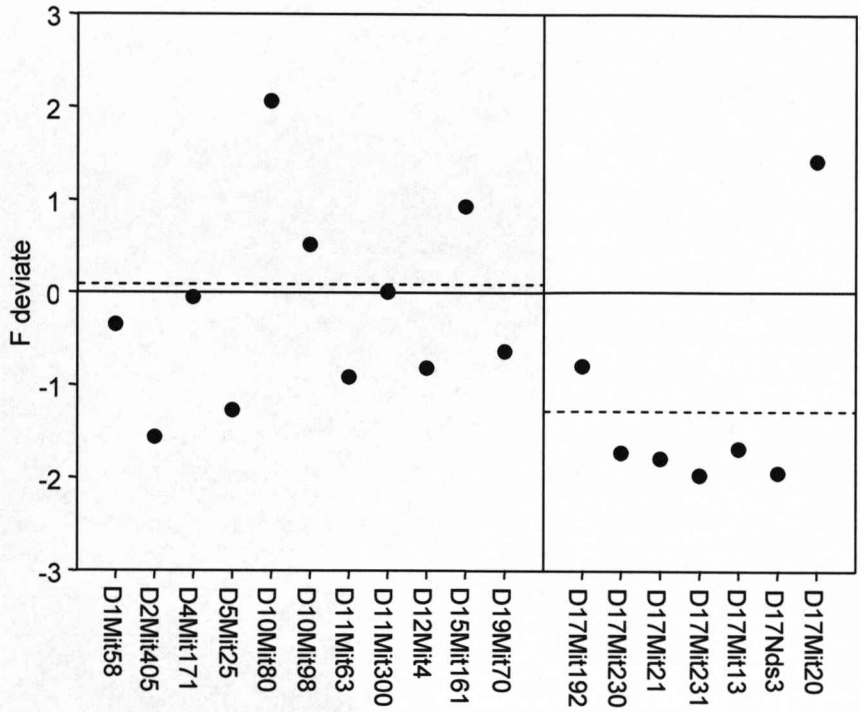



Figure 4.4 F deviate values plotted by marker. The mean for genome-wide markers is represented by the dashed line in the left panel; mean values for MHC-linked markers by the dashed line in the right panel. The connecting line shows the trend across the MHC region.

### 4.4.3 Sequencing

I obtained 10 unique sequences for H2-Ab exon two and 16 for H2-K1 exons two and three. All sequences were obtained from more than one PCR reaction and from more than one individual to confirm polymorphisms were not due to PCR errors. Six of the sequences obtained for H2-Ab have a nine base-pair deletion at positions 182 to 191 of the 273 base pairs coding for exon two. The nine base pairs CCG GAG ATC are replaced by the three base pairs TAC resulting in genes that are six base pairs shorter. The three amino acids proline (hydrophobic), glutamic acid (polar acidic) and isoleucine (hydrophobic) are replaced by tyrosine (polar non-charged).

### 4.4.4 Synonymous and Nonsynonymous substitutions

I used MEGA 4.0 to calculate the rates of substitution per potential synonymous site ($\bar{d}_S$) and potential nonsynonymous site ($\bar{d}_N$) between the unique sequences obtained from the Winsford population (Table 4.5).

Table 4.5 The number of synonymous substitutions per synonymous site from analysis between sequences is shown. Standard error estimates were obtained by a bootstrap procedure (500 replicates). Analyses were conducted using the modified Nei-Gojobori (assumed transition/transversion bias = 2) method in MEGA4 with Jukes-Canter correction for multiple hits. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option).

| | **Non-antigen binding sites** | | | **Antigen binding sites** | | |
|---|---|---|---|---|---|---|
| | $\bar{d}_S$ | $\bar{d}_N$ | $\bar{d}_N/\bar{d}_S$ | $\bar{d}_S$ | $\bar{d}_N$ | $\bar{d}_N/\bar{d}_S$ |
| **All** | **0.0776** ± 0.0172 | **0.0741** ± 0.0206 | **0.954** ± 0.0269 | **0.0545** ± 0.0364 | **0.2185** ± 0.0840 | **4.001** ± 0.0915 |
| **H2-Ab** | 0.0210 ± 0.0154 | 0.0417 ± 0.0202 | 1.986 ± 0.0254 | 0.0306 ± 0.0326 | 0.1512 ± 0.0754 | 4.941 ± 0.0821 |
| **H2-K1** | 0.1343 ± 0.0306 | 0.1064 ± 0.0202 | 0.792 ± 0.0367 | 0.0785 ± 0.0657 | 0.2857 ± 0.1492 | 3.639 ± 0.1630 |

Both $\bar{d}_S$ and $\bar{d}_N$ are significantly higher in the class I gene H2-Ab than the class II gene H2-K1. Analysing both genes together, the rates of synonymous substitution per potential synonymous site is constant in both the non-antigen binding and antigen binding regions and the rate of non synonymous substitution per potential nonsynonymous site in the non-antigen binding region is not significantly different from this value ($F_{2,494}$ = 0.227, p = 0.797). $\bar{d}_N$ does not significantly differ from $\bar{d}_S$ in the non-antigen binding regions, but is significantly greater in the antigen binding regions (Table .4.6).

Table 4.6 Codon-based test of positive selection for analysis between sequences. The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) in favour of the alternative hypothesis (dN > dS) is shown. The variance of the difference was computed using the bootstrap method (500 replicates). Analyses were conducted using the modified Nei-Gojobori (assumed transition/transversion bias = 2) method in MEGA4 with Jukes-Canter correction for multiple hits. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option).

| | **Non-antigen binding sites** | | **Antigen binding sites** | |
|---|---|---|---|---|
| | **Z** | **p** | **Z** | **p** |
| **H2-Ab** | 1.237 | 0.109 | **2.163** | **0.016** |
| **H2-K1** | -1.272 | 1.000 | **2.743** | **0.004** |

Tajima's neutrality test compares the number of segregating sites and the nucleotide diversity in a population. Both genes tested show significantly positive values of Tajima's *D*, with H2-Ab (significant at 99% confidence) having a greater value than H2-K1 (significant at 95% confidence). A positive Tajima's *D* signifies greater numbers of pairwise nucleotide differences than expected from the number of polymorphic sites between sequences. This indicates possible balancing selection or population subdivision. As I have shown that the individuals sampled are likely to be from the same population, Tajima's test suggests that balancing selection is acting on the MHC region.

Table 4.7 Tajima's neutrality test as estimated in MEGA 4.0. The abbreviations used are as follows: m = number of sequences, $\theta$ = number of segregating sites per sequence, $\Theta = \theta/a_1$, and $\pi$ = nucleotide diversity. D is the Tajima test statistic. Significant ($p < 0.05$) values of D are shown in bold.

| | **m** | **$\theta$** | **$\Theta$** | **$\pi$** | **D** |
|---|---|---|---|---|---|
| **H2-Ab** | 75 | 46 | 0.035 | 0.065 | **2.731** |
| **H2-K1** | 75 | 191 | 0.072 | 0.158 | **2.142** |

## 4.5 DISCUSSION

There is no evidence to suggest that the mice sampled from the Winsford population do not belong to a single, stable, panmictic population. I used allele frequencies and levels of heterozygosity to test for evidence of selection acting on the MHC. Population growth and decline affect levels of heterozygosity; for example a recent bottleneck would have been characterised by high levels of heterozygosity due to rapid population expansion (Kimmel et al. 1998). I found no evidence of such an event, thus any variation in allele frequencies are likely to be due to the effects of selection.

Potentially, several generations and family groups could have been sampled in this study, and without information about age this makes parentage reconstructions very difficult. Parentage reconstruction is confounded when the candidate parents are related to one another and the most problematic situation arises when either half- or full-siblings of the progeny are included in the pool of candidate parents (Jones and Ardren 2003). Both could be the case in this study, as house mice live in family demes where a single dominant male defends a territory and females. Individuals trapped in the same area are likely to belong to a single extended family group making parentage assignment difficult and unreliable. Instead I assessed average relatedness within the sampled mice. I found a significant deviation from the null hypothesis of no relatedness. The mean relatedness was below that expected between half-siblings ($r_{xy}$ = 0.25), although the highest pairwise relatedness seen was actually above that expected between full-siblings or parent-offspring pairs ($r_{xy}$ = 0.5).

This suggests that there has been some level of inbreeding within the Winsford population, which agrees with what is already known about breeding patterns of *Mus musculus domesticus* in the wild (Bronson 1979). Blouin et al. (1996) trapped mice in a barn on a dairy farm in California, USA. They typed them at 20 microsatellite loci and investigated relatedness. They found a much lower value of relatedness than in this study (mean $r_{xy} = -0.02$), suggesting that few of the mice captured were relatives. However, they also had some individual relatedness estimates above 0.5. Blouin et al. give few details about the duration and location of trapping, thus it may be that they sampled a larger area than in the current study and thus caught animals from a meta-population or several sub-populations. Another alternative is that the relatedness in the current study is high due to the effects of inbreeding. If relatedness in the Winsford population is such that mice are regularly faced with quite close relatives, there are likely to be inbreeding avoidance mechanisms at work within the population. Thus any evidence of selective pressures acting on the MHC may be due to the use of this multi-gene locus as a representation of genome-wide heterozygosity and/or a marker of close kin.

There was evidence of significant disequilibrium between almost all pairs of markers within and surrounding the MHC. Linkage disequilibrium within the MHC region justifies the use of microsatellite markers to test for balancing selection acting on MHC genes. Sites within the MHC that are under the influence of balancing selection will be linked to the microsatellite markers, thus any evidence of selection could be detected by the markers alone. Watterson's test of neutrality for the MHC-linked markers indicates balancing selection acting at the MHC as found in other studies (Paterson 1998; Huang and Yu 2003). Balancing selection works to maintain multiple alleles within a population, leading to large numbers of alleles with equal frequencies, which results in higher numbers of

homozygotes than expected under neutrality. This can be contrasted to the F deviates for the genome-wide markers, which are spread around zero (with the exception of one marker which is significantly positive, suggesting purifying selection is acting on a linked gene). The genome-wide neutral markers prove a useful tool in investigating selection pressures acting on genes. Any influence of population growth or decline would also be evident in these markers, so it is the comparison between neutral markers and MHC-linked markers that allow deviations from neutral evolution to be observed in the context of the recent demography of the population.

In keeping with previous studies (Hughes and Nei 1988; Hughes and Nei 1989; Hughes and Hughes 1995; Paterson 1998), I found evidence of balancing selection in the pattern of substitution in the antigen binding site. The rate of nonsynonymous evolution is significantly higher than the rate of synonymous substitutions, indicating selection for diversity at residues that are involved in presentation of antigen to cells of the immune system. This pattern does not hold for the rest of the molecule, where the rates of nonsynonymous and synonymous substitutions are approximately equal. The apparent focus of the balancing selection on the antigen binding region of the MHC molecule suggests that it is the ability to bind a diverse array of peptides that is being favoured by natural selection. Tajima's *D* is significantly positive for both MHC genes sequenced, showing the maintenance of alleles at high frequency. High numbers of alleles at equal frequencies is a consequence of balancing selection, confirming that MHC loci in this population are subject to this type of selection pressure.

I have observed evidence of balancing selection in a wild population of mice both in the frequencies of alleles at the MHC, and in the increased rate of nonsynonymous substitution in the antigen binding sites. Elucidating the mechanisms maintaining this selection pressure was not

within the scope of this study. Both heterozygote advantage and frequency dependent selection have been proposed to maintain genetic diversity at the MHC (Doherty and Zinkernagel 1975; Hughes and Nei 1988; Potts and Wakeland 1990; Takahata and Nei 1990; Slade and Mccallum 1992; Potts and Wakeland 1993), and pathogen-host co-evolution and mate choice have been proposed as drivers of these processes. Further work needs to be carried out to disentangle the different pressures maintaining balancing selection at the MHC.

# Chapter 5

# COMPARISON OF THE GENETIC DISTANCE BETWEEN MHC ALLELES IN A NATURAL POPULATION AND IN INBRED STRAINS

## 5.1 INTRODUCTION

A large proportion of studies investigating the functional and genetic properties of the MHC have used inbred laboratory strains of mice as their model species. These mice have been selectively bred for over a hundred years and each genome-wide homozygous strain can be traced back to a small number of founding animals (Atchley and Fitch 1991; Beck et al. 2000). Homogeneity within a strain and heterogeneity between strains means inbred mice have been invaluable in studies investigating inheritance and immunology. However, how useful are these homozygous animals when studying a naturally polymorphic group of loci such as the MHC?

The MHC is a gene cluster characterised by high levels of polymorphism, and rates of mutation and signals of selection pressures at the MHC have been well characterised (see 4.1). There are a greater number of alleles at MHC genes than at most other loci and levels of heterozygosity are high (Hedrick 1994), as is sequence divergence between alleles (Hughes and Yeager 1998). It is believed that there could be between 50 and 100 alleles in the more polymorphic loci such as H2-K and H2-D (Pullen et al. 1992) and several groups have characterised some of the more common sequences (Gotze et al. 1980b; Arnold et al. 1984; Steinmetz 1986; Soper

et al. 1988). The majority of substitutions that make MHC genes so diverse are located within the region that codes for the antigen binding site. The antigen binding site is a specialised groove encoded by approximately 50 amino acids located in the extracellular portion of the molecule, and is the region that interacts with self and foreign antigens.

Inbred strains of mice have been used in studies of pathogenic infections to separate the contributions of MHC and non-MHC genes on resistance (Dick et al. 1988; Else and Wakelin 1988; Else et al. 1990a; Else et al. 1990b), and for that they are a useful tool. However, the outcome of infection is dependent on both the host and pathogen's genotype (Casadevall and Pirofski 1999) and the limited range of MHC and other immune genes found in inbred strains of mice may not be an appropriate model to study the outcome of infection in an out bred population. In a natural population, allelic diversity at the MHC is driven by parasite-mediated balancing selection and disassortative mating at the MHC (Apanius et al. 1997a; Edwards and Hedrick 1998; Hughes and Yeager 1998; Meyer and Thomson 2001; Penn et al. 2002). The guidelines for creating an inbred strain of mice state that the strain must originate from a single mating pair and have been created by brother-sister matings for 20 generations or more (Davisson 1996). The result is that all individuals in a strain are essentially genetically identical. Forcing brother-sister matings removes any mate choice driven selection on the MHC and laboratory bred mice are kept infection free meaning there is no infection driven selection. Thus inbred mice are lacking the diversity within strains that is characteristic of the MHC in the wild.

The MHC has been suggested as a signal of relatedness that is used in the avoidance of mating with close kin (see Brown and Eklund 1994 for a review). Parental MHC type may be learnt in the nest and avoided when mating (Beauchamp et al. 1988; Yamazaki et al. 1988), or a self-referent

system of recognising closely related MHC types could be used. In either case, there may be an ideal genetic distance between an individual and their choice of mate; if two MHC haplotypes are closely related the potential mate is likely to be close kin, if they are distantly related the potential mate is likely to be from a different population and breeding success may be affected by outbreeding depression (Marr et al. 2002). If such a discriminatory system has evolved in the mouse, it is likely to function at the level of genetic distances that are found in natural populations. It has been shown that mice can be trained to distinguish between odours from MHC congenic mice (Yamazaki et al. 1979) and even between MHC congenic strains that differ at a single nucleotide (Yamazaki et al. 1983). Although these studies do not directly test whether mice are using such differences when choosing a mate, they do show that they are able to determine minute changes at the MHC in urine alone. This study investigates whether genetic distances between laboratory and MHC congenic strains of mice are representative of genetic distances between MHC sequences in a natural population, and thus whether they would be encountered and informative in the wild.

We sampled a natural population of *Mus musculus domesticus* from a commercial pig farm in the north-west of England. We collected genetic samples from 75 individuals, including several family groups, spanning several generations. Amino acid analysis of these mice at two MHC genes gives an estimate of the genetic distances between MHC alleles within a natural population. This is then compared to the distances between MHC alleles used in studies of MHC biased mating choices, including our own (chapter 3).

## 5.2 AIMS

To estimate the genetic distances between MHC alleles in several different groups of *Mus musculus domesticus* I sequenced a class I gene (H2-K1) and a class II gene (H2-Ab) for 75 wild caught individuals from a natural population, the 16 MHC haplotypes present in our enclosure study, and I retrieved published sequences for a selection of inbred strains. A comparison was made between the mean genetic distances for each group to investigate whether the genetic distances used in studies of MHC biased mate choice and parasite studies represent the distances that would be encountered by an individual in the wild.

## 5.3 MATERIALS AND METHODS

### 5.3.1 Animals

To investigate MHC sequence diversity in a natural population I sequenced 75 wild house mice (*Mus musculus domesticus*) that were trapped in and around the buildings of Winsford Pig Farm (Cheshire, UK) in seven trapping sessions between September 2003 and January 2005 (see 4.3.1). The animals used to breed founding individuals for the enclosure study were the F1 - F3 offspring of wild house mice caught in five different locations throughout the northwest of England, including the Winsford site (see 3.3.1.1).

### 5.3.2 Sequencing

DNA was extracted from tail snips from mice from the Winsford population and PCR reactions carried out as described in chapter 4 (see 4.3.4). MHC haplotypes of the mice in the enclosure study had previously been determined (see 3.3.1.5), so I was able to identify individuals that

were homozygous at the MHC for 8 of the 16 haplotypes included in the study. Genetic samples were kept for all founders and offspring captured, meaning I could easily sequence directly from PCR for these individuals (see 4.3.4 for details). The MHC gene sequences of the remaining haplotypes were determined using a combination of cloning and direct PCR sequencing.

MHC sequences for inbred strains were found by searching the MGI (Eppig 2007) and NCBI (Wheeler et al. 2008) databases. The IMGT Repertoire (MHC) (Lefranc et al. 2005) provides MHC haplotypes and polymorphism information for each inbred strain of mouse.

### 5.3.3 Genetic distance estimations

The mean number of amino acid changes between pairwise sequence comparisons was calculated in MEGA 4.0 (Tamura et al. 2007) and the standard error of the mean was calculated by the analytical method. The nucleotide sites coding for the amino acids involved in antigen binding were chosen based on the three-dimensional crystal structure of the MHC molecule (Fremont et al. 1992 for H2-K1; Zhu et al. 2003 for H2-Ab1).

I further investigated the types of substitutions in each data set by looking at the rate of synonymous substitutions per potential synonymous site ($\bar{d}_S$) and the rate of nonsynonymous substitutions per potential nonsynonymous site ($\bar{d}_N$) in MEGA. The modified Nei-Gojobori method was used (Nei and Gojobori 1986; Zhang et al. 1998), which weights the transition/transversion ratio in the same way as the Kimura two-parameter method. A Jukes-Cantor correction was used (Jukes 1969) to correct for multiple hits.

### 5.3.4 Phylogenetic tree of inbred sequences

Maximum likelihood (ML) criteria were performed using a heuristic search in PAUP* (Swofford 1993). ML analyses were run with parameter settings identified as optimal for the data set by Modeltest (Posada and Crandall 1998), which compares the model goodness-of-fit by using hierarchical likelihood ratio tests (hLRTs) and the Akaike Information Criterion (AIC) (Posada and Crandall 2001). The substitution model chosen was the TrNef model (Tamura and Nei 1993). Rate variation across sites was estimated to have a gamma (G) distribution with a shape parameter of 0.3224. The proportion of invariable sites was estimated to be 0.3351. The substitution rate matrix was estimated at A–C = 1.0, A–G = 1.6, A–T = 1.0, C–G = 1.0, C–T = 2.3 and G–T = 1.0. Base frequencies were set as equal. Tree searches were carried out using a heuristic search strategy. A total of ten random addition sequence replicates were carried out, with the tree-bisection-reconnection (TBR) branch swapping option in effect. Bootstrap resampling was employed using 100 replicates. The tree was constructed using the 13 inbred strains with published H2-Ab sequences (H2-Ab alone was used as only 8 sequences could be obtained for H2-K1). A H2-Eb sequence as an outgroup to root the tree.

## 5.4 RESULTS

### 5.4.1 Sequencing results

I obtained 10 unique sequences for H2-Ab exon two and 16 for H2-K1 exons two and three (see 4.4.3). The same nine base-pair deletion in six of the H2-Ab sequences is also seen in 10 of the 16 haplotypes used in the enclosure experiment, probably because these animals are likely to have Winsford caught individuals in their ancestry. In fact, three of the

sequences obtained for the enclosure group are identical to sequences in the Winsford population, presumably due to common ancestry. Seven of the inbred sequences retrieved also show the same deletion. None of the Winsford or enclosure sequences match the inbred sequences used.

Mice used in enclosures and their parents had previously been typed using microsatellite markers located in and around the MHC region (Chapter 3). In enclosure B, this typing revealed a recombination of paternal MHC haplotypes within the MHC region (Table 5.1). We sequenced individuals known to carry this recombined haplotype and obtained a H2-K1 sequence corresponding to the k haplotype and a H2-Ab sequence corresponding to the o haplotype confirming that a cross-over event had occurred. Our sequencing results give additional information as to the location of the crossover. The cross-over occurs between marker D17MIT230 and marker D17MIT22. D17MIT230 is located in the H2-K1 coding region and D17MIT22 is in the H2-Eb1 coding region. H2-Ab lies between the two markers (Appendix Figure 5.1). As the H2-K1 sequence corresponds to haplotype k and the H2-Ab sequence corresponds to haplotype o, the cross-over event must be located between the H2-K1 and H2-Ab coding regions.

Table 5.1. Microsatellite sizes across the MHC region for a founding animal from enclosure B and her parents, coloured according to haplotype. The cross-over can be clearly seen between marker D17MIT230 and marker D17MIT22. The box surrounds the markers within the MHC, the other three markers are located outside the region.

| | haplotype | D17MIT23 | D17MIT192 | D17MIT230 | D17MIT22 | D17 MIT13 | D17NDS3 | D17MIT20 |
|---|---|---|---|---|---|---|---|---|
| Sire | k | 184 | 302 | 209 | 181 | 231 | 249 | 172 |
| | o | 187 | 288 | 222 | 187 | 229 | 233 | 176 |
| Dam | p | 187 | 288 | 215 | 181 | 235 | 243 | 180 |
| | l | 187 | 288 | 211 | 191 | 220 | 237 | 180 |
| ID1697 | k/o | 184 | 302 | 209 | 187 | 229 | 233 | 176 |
| | p | 187 | 288 | 215 | 181 | 235 | 243 | 180 |

### 5.4.2 Genetic distance estimates

I calculated the pairwise number of amino acid substitutions per site between unique sequences within each of the three groups: the natural Winsford population, the enclosure experiment and inbred strains. I then took the mean of these distance estimates to represent the average distance between two individuals in a particular group (Figure 5.1 and Table 5.2).

Table 5.2. The mean number of amino acid substitutions (aa subs) between all pairwise comparisons within each test group. H2-Ab (10 sequences) and H2-K1 (16 sequences) are shown separately, as well as all sequence data combined. The standard errors of the mean are included.

| | **Winsford** | **Enclosure** | **Inbred** | **Average** |
|---|---|---|---|---|
| **All** | | | | |
| aa subs | **41.19** | **49.02** | **40.14** | **43.45** |
| std. err. | ± 4.36 | ± 4.24 | ± 3.61 | ± 4.07 |
| **H2-Ab** | | | | |
| aa subs | 8.31 | 11.55 | 12.56 | 10.81 |
| std. err. | ± 2.54 | ± 3.08 | ± 3.21 | ± 2.94 |
| **H2-K1** | | | | |
| aa subs | 32.86 | 37.47 | 27.57 | 32.64 |
| std. err. | ± 5.04 | ± 5.39 | ± 4.75 | ± 5.06 |

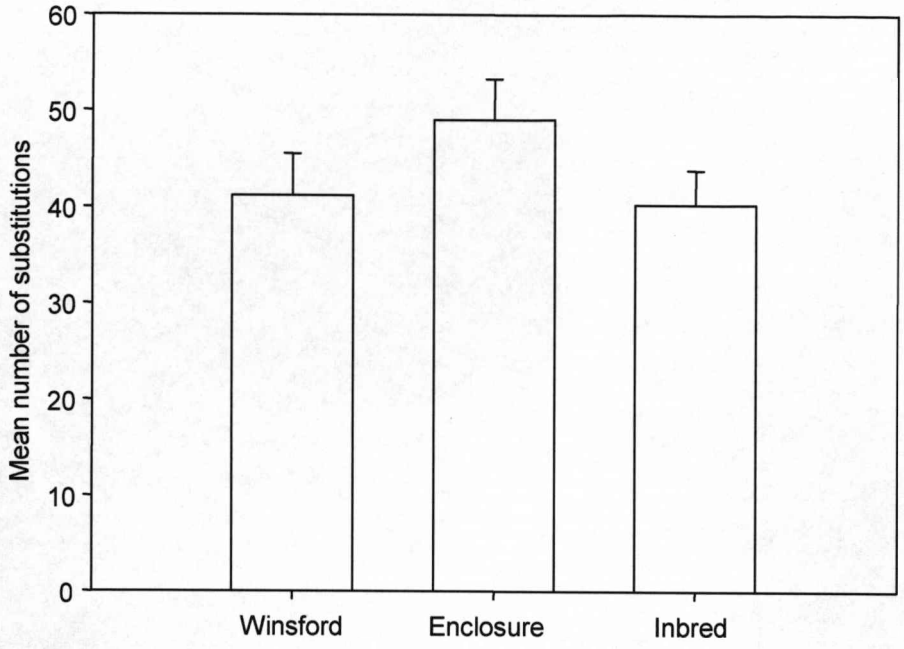



Figure 5.1. Mean of the amino acid substitutions between all pairwise comparisons within each test group. Error bars represent the standard error of the mean.

The enclosure study has the highest mean genetic divergence, and the inbred strains have the lowest, with the natural population between the two. There is a significant difference in the mean number of amino acid substitutions between the natural population and the enclosure group ($Z = 1.80$, $p = 0.036$), but not between the natural population and the inbred strains ($Z = -0.24$, $p = 0.405$). The highest single pair-wise comparison (57 amino acids) is between two H2-K1 sequences in the natural population.

I also looked at the genetic distances between sequences when considering amino acids involved in the antigen binding region of the molecule alone (Table 5.3).

Table 5.3. The mean number of amino acid substitutions (aa subs) between all pairwise comparisons within each test group using antigen binding sites only. H2-Ab (10 sequences) and H2-K1 (16 sequences) are shown separately, as well as all sequence data combined. The standard errors of the mean are included.

| | Winsford | Enclosure | Inbred | Average |
|---|---|---|---|---|
| **All** | | | | |
| aa subs | **9.48** | **11.93** | **13.30** | **11.57** |
| std. err. | ± 1.72 | ± 4.24 | ± 1.95 | ± 2.63 |
| **H2-Ab** | | | | |
| aa subs | 3.96 | 5.62 | 7.41 | 5.66 |
| std. err. | ± 2.11 | ± 3.99 | ± 1.98 | ± 2.18 |
| **H2-K1** | | | | |
| aa subs | 5.53 | 6.32 | 5.89 | 5.91 |
| std. err. | ± 1.80 | ± 5.39 | ± 1.86 | ± 3.02 |

When comparing the number of amino acid substitutions in the antigen binding sites only the inbred strains have the highest genetic diversity, and the natural population the lowest. There is a significant difference between the natural population and the inbred strains ($Z = -2.22$, $p = 0.013$), but not between the natural population and the enclosure study ($Z = -1.42$, $p = 0.077$).

As there are more structural amino acids than there are amino acids involved in the antigen binding site, to compare the rates of substitution, I calculated the number of substitutions per site for non-antigen binding and antigen binding sites (Figure 5.2).

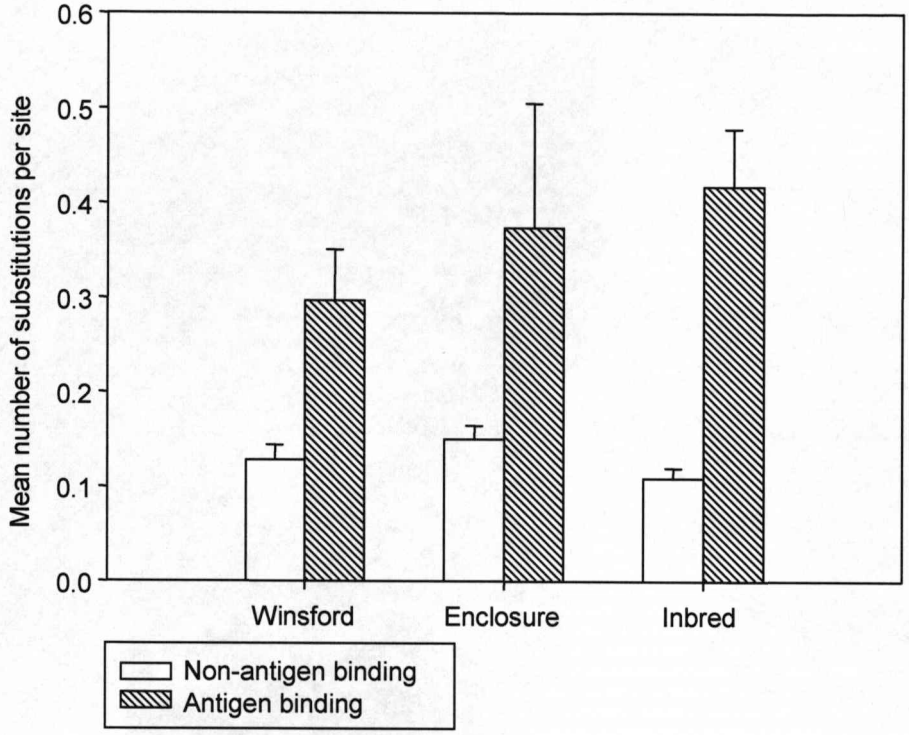



Figure 5.2 Mean of the number of amino acid substitutions per site between all pairwise comparisons within each test group. Comparisons between non-antigen binding sites and antigen binding sites are shown separately. Error bars represent the standard error of the mean.

The genetic distances between binding sites alone is significantly higher than when considering the non-binding sites ($Z = 7.71$, $p < 0.001$). These sites are under higher selection pressure than the rest of the gene, and thus diverge more quickly, reflected in the greater genetic distances seen.

I used the rates of nucleotide substitution per potential synonymous site ($\bar{d}_S$) and potential nonsynonymous site ($\bar{d}_N$) to further investigate the causes of the variance in genetic distances between groups (Figure 5.3 and Table 5.4).

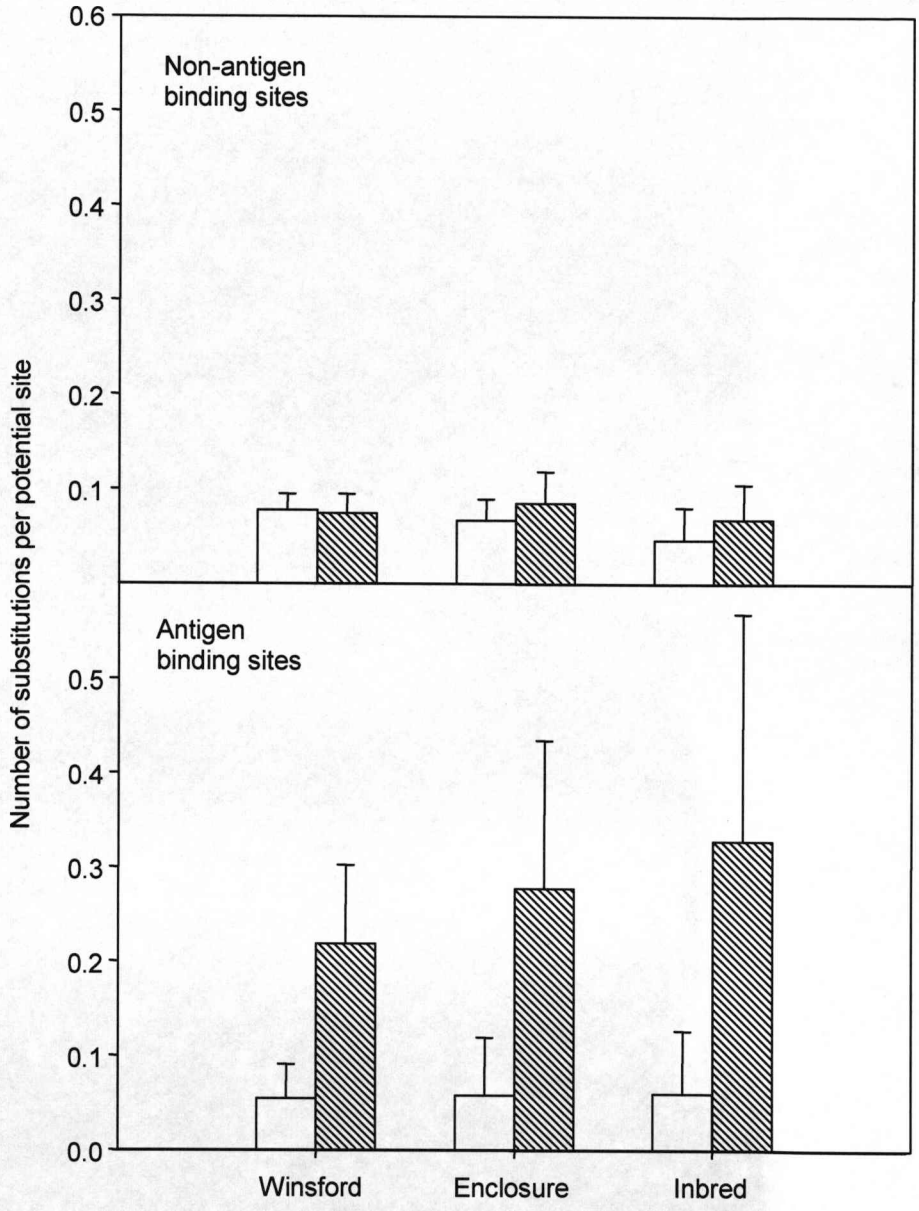



Figure 5.3. The numbers of synonymous substitutions per potential synonymous substitution site, dS (white bars), and the numbers of nonsynonymous substitutions per potential nonsynonymous substitution site, dN (hatched bars), for all three populations. Shown for non-antigen binding regions (upper panel) and antigen binding regions (lower panel) separately.

Table 5.4. The number of synonymous substitutions per synonymous site from analysis between sequences is shown. Standard error estimates were obtained by a bootstrap procedure (500 replicates). Analyses were conducted using the modified Nei-Gojobori (assumed transition/transversion bias = 2) method in MEGA4 with Jukes-Canter correction for multiple hits. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option).

| | Non-antigen binding sites | | | Antigen binding sites | | |
|---|---|---|---|---|---|---|
| | $\bar{d}_S$ | $\bar{d}_N$ | $\bar{d}_N/\bar{d}_S$ | $\bar{d}_S$ | $\bar{d}_N$ | $\bar{d}_N/\bar{d}_S$ |
| **Winsford** | | | | | | |
| **All** | **0.0776** ± 0.0172 | **0.0741** ± 0.0206 | **0.954** ± 0.0269 | **0.0545** ± 0.0364 | **0.2185** ± 0.0840 | **4.001** ± 0.0915 |
| H2-Ab n = 10 | 0.0210 ± 0.0154 | 0.0417 ± 0.0202 | 0.0207 ± 0.0254 | 0.0306 ± 0.0326 | 0.1512 ± 0.0754 | 0.1207 ± 0.0821 |
| H2-K1 n = 16 | 0.1343 ± 0.0306 | 0.1064 ± 0.0202 | -0.0279 ± 0.0367 | 0.0785 ± 0.0657 | 0.2857 ± 0.1492 | 0.2072 ± 0.1630 |
| **Enclosure** | | | | | | |
| **All** | **0.0671** ± 0.0222 | **0.0848** ± 0.0332 | **1.264** ± 0.0399 | **0.0579** ± 0.0613 | **0.2774** ± 0.1569 | **4.791** ± 0.1685 |
| H2-Ab n = 16 | 0.0254 ± 0.0178 | 0.0518 ± 0.0233 | 0.0264 ± 0.0293 | 0.0443 ± 0.0408 | 0.3609 ± 0.1540 | 0.3166 ± 0.1593 |
| H2-K1 n = 16 | 0.0671 ± 0.0748 | 0.0837 ± 0.0166 | 0.0166 ± 0.0766 | 0.0749 ± 0.0748 | 0.2973 ± 0.0166 | 0.2224 ± 0.0766 |
| **Inbred** | | | | | | |
| **All** | **0.0462** ± 0.0344 | **0.0677** ± 0.0366 | **1.465** ± 0.0502 | **0.0596** ± 0.0666 | **0.3291** ± 0.2397 | **5.522** ± 0.2488 |
| H2-Ab n = 13 | 0.0202 ± 0.0154 | 0.0557 ± 0.0239 | 0.0355 ± 0.0284 | 0.0399 ± 0.0435 | 0.2488 ± 0.1117 | 0.2089 ± 0.1199 |
| H2-K1 n = 8 | 0.1139 ± 0.0284 | 0.1139 ± 0.0207 | 0.0000 ± 0.0351 | 0.0758 ± 0.0592 | 0.3059 ± 0.1402 | 0.2301 ± 0.1522 |

The rate of synonymous substitutions per potential synonymous site remains fairly constant when comparing antigen binding and non-antigen binding residues for all three populations and the rate of nonsynonymous substitutions per nonsynonymous site is equivalent when looking at non-antigen binding sites ($F_{8,1510} = 0.063$, $p = 0.999$). $\bar{d}_N$ was higher than $\bar{d}_S$ at all but one comparison (the non-antigen binding residues in the Winsford population $\bar{d}_S = 0.0776 \pm 0.0172$, $\bar{d}_N = 0.0741 \pm 0.0206$). $\bar{d}_N$ and $\bar{d}_S$ values are comparable in the non-antigen binding regions, but are significantly higher when looking at the antigen binding residues in all

three groups (Table 5.5). The inbred strains show the highest $\bar{d}_N$ and the highest $\bar{d}_N/\bar{d}_S$ at the antigen binding sites.

Table 5.5. Codon-based test of positive selection for analysis between sequences. The probability of rejecting the null hypothesis of strict-neutrality ($\bar{d}N = \bar{d}S$) in favour of the alternative hypothesis ($\bar{d}N > \bar{d}S$) is shown with significant differences shown in bold. The variance of the difference was computed using the bootstrap method (500 replicates). Analyses were conducted using the modified Nei-Gojobori (assumed transition/transversion bias = 2) method in MEGA4 with Jukes-Canter correction for multiple hits. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option).

| | **Non-antigen binding sites** | | **Antigen binding sites** | |
|---|---|---|---|---|
| | **Z** | **p** | **Z** | **p** |
| **Winsford** | | | | |
| H2-Ab | 1.237 | 0.109 | **2.163** | **0.016** |
| H2-K1 | -1.272 | 1.000 | **2.743** | **0.004** |
| **Enclosure** | | | | |
| H2-Ab | **1.944** | **0.027** | **3.193** | **0.001** |
| H2-K1 | -0.002 | 1.000 | **2.696** | **0.004** |
| **Inbred** | | | | |
| H2-Ab | 1.324 | 0.094 | **3.758** | **<0.001** |
| H2-K1 | 0.948 | 0.173 | **2.572** | **0.006** |

### 5.4.3 Genetic distance between strains used in previous studies

The functional region of the MHC molecule, the antigen binding site, determine the immunological binding properties of the molecule and is hypothesised to be the region that determines MHC specific odour changes in urine. Thus, the amino acids coding for the antigen binding regions are of particular interest. I looked at the genetic distances between the antigen binding sites of inbred haplotypes used in previous mate choice and familial recognition studies (Table 5.6) and parasite infection studies (Table 5.7). I compared these distances to the mean distance found in the wild population to see whether they were a fair representation of the distances between the MHC alleles mice would encounter in the wild. Many studies use the same set of inbred strains so a limited number of MHC haplotypes have been used. Although when taking all inbred

sequences retrieved into account there is a significant difference in the mean number of amino acid substitutions from that found in a wild population, none of the mean number of amino acid substitutions in the studies analysed differed significantly from the mean of the natural population. However, some pairs of haplotypes used do have lower genetic distances between them e.g., the number of amino acid substitutions per site between haplotypes b and q (mean amino acid substitutions for Winsford = 11.57 ± 2.63, between haplotypes b and q = 16 ± 3.90, Z = -1.68, p = 0.046).

### 5.4.4 Phylogenetic reconstruction of inbred strains

We used sequence data from the MHC gene H2-Ab to construct a weighted maximum parsimony tree of the inbred strains of mice (Figure 5.3). Branch lengths are representative of the number of nucleotide differences between strains. Bootstrap values above 50% are shown at the nodes and provide confidence in relationships. For comparison, a schematic of the relationship between inbred strains based on the actual pedigrees of strains is shown. Only haplotypes whose origin can reliably be assigned are included. Haplotypes that are derived from strains that are completely unrelated to the majority of strains are also not included in the schematic, e.g., the u haplotype originates in the PL/J strain that is derived from non-inbred stock from Princeton and is completely unrelated to other inbred strains. The two trees show similarities at very close relationships such as those between strains carrying s and s2 haplotypes, between those with p and q haplotypes and the d haplotype is monophyletic to p and q in both trees. Other relationships derived from sequence data do not match the schematic showing known pedigrees.

Table 5.6. The mean number of amino acid substitutions (aa subs.) at the antigen binding regions of H2-Ab and H2-K1 between inbred haplotypes used in mate choice studies. Differences between previous studies and the distances found in a natural population (mean number of amino acid differences = 11.57 ± 2.63) have been tested and significant differences shown in bold.

| Study | Result | Haplotypes | Mean no. aa subs. | Z-tests |
|---|---|---|---|---|
| **Sherborne et al.** (2007) | Deficit in matings between mice sharing two MUP haplotypes, no significant effect of MHC. | wild | 11.93 ± 4.24 | Z = -0.14, p = 0.45 |
| **Yamazaki et al.** (1976; 1978) | Male mice from some strains showed disassortative mating preferences, others did not. | a, b, d and k | 11 ± 3.80 | Z = 0.22, p = 0.41 |
| Yamazaki et al. (1988) | Cross fostering of male mice reverses preference seen in two previous studies. | | | |
| Beauchamp et al. (1988) | Cross fostering of female mice does not reverse preference seen in two previous studies. | | | |
| **Yamazaki et al.** (2000) | Maternal recognition of progeny. | b and k | 11 ± 3.89 | Z = 0.22, p = 0.41 |
| Ninomiya and Brown (1995) | Removal of preputial glands removes preference for dissimilar congenic strain. | | | |
| **Potts et al.** (1991) | MHC deficiency in offspring. | b, d, k, and q | 12.17 ± 3.86 | Z = -0.23, p = 0.41 |
| Manning et al. (1992) | Communal nesting with MHC similar females. | | | |

Table 5.7. The mean number of amino acid substitutions (aa subs.) at the antigen binding regions of H2-Ab and H2-K1 between inbred haplotypes used in infection studies. Differences between previous studies and the distances found in a natural population (mean number of amino acid differences = 11.57 ± 2.63) have been tested and significant differences shown in bold.

| Study | Result | Haplotypes | Mean $\hat{d}$ | Significance |
|---|---|---|---|---|
| **Else et al.** (1988; 1990a; 1990b) | The resistance to *Trichuris muris* conferred by the b and q haplotypes was mapped to the H2-A region. | b, d, k, and q | 12.17 ± 3.86 | Z = -0.23, p = 0.41 |
| **Dick et al.** (1988) | Both *Trichinella spiralis* isolate and mouse strain had an effect on development of resistance. | k and q | 13 ± 1.98 | Z = -0.54, p = 0.29 |
| **Good et al.** (1986) | Only mice with the b haplotype recognise the *Plasmodium falciparum* fusion protein being tested as a vaccine. | b, d, k and s | 12.33 ± 1.94 | Z = -0.29, p = 0.39 |
| **Lee and Kasper** (2004) | Inoculum sizes of *Toxoplasma gondii* have different effects on different strains. | b and k | 11 ± 3.89 | Z = 0.22, p = 0.41 |

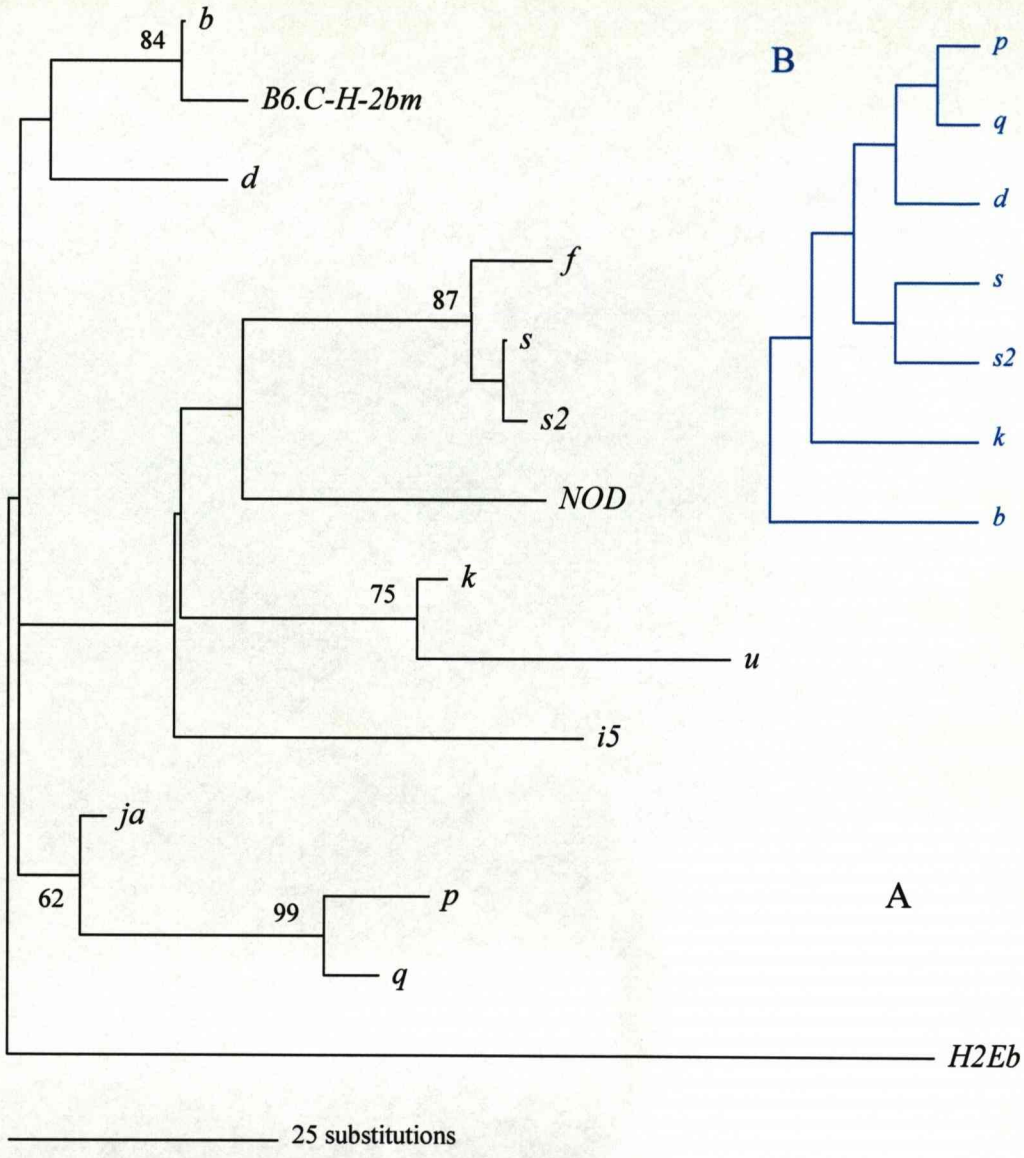



Figure 5.3. A. Parsimony analysis of H2-Ab sequence data for 13 inbred strains of mice. Branch lengths are representative of the number of substitutions between sequences. Values at each node represent the percentage (of 1000 replicates) that give the shown relationship. B. Schematic representing the relationship between MHC haplotypes as reconstructed from the genealogies of mouse inbred strains (Atchley and Fitch 1993; Beck et al. 2000).

## 5.5 DISCUSSION

I found a nine base-pair deletion in exon 2 of H2-Ab in both the Winsford population and individuals used in the enclosure study. An identical deletion is seen in several inbred strains; haplotypes f (strain B10.M), k (strain C3H), s (strain A.SW) and u (strain B10.PL); and has been characterised by Estess *et al.* (1986). Estess hypothesised that sequence similarity between alleles with and without the deletion, including maintenance of a silent base change, indicate that the deletion occurred prior to separation of the alleles in inbred strains. The finding of an identical deletion in a natural population supports this theory. The maintenance of both varieties of allele within a natural population suggests that the deletion has no immunological effect, or confers a complementary fitness benefit.

I analysed the genetic distance between MHC alleles using the amino acid sequences of exons coding for the antigen binding region of a class I (H2-K1) and class II (H2-Ab) gene. When considering the entire sequence, the MHC alleles from in the enclosure study (described in chapter 3) have a significantly higher mean amino acid divergence than found in a natural population. The animals bred for this study were the offspring of a laboratory maintained colony with wild animals from five different locations throughout the north-west of England in their ancestry. In an attempt to avoid the deleterious affects of inbreeding depression, the maintenance of the laboratory colony includes a number of pairings between animals derived from different trapping locations. Thus, although the founders of each enclosure population were the F1 offspring of a single male mated with multiple females, it is not surprising that there is more genetic divergence seen in this group of animals than in a single natural population.

When considering just the antigen binding sites of the MHC molecules a different pattern is observed. The natural population now has the lowest mean number of amino acid differences, with the inbred strains having a significantly more. Mouse strains are known to have mixtures of ancestral genomes from different *Mus musculus* sub-species (Yoshiki and Moriwaki 2006), the main contributors being *M. musculus musculus*, *M. musculus domesticus*, and *M. musculus castaneus*. Contributions of different sub-species would lead to large genetic distances between alleles. However, most inbred lines are derived from a small number of founder animals, in fact phylogenetic analysis of mitochondrial DNA from inbred strains suggests that all common strains descend from a single *Mus musculus domesticus* female (Dai et al. 2005; Goios et al. 2007). This would lead us to expect inbred strains to have lower levels of inter-strain MHC sequence diversity than found between, or even within natural populations. This does not appear to be the case in the functional regions of the MHC molecules sequenced in this study.

The mean genetic distance is significantly higher at the antigen binding site than at the rest of the MHC molecule, both for H2-Ab and H2-K1. These sites code for amino acids that interact with pathogenic antigens and thus the rate of fixation of nonsynonymous changes is expected to be higher at these codons producing large numbers of alleles with different antigen binding sites. The non-antigen binding site residues are structural and are under purifying selection, thus the rate of fixation nonsynonymous substitutions is lower. I observed the same trend as Oliver and Piertney (Oliver and Piertney 2006); the proportion of variable amino acids for the antigen binding region was over three times higher than the proportion or variable sites for the non-antigen binding region. This is manifest in the Kimura 2-parameter genetic distance estimators, as the total number of

nucleotide substitutions per site is higher in the antigen binding regions due to an increase in nonsynonymous changes.

Studies that involve comparisons between inbred strains of mice, including those investigating mate choice decisions, may not be representative of the genetic distances found in natural populations. I tested this directly with a selection of mate choice and parasite studies. Although some pairwise comparisons of the number of amino acid substitutions in the antigen binding site were significantly higher than the mean found in our natural population, the mean of each of the studies I looked at was not significantly different. In the mate choice studies, some of the MHC haplotypes the mice were in contact with have higher divergences than they would find in a natural population. Drawing conclusions from these sorts of comparisons must be therefore done carefully; if animals are unlikely to encounter large distances between MHC types in the wild, it is improbable that they will have evolved mechanisms to compare such distances and any evidence of decisions based on MHC type may instead be due to a correlated gene or genes. Contrary to this, the studies looking at parasite resistance are repeatedly using a limited number of MHC haplotypes. While inbred strains can be useful in determining what aspects of susceptibility are controlled by the MHC, care must be taken if using these strains as a model for host-parasite interactions in the wild as they may not represent the genetic or allelic diversity seen in natural populations.

There is no significant difference between the rates of synonymous substitution per synonymous site ($\bar{d}_S$) and nonsynonymous substitution per nonsynonymous site ($\bar{d}_N$) in the non-antigen binding sites in all three groups. When $d_S$ is equal to $d_N$ a gene is said to be in a state of neutral evolution and this appears to be the case in the regions of the MHC genes coding for structural residues. $\bar{d}_N$ is significantly higher than $\bar{d}_S$ in the

antigen binding region, suggesting these residues are under balancing selection. This is in agreement with previous studies that also find evidence of balancing selection acting on the antigen binding regions (Paterson 1998; Meyer-Lucht and Sommer 2005; Worley et al. 2006).

The $\bar{d}_N$ rate is greatest in the antigen binding site of the inbred strains, giving the highest ratio of synonymous to nonsynonymous substitutions ($\bar{d}_N/\bar{d}_S = 5.522 \pm 0.248$). Thus, there is the greatest evidence of selection acting at the antigen binding regions of the MHC molecule in the inbred strains tested. One explanation for this is that the MHC haplotypes present in inbred strains were derived from very different populations with their own unique pathogenic communities. This would mean that the antigen binding regions of the alleles that went to fixation in each strain would be adapted to their own range of pathogens and thus they would be divergent from each other. In fact, the majority of inbred strains used regularly are derived from crosses between mice of several sub-species (Yoshiki and Moriwaki 2006), so there is a lot of evolutionary time between the MHC alleles introduced into inbred strains and their most recent common ancestor leading to large divergences. Individuals within in a single population such as the one at Winsford are likely to encounter the same repertoire of pathogens and have more recent common ancestral MHC sequences, thus the diversity seen between the functional regions of their MHC alleles would not be as large as that between animals taken from separate populations. Another explanation is that an increase in the maintenance of nonsynonymous mutations has occurred after the formation of the inbred strains. There is supposedly a lack of pathogen driven selection acting on the MHC of inbred strains, which are maintained in near sterile conditions. Detrimental mutations in antigen binding site residues will not be selected against and will be maintained in the strain, where they would have been eradicated due to a loss in fitness in a natural

population. However, Carroll et al. (2002) estimated that the number of new mutations between strains separated by 100 years would be as low as 0.068 for each MHC gene, even taking into account the high mutation rate at these genes. Thus it may be the former hypothesis, or a combination of the two, that result in the higher $\bar{d}_N/\bar{d}_S$ ratio seen.

I reconstructed the phylogeny of inbred stains of mice using MHC sequence data. Although the phylogeny mirrors the very close relationships calculated from known pedigrees, the majority of more distant associations disagree with pedigree information. This is in agreement with a study by Atchley and Finch (1991) who used different subsets of genetic data to reconstruct phylogenetic relationships of inbred strains which they then compared to genealogical data. When using immune genes alone they generated a tree that significantly differed from relationships derived from pedigree data. There are several hypotheses explaining this lack of correlation. One is that immune genes, and in particular the MHC, had large numbers of alleles in natural outbred populations that were independently fixed by random drift during the development of inbred strains. Thus two different strains generated from the same population could have very different MHC alleles due to the high diversity of the MHC in the founding population. In fact results from a genome wide study looking at genetic diversity between two inbred strains describe a mosaic structure, with areas of low diversity where the strains show a very recent common origin, and areas of high diversity where ancestry can be traced back to the three subspecies that generated inbred strains: *M. m. musculus*, *M. m. domesticus* and *M. m. castaneus* (Wade et al. 2002). Thus when we see distant relationships between MHC alleles it may be because there were different subspecies contributing to the generation of the two strains.

I have measured the number of amino acid substitutions in pairwise comparisons for class I and class II MHC genes for a population of wild

mice, mice used in our enclosure study and inbred strains. When considering the antigen binding residues alone, I found a significant increase in the mean number of amino acid substitutions in the inbred sequences when compared to the natural population.

# Chapter 6

## GENERAL DISCUSSION

The MHC has many unique genetic and functional properties that make it a fascinating region of the genome. It has therefore been much researched, with studies investigating the high level of diversity of MHC genes, the influence of pathogen driven selection on the maintenance of this diversity and, most recently, the hypothesised role of the MHC in mate choice and sexual selection (reviewed by Meyer and Thomson (2001), Jeffery and Bangham (2000) and Penn and Potts (1999) respectively). Many of these studies have used inbred strains of mice as their model species. While inbred strains, and especially MHC congenic strains, can be invaluable when investigating the immunological effects of individual MHC genes, studies investigating MHC variation and the behavioural consequences of MHC variation should use caution when extrapolating results from strains that have been experimentally inbred for hundreds of generations. My thesis uniquely investigates MHC diversity, parasite resistance and mate choice driven selection using completely wild-derived *Mus musculus domesticus*.

Despite substantial interest in the MHC of wild-populations (reviewed by Piertney and Oliver (2006)), there has been little recent investigation of diversity and selection of the MHC in wild house mice. A study of some 480 mice trapped from 12 different populations in North America, South America, Europe, North Africa and the Middle East used serological methods to characterise diversity at the MHC (Duncan et al. 1979; Gotze et al. 1980a; Nadeau et al. 1981; Neufeld et al. 1986). They predicted that there were potentially up to 100 alleles at the H2-K and H2-D loci in

mouse populations world-wide. They also found striking levels of heterozygosity at H2-K and H2-D (95%), and at H2-Ab and H2-Eb loci (85%) (reviewed by Klein and Figueroa (1981)). Heterozygosity levels exceeding those expected under neutral evolution are indicative of balancing selection acting on the MHC (Hedrick and Thomson 1983; Takahata et al. 1992). The work in Chapter 4 found similarly increased levels of heterozygosity and evidence of balancing selection acting on the MHC of a wild population of house mice in the northwest of England. Tests using MHC-linked microsatellite markers (Watterson's test, Table 4.4 and Figure 4.4) found evidence of balancing selection, as did tests looking at polymorphisms and allelic frequencies (Tajima's *D*, Table 4.7). Comparing rates of synonymous and nonsynonymous substitution allowed the selection pressure to be traced to the regions coding for the antigen binding sites in the wild population (Table 4.5 and Table 4.6) and also within a sub-sample of sequences from inbred strains of mice (Table 5.5). Only one class I and one class II gene were sequenced in this study. Characterisation at further genes would give an indication of how selection varies across the MHC region. Future investigation of MHC gene sequences in different populations could offer support to the current study and allow comparison of between population diversity. The driving force maintaining balancing selection at the MHC is also a topic for further investigation.

Both natural and sexual selection have been implicated in the maintenance of MHC polymorphism (Potts et al. 1994), although the exact mechanisms have not been determined. Natural selection may be mediated through host-parasite interactions (Hughes and Hughes 1995; Meyer and Thomson 2001) and both laboratory based studies and studies of natural populations have found effects of MHC heterozygosity on resistance to pathogens (Nevo and Beiles 1992; Penn et al. 2002; McClelland et al. 2003; Wegner

et al. 2003; Wegner et al. 2004; Meyer-Lucht and Sommer 2005). Sexual selection could maintain high levels of MHC heterozygosity through MHC disassortative mating (Penn and Potts 1999). MHC-mediated mate choice was first investigated when Yamazaki (1976) noticed male preference for MHC-dissimilar females in inbred laboratory strains of mice. Since then there has been evidence both for (Yamazaki et al. 1976; Beauchamp et al. 1988; Yamazaki et al. 1988; Potts et al. 1991; Potts et al. 1994; Wedekind et al. 1995; Reusch et al. 2001; Milinski et al. 2005), and against the role of the MHC in mate choice (Hedrick and Black 1997; Paterson and Pemberton 1997; Landry et al. 2001; Sauermann et al. 2001; Ekblom et al. 2004; Westerdahl 2004). There was no evidence of MHC-biased mating in a large-scale study investigating inbreeding avoidance in semi-natural populations of wild-derived mice (Chapter 4 and Sherborne et al. 2007). Instead, the mice appeared to be using another highly polymorphic locus, MUP, to avoid mating with close kin.

Major urinary proteins are similar to MHC molecules in that they are encoded by a highly polymorphic gene cluster and they have a specialised binding groove that binds short peptides (Robertson et al. 1993; Beynon et al. 2002). The sole known role of MUPs is in chemical signalling and an individual's MUP expression pattern may contain up to a dozen different proteins. It is these properties that have made MUPs a candidate for individual recognition and mate choice in the wild (Hurst et al. 2001; Brennan 2004; Cheetham et al. 2007). The finding of the enclosure study (Chapter 3), that the dearth in matings between full siblings can be fully explained by a significant deficit in the number of successful matings between individuals that share both MUP haplotypes (Table 3.2, Model 4), supports these previous hypotheses, and suggests a role for MUPs in the avoidance of inbreeding. This study generated a huge amount of data as we took genetic, blood and urine samples for over 80 founding individuals and

nearly 900 offspring. Some of this data is already being analysed, for example the MUP phenotypes are currently being determined and related to genotypes. As parentage was determined, individual reproductive success can be correlated to measurements such as body mass, MHC and MUP genotype or genome-wide heterozygosity.

Genome-wide heterozygosity is believed to confer fitness benefits through heterozygote advantage and the masking of deleterious recessive alleles (Mitton 1993; David 1998; Hansson and Westerberg 2002). It has long been accepted that females value heterozygosity in their offspring. More recently it has been suggested that they may also value heterozygosity in their mates, as heterozygosity correlates with indicators of fitness such as dominance, male ornaments and sperm quality (Brown 1997; Kempenaers 2007b). Both the MHC and MUP have been suggested as signalling systems that advertise the level of genome-wide heterozygosity in a potential mate. MHC and MUP heterozygosity is likely to correspond to genome-wide heterozygosity in a natural population, and thus assessing the levels of heterozygosity at these loci could provide a mechanism for assessing genome-wide heterozygosity. However, no evidence was found to suggest that MHC and MUP heterozygosity had an effect on the reproductive success of male mice (Table 3.4). Instead, weight and genome-wide heterozygosity itself were found to have a significant effect on the numbers of offspring sired by males. Heavier and more heterozygous males sired significantly more offspring than lighter and more homozygous males. This is another blow for the role of the MHC in mate choice. As well as finding no support for the role of the MHC in the recognition and avoidance of mates as potential kin, no evidence that MHC heterozygosity has any intrinsic benefits for reproductive success was found.

The urine scent trails in Chapter 2 failed to find any female preference for MHC heterozygous males over MHC homozygous males (Figure 2.5), a further indication that MHC heterozygosity may not be used to represent genome-wide heterozygosity. Interestingly, females did show a preference for the urine of uninfected males over males infected with the intestinal nematode *T. muris* (Figure 2.4). Further work is currently underway to look for the molecular signal of infection using mass spectrometry on the urine of infected and uninfected males. This aversion for the odour of infected individuals is in support of previous studies showing that animals will avoid the odour of infected cohorts (Kavaliers and Colwell 1993; Kavaliers and Colwell 1995b; Willis and Poulin 2000). It also suggests that females are using a much more direct assessment of male immunological fitness than assessing MHC heterozygosity or phenotype; they are simply avoiding males whose susceptibility to parasitic infection has indicated that they may have substandard immune genes. This is similar to the outcome of the reproductive success study (Chapter 3) where there was no evidence of a complex reporter system for genome-wide heterozygosity; instead body-size and genome-wide heterozygosity itself correlated with reproductive success. It seems intuitive that females should assess potential mates by the simplest means possible, instead of investing resources in investigating genetic markers that may be correlated with heterozygosity or fitness.

In chapter 2, I show that a parasite that has been an effective infectious agent in inbred strains of mice, *Trichuris muris* (Wakelin 1967; Else and deSchoolmeester 2003; Cliffe and Grencis 2004), is readily expelled from wild-derived mice. Susceptibility to *T. muris* has been shown to have both MHC and non-MHC determined components (Else and Wakelin 1988). Differing MHC genotypes were compared on similar genetic backgrounds, thus this study had the potential to determine whether resistance was

mediated by the MHC genes present in the wild population or other immune genes. In fact, there was no influence of MHC heterozygosity or genotype on the expulsion of *T. muris* (all individuals completely expelled the parasite) or on the level of parasite-specific antibody produced (Figure 2.2), suggesting that either the three MHC haplotypes tested were equally good at recognising pathogenic *T. muris* antigen, or background genes had a greater effect on resistance than genes mapping to the MHC. Further serological techniques would help to define the role of different genes in resistance to the parasite. For example, as in a study on *T. muris* infections in MHC-congenic inbred strains (Else et al. 1990b), immunoblots would show which pathogenic antigens are being recognised by which antibodies during infection, control of which Else et al. showed mapped to the MHC region. A larger study in which worm burdens could be assessed at different time points post infection would allow the speed of expulsion to be determined. Further work on this system would help to disentangle the effects of MHC and non-MHC linked genes on resistance and could elucidate why some inbred strains appear to be completely susceptible to *T. muris*, even though susceptibility appears to be rare in the wild.

Inbred strains of mice are one of the most widely used model organisms in animal studies. Complete homozygosity within strains and genetic variation between strains means they are useful tools in determining the influence of genetic polymorphisms on phenotypes. I analysed the average genetic distance between MHC sequences in a natural population and between MHC sequences of inbred strains. I found no significant difference between the mean genetic distance found in a natural population and those found in eight studies investigating mate choice, and four studies investigating parasite resistance using inbred strains of mice. However, in some of the studies that test female reaction to two different male odours, the genetic distance between the MHC sequences used was greater than

those that would be encountered in a natural population. Some interesting future work would be to recreate previous studies that have used inbred mice using animals from the wild-derived population maintained at Leahurst to further test the relevance of using inbred animals to represent wild behaviour or immuno-competence (care would need to be taken in experimental design to take into account the variation across the genome present in wild animals, but absent in inbred strains).

The finding that house mice use the MHC in the avoidance of inbreeding spurred the investigation of the role of the MHC in mate choice in other species; including birds, (Zelano and Edwards 2002; Ekblom et al. 2004; Westerdahl 2004), fish (Milinski et al. 2005; Pitcher and Neff 2006) and humans (Roberts et al. 2005). If mate choice is not influenced by the MHC in mice, what does that mean for its role in other species? The MHC was shown to play a role in mate choice in some non-rodent species, but some studies found no influence of MHC genes on mating patterns (Westerdahl 2004; Paterson 2005). If the MHC is not acting as a cue for inbreeding avoidance in all species investigated, could MUP? MUPs have been shown to influence individual recognition and inbreeding avoidance (Chapter 3) in mice, however MUP orthologs are poorly investigated in non-rodent species (although see Dandeu et al. 1995). It is unlikely that MUPs are acting as a mechanism for inbreeding avoidance in all vertebrate species however, as even within the Mus genus they have proved to be monomorphic in some sub-species (Robertson et al. 2007). Whether inbreeding avoidance mechanisms are required at all is dependent on the mating and dispersal system of a species (Moore and Ali 1984; Caballero et al. 1996; Perrin and Mazalov 2000; Lehmann and Perrin 2003). It is possible that species or taxa-specific mechanisms have evolved to facilitate the avoidance of inbreeding at a level required by the mating behaviour of a particular species. It may be important for future studies to focus on

species-specific cues of inbreeding avoidance, rather than looking for a vertebrate wide marker.

In conclusion, the work in this thesis describes the role of balancing selection in maintaining the high levels of polymorphism and heterozygosity found at the major histocompatibility complex of mice. The cause of this selection pressure was investigated, and there was no evidence to suggest a role for mate choice and sexual selection. Females did not appear to be assessing mate quality through MHC mediated changes in urinary odour. MHC heterozygosity had no effect on reproductive success in males. There was no evidence of a role for the MHC in the avoidance of inbreeding. This suggests that the balancing selection seen at the MHC must be a result of natural parasite driven selection.

# References


Able, D. J. (1996). The contagion indicator hypothesis for parasite-mediated sexual selection. *Proceedings of the National Academy of Sciences of the United States of America 93:* 2229-2233

Acevedo-Whitehouse, K., Gulland, F., Greig, D., and Amos, W. (2003). Disease susceptibility in California sea lions. *Nature 422:* 35-35

Aidoo, M., Terlouw, D. J., Kolczak, M. S., McElroy, P. D., ter Kuile, F. O., Kariuki, S., Nahlen, B. L., Lal, A. A., and Udhayakumar, V. (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet 359:* 1311-2

Alberts, S. C. and Ober, C. (1993). Genetic-variability in the major histocompatibility complex - a review of non-pathogen-mediated selective mechanisms. *Yearbook of Physical Anthropology, Vol 36 1993*, pp. 71-89

Allison, A. C. (1964). Polymorphism and Natural Selection in Human Populations. *Cold Spring Harb Symp Quant Biol 29:* 137-49

Andersson, M. (1994). Sexual Selection. Princeton University Press, Princeton

Apanius, V., Penn, D., Slev, P. R., Ruff, L. R., and Potts, W. K. (1997a). The nature of selection on the major histocompatibility complex. *Crit Rev Immunol 17:* 179-224

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

Arcaro, K. F. and Eklund, A. (1998). A review of MHC-based mate preferences and fostering experiments in two congenic strains of mice. *Genetica 104:* 241-244

Arkush, K. D., Giese, A. R., Mendonca, H. L., McBride, A. M., Marty, G. D., and Hedrick, P. W. (2002). Resistance to three pathogens in the endangered winter-run chinook salmon (Oncorhynchus tshawytscha): effects of inbreeding and major histocompatibility complex genotypes. *Canadian Journal of Fisheries and Aquatic Sciences 59:* 966-975

Arnold, B., Burgert, H. G., Archibald, A. L., and Kvist, S. (1984). Complete nucleotide sequence of the murine H-2Kk gene. Comparison of three H-2K locus alleles. *Nucleic Acids Res 12:* 9473-87

Atchley, W. R. and Fitch, W. (1993). Genetic Affinities of Inbred Mouse Strains of Uncertain Origin. *Molecular Biology and Evolution 10:* 1150-1169

Atchley, W. R. and Fitch, W. M. (1991). Gene trees and the origins of inbred strains of mice. *Science 254:* 554-8

Bailey, D. W. and Kohn, H. I. (1965). Inherited Histocompatibility Changes in Progeny of Irradiated and Unirradiated Inbred Mice. *Genetical Research 6:* 330-&

Bancroft, A. J., Else, K. J., Humphreys, N. E., and Grencis, R. K. (2001). The effect of challenge and trickle Trichuris muris infections on the polarisation of the immune response. *International Journal for Parasitology 31:* 1627-1637

Bard, J., Yamazaki, K., Curran, M., Boyse, E. A., and Beauchamp, G. K. (2000). Effect of B2m gene disruption on MHC-determined odortypes. *Immunogenetics 51:* 514-518

Bateson, P. (1978). Sexual Imprinting and Optimal Outbreeding. *Nature 273:* 659-660

Beauchamp, G. K., Yamazaki, K., Bard, J., and Boyse, E. A. (1988). Preweaning experience in the control of mating preferences by genes in the major histocompatibility complex of the mouse. *Behav Genet 18:* 537-47

Beck, J. A., Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J. T., Festing, M. F., and Fisher, E. M. (2000). Genealogies of mouse inbred strains. *Nat Genet 24:* 23-5

Behnke, J. M. and Wakelin, D. (1973). The survival of Trichuris muris in wild populations of its natural hosts. *Parasitology 67:* 157-64

Belkhir, K., Castric, V., and Bonhomme, F. (2002). IDENTIX, a software to test for relatedness in a population using permutation methods. *Molecular Ecology Notes 2:* 611-614

Beynon, R. J. and Hurst, J. L. (2003). Multiple roles of major urinary proteins in the house mouse, Mus domesticus. *Biochemical Society Transactions 31:* 142-146

Beynon, R. J., Veggerby, C., Payne, C. E., Robertson, D. H. L., Gaskell, S. J., Humphries, R. E., and Hurst, J. L. (2002). Polymorphism in major urinary proteins: Molecular heterogeneity in a wild mouse population. *Journal of Chemical Ecology 28:* 1429-1446

Bjorkman, P. J. and Parham, P. (1990). Structure, Function, and Diversity of Class I Major Histocompatibility Complex Molecules. *Annual Review of Biochemistry 59:* 253-288

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., and Wiley, D. C. (1987). Structure of the Human Class-I Histocompatibility Antigen, Hla-A2. *Nature 329:* 506-512

Blackwell, N. M. and Else, K. J. (2001). B cells and antibodies are required for resistance to the parasitic gastrointestinal nematode Trichuris muris. *Infection and Immunity 69:* 3860-3868

Blackwell, N. M. and Else, K. J. (2002). A comparison of local and peripheral parasite-specific antibody production in different strains of mice infected with Trichuris muris. *Parasite Immunology 24:* 203-211

Blouin, M. S., Parsons, M., Lacaille, V., and Lotz, S. (1996). Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology 5:* 393-401

Bodmer, W. F. (1972). Evolutionary significance of the HL-A system. *Nature 237:* 139-45 passim

Brennan, P. A. (2004). The nose knows who's who: chemosensory individuality and mate recognition in mice. *Horm Behav 46:* 231-40

Brennan, P. A. and Kendrick, K. M. (2006). Mammalian social odours: attraction and individual recognition. *Philosophical Transactions of the Royal Society B-Biological Sciences 361:* 2061-2078

Bretman, A., Wedell, N., and Tregenza, T. (2004). Molecular evidence of post-copulatory inbreeding avoidance in the field cricket Gryllus bimaculatus. *Proceedings of the Royal Society of London Series B-Biological Sciences 271:* 159-164

Bronson, F. H. (1979). Reproductive Ecology of the House Mouse. *Quarterly Review of Biology 54:* 265-299

Brown, J. H., Jardetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J., and Wiley, D. C. (1988). A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature 332:* 845-50

Brown, J. L. (1997). A theory of mate choice based on heterozygosity. *Behavioral Ecology 8:* 60-65

Brown, J. L. (1998). The new heterozygosity theory of mate choice and the MHC. *Genetica 104:* 215-21

Brown, J. L. and Eklund, A. (1994). Kin Recognition and the Major Histocompatibility Complex - an Integrative Review. *American Naturalist 143:* 435-461

Buckle, G. R. and Greenberg, L. (1981). Nestmate recognition in sweat bees (Lasioglossum zephyrum) - does an individual recognize its own odour or only odours of its nestmates. *Animal Behaviour 29:* 802-809

Caballero, A., Santiago, E., and Toro, M. A. (1996). Systems of mating to reduce inbreeding in selected populations. *Animal Science 62:* 431-442

Callaway, E. (2007). The mouse map gets a lot more signposts. *Nature 448:* 516-517

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K., and O'Brien, S. J. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science 283:* 1748-52

Carroll, L. S., Penn, D. J., and Potts, W. K. (2002). Discrimination of MHC-derived odors by untrained mice is consistent with divergence in peptide-binding region residues. *Proc Natl Acad Sci U S A 99:* 2187-92

Casadevall, A. and Pirofski, L. A. (1999). Host-pathogen interactions: Redefining the basic concepts of virulence and pathogenicity. *Infection and Immunity 67:* 3703-3713

Cavaggioni, A. and Mucignat-Caretta, C. (2000). Major urinary proteins, alpha(2u)-globulins and aphrodisin. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1482:* 218-228

Chamero, P., Marton, T. F., Logan, D. W., Flanagan, K., Cruz, J. R., Saghatelian, A., Cravatt, B. F., and Stowers, L. (2007). Identification of protein pheromones that promote aggressive behaviour. *Nature 450:* 899-U23

Charlesworth, B. and Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetical Research 74:* 329-340

Charlesworth, D. and Awadalla, P. (1998). Flowering plant self-incompatibility: the molecular population genetics of Brassica S-loci. *Heredity 81:* 1-9

Charlesworth, D. and Charlesworth, B. (1987). Inbreeding Depression and Its Evolutionary Consequences. *Annual Review of Ecology and Systematics 18:* 237-268

Cheetham, S. A., Thom, M. D., Jury, F., Ollier, W. E., Beynon, R. J., and Hurst, J. L. (2007). The genetic basis of individual-recognition signals in the mouse. *Curr Biol 17:* 1771-7

Cliffe, L. J. and Grencis, R. K. (2004). The Trichuris muris system: a paradigm of resistance and susceptibility to intestinal nematode infection. *Advances in Parasitology, Vol 57*, pp. 255-307

Colegrave, N., Kotiaho, J. S., and Tomkins, J. L. (2002). Mate choice or polyandry: reconciling genetic compatibility and good genes sexual selection. *Evolutionary Ecology Research 4:* 911-917

Coltman, D. W., Bowen, W. D., and Wright, J. M. (1998). Birth weight and neonatal survival of harbour seal pups ape positively correlated with genetic variation measured by microsatellites. *Proceedings of the Royal Society of London Series B-Biological Sciences 265:* 803-809

Comings, D. E. and MacMurray, J. P. (2000). Molecular heterosis: A review. *Molecular Genetics and Metabolism 71:* 19-31

Cordero, P. J., Aparico, J. M., and Veiga, J. P. (2004). Parental genetic characteristics and hatching success in the spotless starling, Sturnus unicolor. *Animal Behaviour 67:* 637-642

Cornuet, J. M. and Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics 144:* 2001-2014

Dai, J. G., Min, J. X., Xiao, Y. B., Lei, X., Shen, W. H., and Wei, H. (2005). The absence of mitochondrial DNA diversity among common laboratory inbred mouse strains. *Journal of Experimental Biology 208:* 4445-4450

Dandeu, J. P., Rabillon, J., Carmileroy, A., Divanovic, A., Camoin, L., and David, B. (1995). EQU-CL, the horse major allergen and its close structural relationship to the mouse and rat major urinary protein. *Journal of Allergy and Clinical Immunology 95:* 348-348

David, P. (1998). Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity 80:* 531-537

Davisson, M. T. (1996). Rules for nomenclature of inbred strains. *In* M. F. Lyon, Rastan, S. & Brown, S.D.M. (ed.): *Genetic Variants and Strains of the Laboratory Mouse* pp. 1532–1536, Oxford University Press, Oxford

Dawkins, R. (1982). The extended phenotype. W. H. Freeman, San Francisco

Defries, J. C. and McClearn, G. E. (1970). Social dominance and darwinian fitness in laboratory mouse. *American Naturalist 104:* 408-&

Desjardins, C., Maruniak, J. A., and Bronson, F. H. (1973). Social rank in house mice: differentiation revealed by ultraviolet visualization of urinary marking patterns. *Science 182:* 939-41

Dewsbury, D. A. (1988). Kin Discrimination and Reproductive-Behavior in Muroid Rodents. *Behavior Genetics 18:* 525-536

Dick, T. A., Dougherty, D. A., and Wassom, D. L. (1988). Trichinella spiralis infections of inbred mice - genetics of the host response following infection with different Trichinella isolates. *Journal of Parasitology 74:* 665-669

Ditchkoff, S. S., Hoofer, S. R., Lochmiller, R. L., Masters, R. E., and Van Den Bussche, R. A. (2005). MHC-DRB evolution provides insight into parasite resistance in white-tailed deer. *Southwestern Naturalist 50:* 57-64

Dobzhansky, T. (1948). Genetics of Natural Populations .18. Experiments on Chromosomes of Drosophila-Pseudoobscura from Different Geographic Regions. *Genetics 33:* 588-602

Doherty, P. C. and Zinkernagel, R. M. (1975). Enhanced Immunological Surveillance in Mice Heterozygous at H-2 Gene Complex. *Nature 256:* 50-52

Duncan, W. R., Wakeland, E. K., and Klein, J. (1979). Heterozygosity of H-2 loci in wild mice. *Nature 281:* 603-5

Durel, C. E., Bertin, P., and Kremer, A. (1996). Relationship between inbreeding depression and inbreeding coefficient in maritime pine (Pinus pinaster). *Theoretical and Applied Genetics 92:* 347-356

Edwards, S. V. and Hedrick, P. W. (1998). Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution 13:* 305-311

Egid, K. and Brown, J. L. (1989). The Major Histocompatibility Complex and Female Mating Preferences in Mice. *Animal Behaviour 38:* 548-550

Ehman, K. D. and Scott, M. E. (2002). Female mice mate preferentially with non-parasitized males. *Parasitology 125:* 461-6

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

Eklund, A., Egid, K., and Brown, J. L. (1991). The major histocompatibility complex and mating preferences of male mice. *Animal Behaviour 42:* 693-694

Else, K. and Wakelin, D. (1988). The effects of H-2 and non-H-2 genes on the expulsion of the nematode Trichuris muris from inbred and congenic mice. *Parasitology 96 ( Pt 3):* 543-50

Else, K. J. and deSchoolmeester, M. L. (2003). Immunity to Trichuris muris in the laboratory mouse. *J Helminthol 77:* 95-8

Else, K. J. and Grencis, R. K. (1991). Cellular Immune-Responses to the Murine Nematode Parasite Trichuris-Muris .1. Differential Cytokine Production during Acute or Chronic Infection. *Immunology 72:* 508-513

Else, K. J. and Wakelin, D. (1990). Genetically-Determined Influences on the Ability of Poor Responder Mice to Respond to Immunization against Trichuris-Muris. *Parasitology 100:* 479-489

Else, K. J., Wakelin, D., Wassom, D. L., and Hauda, K. M. (1990a). The influence of genes mapping within the major histocompatibility complex on resistance to Trichuris muris infections in mice. *Parasitology 101 Pt 1:* 61-7

Else, K. J., Wakelin, D., Wassom, D. L., and Hauda, K. M. (1990b). MHC-restricted antibody responses to Trichuris Muris excretory/secretory (E/S) antigen. *Parasite Immunology 12:* 509-527

Eppig, J. T., Blake, J.A., Bult, C.J., Kadin, J.A., Richardson, J.E. and the members of the Mouse Genome Database Group (2007). The Mouse Genome Database (MGD): new features facilitating a model system. *Nucleic Acids Research 35:* D630-D637

Estess, P., Begovich, A. B., Koo, M., Jones, P. P., and McDevitt, H. O. (1986). Sequence analysis and structure-function correlations of murine q, k, u, s, and f haplotype I-A beta cDNA clones. *Proc Natl Acad Sci U S A 83:* 3594-8

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology 3:* 87-&

Figueroa, F., Gunther, E., and Klein, J. (1988). Mhc Polymorphism Predating Speciation. *Nature 335:* 265-267

Firman, R. C. and Simmons, L. W. (2008). Polyandry facilitates postcopulatory inbreeding avoidance in house mice. *Evolution 62:* 603-611

Fischer Lindahl, K. (1997). On naming H2 haplotypes: functional significance of MHC class Ib alleles. *Immunogenetics 46:* 53-62

Franks, N. R. and Scovell, E. (1983). Dominance and reproductive success among slave-making worker ants. *Nature 304:* 724-725

Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A., and Wilson, I. A. (1992). Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science 257:* 919-27

Gage, M. J. G., Surridge, A. K., Tomkins, J. L., Green, E., Wiskin, L., Bell, D. J., and Hewitt, G. M. (2006). Reduced heterozygosity depresses sperm quality in wild rabbits, Oryctolagus cuniculus. *Current Biology 16:* 612-617

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E., and Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature 384:* 134-141

Garrigan, D. and Hedrick, P. W. (2003). Perspective: Detecting adaptive molecular polymorphism: Lessons from the MHC. *Evolution 57:* 1707-1722

Gerlinskaya, L. A., Frolova, Y. A., Kondratyuk, E. Y., and Moshkin, M. P. (2007). Scent-marking costs and reproductive success in male mice of outbred strain ICR. *Zhurnal Obshchei Biologii 68:* 296-306

Goios, A., Pereira, L., Bogue, M., Macaulay, V., and Amorim, A. (2007). mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Research 17:* 293-298

Good, M. F., Berzofsky, J. A., Maloy, W. L., Hayashi, Y., Fujii, N., Hockmeyer, W. T., and Miller, L. H. (1986). Genetic control of the immune response in mice to a Plasmodium falciparum sporozoite vaccine. *Journal of Experimental Medicine 164:* 655-660

Gotze, D., Nadeau, J., Wakeland, E. K., Berry, R. J., Bonhomme, F., Egorov, I. K., Hjorth, J. P., Hoogstraal, H., Vives, J., Winking, H., and Klein, J. (1980a). Histocompatibility-2 System in Wild Mice .10. Frequencies of H-2 and Ia Antigens in Wild Mice from Europe and Africa. *Journal of Immunology 124:* 2675-2681

Gotze, D., Nadeau, J., Wakeland, E. K., Berry, R. J., Bonhomme, F., Egorov, I. K., Hjorth, J. P., Hoogstraal, H., Vives, J., Winking, H., and Klein, J. (1980b). Histocompatibility-2 system in wild mice. X. Frequencies of H-2 and Ia antigens in wild mice from Europe and Africa. *J Immunol 124:* 2675-81

Greenberg, L. (1979). Genetic Component of Bee Odor in Kin Recognition. *Science 206:* 1095-1097

Grob, B., Knapp, L. A., Martin, R. D., and Anzenberger, G. (1998). The major histocompatibility complex and mate choice: Inbreeding avoidance and selection of good genes. *Experimental and Clinical Immunogenetics 15:* 119-129

Guo, S. W. and Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics 48:* 361-372

Hale, R. E. and St Mary, C. M. (2007). Nest tending increases reproductive success, sometimes: environmental effects on paternal care and mate choice in flagfish. *Animal Behaviour 74:* 577-588

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

Hamilton, W. D. and Zuk, M. (1982). Heritable True Fitness and Bright Birds - a Role for Parasites. *Science 218:* 384-387

Hansson, B. and Westerberg, L. (2002). On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology 11:* 2467-2474

Hawley, D. M., Sydenstricker, K. V., Kollias, G. V., and Dhondt, A. A. (2005). Genetic diversity predicts pathogen resistance and cell-mediated immunocompetence in house finches. *Biology Letters 1:* 326-329

Hayes, K. S., Bancroft, A. J., and Grencis, R. K. (2004). Immune-mediated regulation of chronic intestinal nematode infection. *Immunological Reviews 201:* 75-88

Hedrick, P. W. (1992). Female Choice and Variation in the Major Histocompatibility Complex. *Genetics 132:* 575-581

Hedrick, P. W. (1994). Evolutionary genetics of the major histocompatibility complex. *American Naturalist 143:* 945-964

Hedrick, P. W. and Black, F. L. (1997). HLA and mate selection: No evidence in South Amerindians. *American Journal of Human Genetics 61:* 505-511

Hedrick, P. W. and Kalinowski, S. T. (2000). Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics 31:* 139-162

Hedrick, P. W. and Thomson, G. (1983). Evidence for Balancing Selection at Hla. *Genetics 104:* 449-456

Hoffman, J. I., Boyd, I. L., and Amos, W. (2004). Exploring the relationship between parental relatedness and male reproductive success in the antarctic fur seal Arctocephalus gazella. *Evolution 58:* 2087-2099

Hoogland, J. L. (1992). Levels of inbreeding among prairie dogs. *American Naturalist 139:* 591-602

Huang, S. W. and Yu, H. T. (2003). Genetic variation of microsatellite loci in the major histocompatibility complex (MHC) region in the southeast Asian house mouse (Mus musculus castaneus). *Genetica 119:* 201-218

Hubbard, T. J. P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E.,

Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Overduin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., and Birney, E. (2007). Ensembl 2007. *Nucleic Acids Research 35:* D610-D617

Huck, U. W., Lisk, R. D., and McKay, M. V. (1988). Social-dominance and reproductive success in pregnant and lactating golden-hamsters (Mesocicetus-Auratus) under semi-natural conditions. *Physiology & Behavior 44:* 313-319

Hughes, A. L. and Hughes, M. K. (1995). Natural Selection on the Peptide-Binding Regions of Major Histocompatibility Complex-Molecules. *Immunogenetics 42:* 233-243

Hughes, A. L. and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature 335:* 167-70

Hughes, A. L. and Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci U S A 86:* 958-62

Hughes, A. L. and Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics 32:* 415-435

Hurst, J. L. and Beynon, R. J. (2004). Scent wars: the chemobiology of competitive signalling in mice. *Bioessays 26:* 1288-1298

Hurst, J. L., Payne, C. E., Nevison, C. M., Marie, A. D., Humphries, R. E., Robertson, D. H., Cavaggioni, A., and Beynon, R. J. (2001). Individual recognition in mice mediated by major urinary proteins. *Nature 414:* 631-4

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

Jennions, M. D., Moller, A. P., and Petrie, M. (2001). Sexually selected traits and adult survival: A meta-analysis. *Quarterly Review of Biology 76:* 3-36

Jones, A. G. and Ardren, W. R. (2003). Methods of parentage analysis in natural populations. *Mol Ecol 12:* 2511-23

Jordan, W. C. and Bruford, M. W. (1998). New perspectives on mate choice and the MHC. *Heredity 81:* 239-245

Jukes, T. H., Cantor, C. R (1969). Evolution of protein molecules. *In* H. N. Munro (ed.): *Mammalian protein metabolism*, pp. 21-132, Academic, New York

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol 16:* 1099-106

Karlin, S. and Feldman, M. W. (1968). Further analysis of negative assortative mating. *Genetics 59:* 117-&

Kavaliers, M., Choleris, E., Agmo, A., Muglia, L. J., Ogawa, S., and Pfaff, D. W. (2005). Involvement of the oxytocin gene in the recognition and avoidance of parasitized males by female mice. *Animal Behaviour 70:* 693-702

Kavaliers, M., Choleris, E., Agmo, A., and Pfaff, D. W. (2004). Olfactory-mediated parasite recognition and avoidance: linking genes to behavior. *Hormones and Behavior 46:* 272-283

Kavaliers, M. and Colwell, D. D. (1993). Aversive Responses of Female Mice to the Odors of Parasitized Males - Neuromodulatory Mechanisms and Implications for Mate Choice. *Ethology 95:* 202-212

Kavaliers, M. and Colwell, D. D. (1995a). Discrimination by female mice between the odours of parasitized and non-parasitized males. *Proc Biol Sci 261:* 31-5

Kavaliers, M. and Colwell, D. D. (1995b). Odors of parasitized males induce aversive responses in female mice. *Animal Behaviour 50:* 1161-1169

Kavaliers, M., Colwell, D. D., Braun, W. J., and Choleris, E. (2003). Brief exposure to the odour of a parasitized male alters the subsequent mate odour responses of female mice. *Animal Behaviour 65:* 59-68

Kavaliers, M., Colwell, D. D., and Choleris, E. (1998). Parasitized female mice display reduced aversive responses to the odours of infected males. *Proc Biol Sci 265:* 1111-8

Kellenberger, C., Roussel, A., and Malissen, B. (2005). The H-2Kk MHC peptide-binding groove anchors the backbone of an octameric antigenic peptide in an unprecedented mode. *J Immunol 175:* 3819-25

Keller, L. F. (1998). Inbreeding and its fitness effects in an insular population of song sparrows (Melospiza melodia). *Evolution 52:* 240-250

Keller, M., Douhard, Q., Baum, M. J., and Bakker, J. (2006). Destruction of the main olfactory epithelium reduces female sexual behavior and olfactory investigation in female mice. *Chemical Senses 31:* 315-323

Kempenaers, B. (2007a). Mate choice and genetic quality: A review of the heterozygosity theory. *Advances in the Study of Behavior, Vol 37 37:* 189-278

Kempenaers, B. (2007b). Mate choice and genetic quality: A review of the heterozygosity theory. *Advances in the Study of Behavior, Vol 37*, pp. 189-278

Kempenaers, B., Adriaensen, F., vanNoordwijk, A. J., and Dhondt, A. A. (1996). Genetic similarity, inbreeding and hatching failure in blue

tits: Are unhatched eggs infertile? *Proceedings of the Royal Society of London Series B-Biological Sciences 263:* 179-185

Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S., and Jorde, L. B. (1998). Signatures of population expansion in microsatellite repeat data. *Genetics 148:* 1921-1930

Kimura, M. and Crow, J. F. (1963). On Maximum Avoidance of Inbreeding. *Genetical Research 4:* 399-&

Klein, J. (1980). Generation of diversity at MHC loci: Implications for T-cell receptor repertoires. *In* M. Fougereau, Dausset, J. (ed.): *Immunology '80 - Progress in Immunology IV*, pp. 239-253, Academic Press, London

Klein, J. (1986). Natural History of the Major Histocompatibility Complex. John Wiley & Sons, New York

Klein, J. and Figueroa, F. (1981). Polymorphism of the Mouse H-2 Loci. *Immunological Reviews 60:* 23-57

Klein, S. L., Gamble, H. R., and Nelson, R. J. (1999). Trichinella spiralis infection in voles alters female odor preference but not partner preference. *Behavioral Ecology and Sociobiology 45:* 323-329

Klein, S. L. and Nelson, R. J. (1999). Activation of the immune-endocrine system with lipopolysaccharide reduces affiliative behaviors in voles. *Behavioral Neuroscience 113:* 1042-1048

Kokko, H. and Ots, I. (2006). When not to avoid inbreeding. *Evolution 60:* 467-475

Koyama, K., Tamauchi, H., Tomita, M., Kitajima, T., and Ito, Y. (1999). B-cell activation in the mesenteric lymph nodes of resistant BALB/c mice infected with the murine nematode parasite Trichuris muris. *Parasitology Research 85:* 194-199

Lamprecht, J. (1986). Social-dominance and reproductive success in a goose flock (Anser-Indicus). *Behaviour 97:* 50-65

Landry, C., Garant, D., Duchesne, P., and Bernatchez, L. (2001). 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (Salmo salar). *Proceedings of the Royal Society of London Series B-Biological Sciences 268:* 1279-1285

Lanuza, E., Novejarque, A., Martinez-Ricos, J., Martinez-Hernandez, J., Agustin-Pavon, C., and Martinez-Garcia, F. (2008). Sexual pheromones and the evolution of the reward system of the brain: The chemosensory function of the amygdala. *Brain Research Bulletin 75:* 460-466

Lee, Y. H. and Kasper, L. H. (2004). Immune responses of different mouse strains after challenge with equivalent lethal doses of Toxoplasma gondii. *Parasite-Journal De La Societe Francaise De Parasitologie 11:* 89-97

Lefranc, M. P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., Ginestoux, C., Clement, O., Chaume, D., and

Lefranc, G. (2005). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res 33:* D593-7

Lehmann, L. and Perrin, N. (2003). Inbreeding avoidance through kin recognition: choosy females boost male dispersal. *Am Nat 162:* 638-52

Leonard, M. L. and Picman, J. (1988). Mate Choice by Marsh Wrens - the Influence of Male and Territory Quality. *Animal Behaviour 36:* 517-528

Levine, L., Barsel, G. E., and Diakow, C. A. (1966). Mating behaviour of two inbred strains of mice. *Animal Behaviour 14:* 1-&

Lieutenant-Gosselin, M. and Bernatchez, L. (2006). Local heterozygosity-fitness correlations with global positive effects on fitness in threespine stickleback. *Evolution 60:* 1658-1668

Lin, D. Y., Zhang, S. Z., Block, E., and Katz, L. C. (2005). Encoding social signals in the mouse main olfactory bulb. *Nature 434:* 470-477

Litvinova, E. A., Kudaeva, O. T., Mershieva, L. V., and Moshkin, M. P. (2005). High level of circulating testosterone abolishes decline in scent attractiveness in antigen-treated male mice. *Animal Behaviour 69:* 511-517

Lynch, M. (1991). The Genetic Interpretation of Inbreeding Depression and Outbreeding Depression. *Evolution 45:* 622-629

Manning, C. J., Wakeland, E. K., and Potts, W. K. (1992). Communal nesting patterns in mice implicate MHC genes in kin recognition. *Nature 360:* 581-3

Marie, A. D., Veggerby, C., Robertson, D. H. L., Gaskell, S. J., Hubbard, S. J., Martinsen, L., Hurst, J. L., and Beynon, R. J. (2001). Effect of polymorphisms on ligand binding by mouse major urinary proteins. *Protein Science 10:* 411-417

Marr, A. B., Keller, L. F., and Arcese, P. (2002). Heterosis and outbreeding depression in descendants of natural immigrants to an inbred population of song sparrows (Melospiza melodia). *Evolution 56:* 131-142

Marsden, H. M. and Bronson, F. H. (1964). Estrous Synchrony in Mice - Alteration by Exposure to Male Urine. *Science 144:* 1469-&

Martel, K. L. and Baum, M. J. (2007). Sexually dimorphic activation of the accessory, but not the main, olfactory bulb in mice by urinary volatiles. *European Journal of Neuroscience 26:* 463-475

Martinez-Ricos, J., Agustin-Pavon, C., Lanuza, E., and Martinez-Garcia, F. (2007). Intraspecific communication through chemical signals in female mice: Reinforcing properties of involatile male sexual pheromones. *Chemical Senses 32:* 139-148

Maruyama, T. and Nei, M. (1981). Genetic-Variability Maintained by Mutation and over-Dominant Selection in Finite Populations. *Genetics 98:* 441-459

Mateo, J. M. and Johnston, R. E. (2000). Kin recognition and the 'armpit effect': evidence of self-referent phenotype matching. *Proceedings of the Royal Society of London Series B-Biological Sciences 267:* 695-700

Mays, H. L. and Hill, G. E. (2004). Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology & Evolution 19:* 554-559

McClelland, E. E., Penn, D. J., and Potts, W. K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infect Immun 71:* 2079-86

Mcconnell, T. J., Talbot, W. S., Mcindoe, R. A., and Wakeland, E. K. (1988). The Origin of Mhc Class-II Gene Polymorphism within the Genus Mus. *Nature 332:* 651-654

McGregor, P. K. and Krebs, J. R. (1982). Mating and song sharing in the great tit. *Nature 297:* 60-61

Meagher, S., Penn, D. J., and Potts, W. K. (2000). Male-male competition magnifies inbreeding depression in wild house mice. *Proceedings of the National Academy of Sciences of the United States of America 97:* 3324-3329

Melvold, R. W., Wang, K. G., and Kohn, H. I. (1997). Histocompatibility gene mutation rates in the mouse: a 25-year review. *Immunogenetics 47:* 44-54

Merila, J., Sheldon, B. C., and Griffith, S. C. (2003). Heterotic effects on fitness in a wild bird population. *Annales Zoologici Fennici 40:* 269-280

Meyer-Lucht, Y. and Sommer, S. (2005). MHC diversity and the association to nematode parasitism in the yellow-necked mouse (Apodemus flavicollis). *Mol Ecol 14:* 2233-43

Meyer, D. and Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: a review. *Ann Hum Genet 65:* 1-26

Milinski, M. (2006). The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology Evolution and Systematics 37:* 159-186

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

Mitton, J. B. (1993). Enzyme Heterozygosity, Metabolism, and Developmental Stability. *Genetica 89:* 47-65

Moller, A. P. and Alatalo, R. V. (1999). Good-genes effects in sexual selection. *Proceedings of the Royal Society of London Series B-Biological Sciences 266:* 85-91

Moncho-Bogani, J., Lanuza, E., Hernandez, A., Novejarque, A., and Martinez-Garcia, F. (2002). Attractive properties of sexual pheromones in mice: Innate or learned? *Physiology & Behavior 77:* 167-176

Moncho-Bogani, J., Lanuza, E., Lorente, M. J., and Martinez-Garcia, F. (2004). Attraction to male pheromones and sexual behaviour show different regulatory mechanisms in female mice. *Physiology & Behavior 81:* 427-434

Moncho-Bogani, J., Martinez-Garcia, F., Novejarque, A., and Lanuza, E. (2005). Attraction to sexual pheromones and associated odorants in female mice involves activation of the reward system and basolateral amygdala. *European Journal of Neuroscience 21:* 2186-2198

Moore, J. and Ali, R. (1984). Are Dispersal and Inbreeding Avoidance Related. *Animal Behaviour 32:* 94-112

More, L. (2006). Mouse major urinary proteins trigger ovulation via the vomeronasal organ. *Chemical Senses 31:* 393-401

Nadeau, J. H., Wakeland, E. K., Gotze, D., and Klein, J. (1981). The Population-Genetics of the H-2 Polymorphism in European and North African Populations of the House Mouse (Mus-Musculus L). *Genetical Research 37:* 17-31

Nathenson, S. G., Geliebter, J., Pfaffenbach, G. M., and Zeff, R. A. (1986). Murine major histocompatibility complex class-I mutants: molecular analysis and structure-function implications. *Annu Rev Immunol 4:* 471-502

Neff, B. D. and Pitcher, T. E. (2005). Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology 14:* 19-38

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

Neufeld, E., Ritte, U., Figueroa, F., and Klein, J. (1986). Low H2 polymorphism in some Israeli wild mouse populations. *Immunogenetics 24:* 374-380

Nevison, C. M., Barnard, C. J., Beynon, R. J., and Hurst, J. L. (2000). The consequences of inbreeding for recognizing competitors. *Proceedings of the Royal Society of London Series B-Biological Sciences 267:* 687-694

Nevo, E. and Beiles, A. (1992). Selection for class II Mhc heterozygosity by parasites in subterranean mole rats. *Experientia 48:* 512-5

Ninomiya, K. and Brown, R. E. (1995). Removal of the preputial gland alters the individual odors of male MHC-congenic mice and the preferences of females. *Physiology & Behavior 58:* 191-194

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

Pankevich, D. E., Cherry, J. A., and Baum, M. J. (2006). Effect of vomeronasal organ removal from male mice on their preference for and neural fos responses to female urinary odors. *Behavioral Neuroscience 120:* 925-936

Parham, P. (2000). The Immune System. Garland Publishing, New York

Paterson, S. (1998). Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *Journal of Heredity 89:* 289-294

Paterson, S. (2005). No evidence for specificity between host and parasite genotypes in experimental Strongyloides ratti (Nematoda) infections. *Int J Parasitol*

Paterson, S. and Pemberton, J. M. (1997). No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proc R Soc Lond B Biol Sci 264:* 1813-9

Penn, D. and Potts, W. K. (1998a). Chemical signals and parasite-mediated sexual selection. *Trends in Ecology & Evolution 13:* 391-396

Penn, D. and Potts, W. K. (1998b). Untrained mice discriminate MHC-determined odors. *Physiol Behav 64:* 235-43

Penn, D. J. (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. *Ethology 108:* 1-21

Penn, D. J., Damjanovich, K., and Potts, W. K. (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A 99:* 11260-4

Penn, D. J. and Potts, W. K. (1999). The evolution of mating preferences and major histocompatibility complex genes. *American Naturalist 153:* 145-164

Perrin, N. and Mazalov, V. (2000). Local competition, inbreeding, and the evolution of sex-biased dispersal. *American Naturalist 155:* 116-127

Petrie, M. and Halliday, T. (1994). Experimental and Natural Changes in the Peacocks (Pave Cristatus) Train Can Affect Mating Success. *Behavioral Ecology and Sociobiology 35:* 213-217

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

Pitcher, T. E. and Neff, B. D. (2006). MHC class IIB alleles contribute to both additive and nonadditive genetic effects on survival in Chinook salmon. *Molecular Ecology 15:* 2357-2365

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

Posada, D. and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations.

*Proceedings of the National Academy of Sciences of the United States of America 98:* 13757-13762

Potts, W. K., Manning, C. J., and Wakeland, E. K. (1991). Mating Patterns in Seminatural Populations of Mice Influenced by Mhc Genotype. *Nature 352:* 619-621

Potts, W. K., Manning, C. J., and Wakeland, E. K. (1992). Sexual Selection and Mhc Genes - Reply. *Nature 356:* 294-294

Potts, W. K., Manning, C. J., and Wakeland, E. K. (1994). The Role of Infectious-Disease, Inbreeding and Mating Preferences in Maintaining Mhc Genetic Diversity - an Experimental Test. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 346:* 369-378

Potts, W. K. and Slev, P. R. (1995a). Pathogen-Based Models Favoring MHC Genetic Diversity. *Immunological Reviews 143:* 181-197

Potts, W. K. and Slev, P. R. (1995b). Pathogen-based models favoring MHC genetic diversity. *Immunol Rev 143:* 181-97

Potts, W. K. and Wakeland, E. K. (1990). Evolution of Diversity at the Major Histocompatibility Complex. *Trends in Ecology & Evolution 5:* 181-187

Potts, W. K. and Wakeland, E. K. (1993). Evolution of Mhc Genetic Diversity - a Tale of Incest, Pestilence and Sexual Preference. *Trends in Genetics 9:* 408-412

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics 155:* 945-959

Pullen, J. K., Horton, R. M., Cai, Z. L., and Pease, L. R. (1992). Structural diversity of the classical H-2 genes: K, D, and L. *J Immunol 148:* 953-67

Pusey, A., Williams, J., and Goodall, J. (1997). The influence of dominance rank on the reproductive success of female chimpanzees. *Science 277:* 828-831

Pusey, A. and Wolf, M. (1996). Inbreeding avoidance in animals. *Trends in Ecology & Evolution 11:* 201-206

Queller, D. C. and Goodnight, K. F. (1989). Estimating Relatedness Using Genetic-Markers. *Evolution 43:* 258-275

Ramm, S., Cheetham, S., Hurst, J. (2008). Encoding choosiness: female attraction requires prior physical contact with individual male scents in mice. *Proc. R. Soc. B In publication*

Ramsey, M., Vaughton, G., and Peakall, R. (2006). Inbreeding avoidance and the evolution of gender dimorphism in Wurmbea biglandulosa (Colchicaceae). *Evolution 60:* 529-537

Raymond, M. and Rousset, F. (1995). Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity 86:* 248-249

Reusch, T. B., Haberli, M. A., Aeschlimann, P. B., and Milinski, M. (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature 414:* 300-2

Reynolds, J. D. and Gross, M. R. (1992). Female Mate Preference Enhances Offspring Growth and Reproduction in a Fish, Poecilia-Reticulata. *Proceedings of the Royal Society of London Series B-Biological Sciences 250:* 57-62

Rich, T. J. and Hurst, J. L. (1998). Scent marks as reliable signals of the competitive ability of mates. *Anim Behav 56:* 727-735

Roberts, S. C. and Gosling, L. M. (2003). Genetic similarity and quality interact in mate choice decisions by female mice. *Nat Genet 35:* 103-6

Roberts, S. C., Little, A. C., Gosling, L. M., Perrett, D. I., Carter, V., Jones, B. C., Penton-Voak, I., and Petrie, M. (2005). MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior 26:* 213-226

Robertson, D. H. L., Beynon, R. J., and Evershed, R. P. (1993). Extraction, Characterization, and Binding Analysis of two Pheromonally Active Ligands Associated with Major Urinary Protein of House Mouse (Mus-Musculus). *Journal of Chemical Ecology 19:* 1405-1416

Robertson, D. H. L., Cox, K. A., Gaskell, S. J., Evershed, R. P., and Beynon, R. J. (1996). Molecular heterogeneity in the major urinary proteins of the house mouse Mus musculus. *Biochemical Journal 316:* 265-272

Robertson, D. H. L., Hurst, J. L., Searle, J. B., Gunduz, I., and Beynon, R. J. (2007). Characterization and comparison of major urinary proteins from the house mouse, Mus musculus domesticus, and the aboriginal mouse, Mus macedonicus. *Journal of Chemical Ecology 33:* 613-630

Rolland, C., MacDonald, D. W., De Fraipont, M., and Berdoy, M. (2003). Free female choice in house mice: Leaving best for last. *Behaviour 140:* 1371-1388

Saha, B. K. (1996). Typing of murine major histocompatibility complex with a microsatellite in the class II Eb gene. *J Immunol Methods 194:* 77-83

Sauermann, U., Nurnberg, P., Bercovitch, F. B., Berard, J. D., Trefilov, A., Widdig, A., Kessler, M., Schmidtke, J., and Krawczak, M. (2001). Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics 108:* 249-254

Sharrow, S. D., Vaughn, J. L., Zidek, L., Novotny, M. V., and Stone, M. J. (2002). Pheromone binding by polymorphic mouse major urinary proteins. *Protein Science 11:* 2247-2256

Sherborne, A. L., Thom, M. D., Paterson, S., Jury, F., Ollier, W. E., Stockley, P., Beynon, R. J., and Hurst, J. L. (2007). The Genetic Basis of Inbreeding Avoidance in House Mice. *Curr Biol*

Sherman, P. W., Reeve, H. K., Pfennig, D. W. (1997). Recognition systems. *In* J. R. Krebs, Davies, N. B. (ed.): *Behavioural ecology: an evolutionary approach*, pp. 69 - 96, Blackwell Scientific, Oxford, UK

Simmons, L. W. (2005). The evolution of polyandry: Sperm competition, sperm selection, and offspring viability. *Annual Review of Ecology Evolution and Systematics 36:* 125-146

Slade, R. W. and Mccallum, H. I. (1992). Overdominant Vs Frequency-Dependent Selection at Mhc Loci. *Genetics 132:* 861-862

Slev, P. R., Nelson, A. C., and Potts, W. K. (2006). Sensory neurons with MHC-like peptide binding properties: disease consequences. *Curr Opin Immunol 18:* 608-16

Snell, G. D. (1958a). Histocompatibility genes of the mouse. 1. Demonstration of weak histocompatibility differences by immunization and controlled tumor dosage. *Journal of the National Cancer Institute 20:* 787-824

Snell, G. D. (1958b). Histocompatibility genes of the mouse. 2. Production and analysis of isogenic resistant lines. *Journal of the National Cancer Institute 21:* 843-877

Soper, L. A., Stein, M. B., and Passmore, H. C. (1988). The E beta hot spot of recombination in wild-derived natural recombinant MHC haplotypes. Cross-over site mapping and the identification of a 1.0-kb E beta deletion in the p and w14 haplotypes. *J Immunol 140:* 984-90

Steinmetz, M. (1986). Polymorphism and recombinational hot-spots in the murine MHC. *Current Topics in Microbiology and Immunology 127:* 279-284

Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L., and Wiley, D. C. (1994). Crystal-Structure of the Human Class-Ii Mhc Protein Hla-Dr1 Complexed with an Influenza-Virus Peptide. *Nature 368:* 215-221

Stockley, P., Searle, J. B., Macdonald, D. W., and Jones, C. S. (1993). Female Multiple Mating-Behavior in the Common Shrew as a Strategy to Reduce Inbreeding. *Proceedings of the Royal Society of London Series B-Biological Sciences 254:* 173-179

Stopka, P., Janotova, K., and Heyrovsky, D. (2007). The advertisement role of major urinary proteins in mice. *Physiology & Behavior 91:* 667-670

Stopkova, R., Stopka, P., Janotova, K., and Jedelsky, P. L. (2007). Species-specific expression of major urinary proteins in the house mice (Mus musculus musculus and Mus musculus domesticus). *Journal of Chemical Ecology 33:* 861-869

Swofford, D. L. (1993). Paup - a Computer-Program for Phylogenetic Inference Using Maximum Parsimony. *Journal of General Physiology 102:* A9-A9

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics 123:* 585-595

Takahata, N. and Nei, M. (1990). Allelic Genealogy under Overdominant and Frequency-Dependent Selection and Polymorphism of Major Histocompatibility Complex Loci. *Genetics 124:* 967-978

Takahata, N., Satta, Y., and Klein, J. (1992). Polymorphism and Balancing Selection at Major Histocompatibility Complex Loci. *Genetics 130:* 925-938

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution 24:* 1596-1599

Tamura, K. and 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 10:* 512-526

Thom, M. D., Stockley, P., Jury, F., Ollier, W. E. R., Beynon, R. J., and Hurst, J. L. (2008). The direct assessment of genetic heterozygosity through scent in the mouse. *Current Biology 18:* 619-623

Thomas, L. (1975). *In* E. M. Neter, F (ed.): *Fourth International Congress of Immunology*, pp. 2, Karger, Basel

Tregenza, T. and Wedell, N. (2000). Genetic compatibility, mate choice and patterns of parentage: Invited review. *Molecular Ecology 9:* 1013-1027

Trivers, R. L. (1972). Parental investment and sexual selection. *In* B. Campbell (ed.): *Sexual selection and the descent of man, 1871-1971*, pp. 136-179, Aldine, Chicago, IL

Trowsdale, J. (1993). Genomic structure and function in the MHC. *Trends in Genetics 9:* 117-122

Van Oortmeerssen, G. A. (1971). Biological significance, genetics and evolutionary origin of variability in behaviour within and between inbred strains of mice (Mus musculus). *Behaviour 38:* 1-92

Wade, C. M., Kulbokas, E. J., Kirby, A. W., Zody, M. C., Mullikin, J. C., Lander, E. S., Lindblad-Toh, K., and Daly, M. J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature 420:* 574-578

Wagner, W. E. (1998). Measuring female mating preferences. *Animal Behaviour 55:* 1029-1042

Wakelin, D. (1967). Acquired immunity to Trichuris muris in the albino laboratory mouse. *Parasitology 57:* 515-24

Wakelin, D. (1970). Stimulation of Immunity and Induction of Unresponsiveness to Trichuris-Muris in Various Strains of Laboratory Mice. *Zeitschrift Fur Parasitenkunde 35:* 162-&

Wakelin, D. (1975). Genetic-Control of Immune-Responses to Parasites - Immunity to Trichuris-Muris in Inbred and Random-Bred Strains of Mice. *Parasitology 71:* 51-60

Wang, H. W., Wysocki, C. J., and Gold, G. H. (1993). Induction of olfactory receptor sensitivity in mice. *Science 260:* 998-1000

Watterson, G. A. (1978). The homozygosity test of neutrality. *Genetics 88:* 405-417

Watts, S., Vogel, J. M., Harriman, W. D., Itoh, T., Stauss, H. J., and Goodenow, R. S. (1987). DNA sequence analysis of the C3H H-2Kk and H-2Dk loci. Evolutionary relationships to H-2 genes from four other mouse strains. *J Immunol 139:* 3878-85

Wedekind, C., Seebeck, T., Bettens, F., and Paepke, A. J. (1995). Mhc-Dependent Mate Preferences in Humans. *Proceedings of the Royal Society of London Series B-Biological Sciences 260:* 245-249

Wegner, K. M., Kalbe, M., Schaschl, H., and Reusch, T. B. (2004). Parasites and individual major histocompatibility complex diversity--an optimal choice? *Microbes Infect 6:* 1110-6

Wegner, K. M., Reusch, T. B., and Kalbe, M. (2003). Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *J Evol Biol 16:* 224-32

Westerdahl, H. (2004). No evidence of an MHC-based female mating preference in great reed warblers. *Molecular Ecology 13:* 2465-2470

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Ostell, J., Pruitt, K. D., Schuler, G. D., Shumway, M., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusova, T. A., Wagner, L., and Yaschenko, E. (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res 36:* D13-21

Widera, G. and Flavell, R. A. (1984). The Nucleotide-Sequence of the Murine I-E-Beta-B Immune-Response Gene - Evidence for Gene Conversion Events in Class-Ii Genes of the Major Histocompatibility Complex. *Embo Journal 3:* 1221-1225

Williams, J. R. and Lenington, S. (1993). Factors modulating preferences of female house mice for male differing in t-complex genotype – role of t-complex genotype, genetic background, and estrous condition of females. *Behavior Genetics 23:* 51-58

Willis, C. and Poulin, R. (2000). Preference of female rats for the odours of non-parasitised males: the smell of good genes? *Folia Parasitologica 47:* 6-10

Worley, K., Carey, J., Veitch, A., and Coltman, D. W. (2006). Detecting the signature of selection on immune genes in highly structured

populations of wild sheep (Ovis dalli). *Molecular Ecology 15:* 623-637

Wyatt, T. D. (2003). Pheromones and animal behaviour: communication by smell and taste. Cambridge University Press

Yamaguchi, M., Yamazaki, K., Beauchamp, G. K., Bard, J., Thomas, L., and Boyse, E. A. (1981). Distinctive urinary odors governed by the major histocompatibility locus of the mouse. *Proc Natl Acad Sci U S A 78:* 5817-20

Yamazaki, K., Beauchamp, G. K., Curran, M., Bard, J., and Boyse, E. A. (2000). Parent-progeny recognition as a function of MHC odortype identity. *Proc Natl Acad Sci U S A 97:* 10500-2

Yamazaki, K., Beauchamp, G. K., Egorov, I. K., Bard, J., Thomas, L., and Boyse, E. A. (1983). Sensory distinction between H-2B and H-2BM1 mutant mice. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 80:* 5685-5688

Yamazaki, K., Beauchamp, G. K., Kupniewski, D., Bard, J., Thomas, L., and Boyse, E. A. (1988). Familial imprinting determines H-2 selective mating preferences. *Science 240:* 1331-1332

Yamazaki, K., Beauchamp, G. K., Shen, F. W., Bard, J., and Boyse, E. A. (1994). Discrimination of odortypes determined by the major histocompatibility complex among outbred mice. *Proc Natl Acad Sci U S A 91:* 3735-8

Yamazaki, K., Boyse, E. A., Bard, J., Curran, M., Kim, D., Ross, S. R., and Beauchamp, G. K. (2002). Presence of mouse mammary tumor virus specifically alters the body odor of mice. *Proceedings of the National Academy of Sciences of the United States of America 99:* 5612-5615

Yamazaki, K., Boyse, E. A., Mike, V., Thaler, H. T., Mathieson, B. J., Abbott, J., Boyse, J., Zayas, Z. A., and Thomas, L. (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. *J Exp Med 144:* 1324-35

Yamazaki, K., Singer, A., and Beauchamp, G. K. (1998). Origin, functions and chemistry of H-2 regulated odorants. *Genetica 104:* 235-40

Yamazaki, K., Yamaguchi, M., Andrews, P. W., Peake, B., and Boyse, E. A. (1978). Mating Preferences of F2 Segregants of Crosses between Mhc-Congenic Mouse Strains. *Immunogenetics 6:* 253-259

Yamazaki, K., Yamaguchi, M., Baranoski, L., Bard, J., Boyse, E. A., and Thomas, L. (1979). Recognition among mice. Evidence from the use of a Y-maze differentially scented by congenic mice of different major histocompatibility types. *Journal of Experimental Medicine 150:* 755-760

Yoshikage, M., Toshiaki, I., Seiichi, K., Nobuo, K., and Nishimura, M. (2007). Sex steroids modulate the signals from volatile female

odors in the accessory olfactory bulb of male mice. *Neuroscience Letters 413:* 11-15

Yoshiki, A. and Moriwaki, K. (2006). Mouse phenome research: implications of genetic background. *Ilar Journal 47:* 94-102

Zala, S. M., Potts, W. K., and Penn, D. J. (2004). Scent-marking displays provide honest signals of health and infection. *Behavioral Ecology 15:* 338-344

Zeh, J. A. and Zeh, D. W. (1996). The evolution of polyandry I: Intragenomic conflict and genetic incompatibility. *Proceedings of the Royal Society of London Series B-Biological Sciences 263:* 1711-1717

Zeh, J. A. and Zeh, D. W. (1997). The evolution of polyandry 2. Post-copulatory defences against genetic incompatibility. *Proceedings of the Royal Society of London Series B-Biological Sciences 264:* 69-75

Zelano, B. and Edwards, S. V. (2002). An Mhc component to kin recognition and mate choice in birds: Predictions, progress, and prospects. *American Naturalist 160:* S225-S237

Zhang, J. Z., Rosenberg, H. F., and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences of the United States of America 95:* 3708-3713

Zhu, Y., Rudensky, A. Y., Corper, A. L., Teyton, L., and Wilson, I. A. (2003). Crystal structure of MHC class II I-Ab in complex with a human CLIP peptide: prediction of an I-Ab peptide-binding motif. *J Mol Biol 326:* 1157-74

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

Zufall, F. and Leinders-Zufall, T. (2007). Mammalian pheromone sensing. *Current Opinion in Neurobiology 17:* 483-489

Zuk, M. (1992). The Role of Parasites in Sexual Selection - Current Evidence and Future-Directions. *Advances in the Study of Behavior 21:* 39-68

# Appendix 1: Tables and Figures

Appendix Table 1. Microsatellite markers used for MHC genotyping in three studies (inf = infection study (Chapter 2), enc = enclosure study (Chapter 3) and win = Winsford population study (Chapter 4)).

| MHC Marker | Location | Forward Primer | Reverse Primer | Size | Repeat | Label | Study | F1* |
|---|---|---|---|---|---|---|---|---|
| **D17Mit192** | 17 B1 32.3Mb | CAGACAACACCTCTGAACTGG | GAGGCCCCTATGTGTAGCAG | 298 | GT | FAM | win | |
| **D17Mit230** | 17 B2 32.6Mb | TGACATAAACCTCTGGCTTCC | CCAGCCCATCTAAAGTCATTTC | 288 | CA | NED | inf/enc/win | A,B,C,D |
| **D17Mit21** | 17 B2 33.8Mb | ACCTCTGGTCTGTGGGAGTG | TGGACCCAACTTCTTAAAACC | 175 | CA | HEX | win | |
| **D17Mit22** | 17 B2 33.9Mb | GCATTAGATAGAGAGTAGATGGGTTG | ATGGATGGCGAGAATGAGAC | 216 | GT | HEX | enc | A,B,D |
| **D17Mit231** | 17 B2 34.1Mb | GCCTCAGCAAGACCCTAAAC | ACTCCTCCTTTTCCCTCTCC | 285 | GT | HEX | enc/win | |
| **D17Mit13** | 17 B2 34.6Mb | TGCAGGCAAGATCCAAGAAG | GAAAGAGGGTGTCGATGCTC | 239 | GT | HEX | enc/win | |
| **D17Nds3** | 17 B3 34.8Mb | CAGCCTTAATGGGTCTGGTC | ACAGAGGGGAAGAGGAAAGC | 222 | CT | NED | enc/win | B |
| **D17Mit47** | 17 B3 35.8Mb | CTGAGACCAGTGCAGTGGAA | TTTTTCAATATGTGAGCATGTGC | 238 | CA | FAM | enc | C |
| **D17Mit24** | 17 B3 37.0Mb | ACCTCTCACCTCTCTCTGTG | GCAAGTTTAGGGATCTTTTCTCC | 131 | CA | FAM | inf/enc | A,B,C,D |
| **D17Mit20** | 17 E1.1 56.9Mb | AGAACAGGACACCGGACATC | TCATAAGTAGGCACACCAATGC | 165 | CA | NED | win | |

* In the enclosure study, after identifying MHC haplotypes in founders and their parents using all listed markers, only a subset were necessary to identify haplotypes in F1 offspring for each of populations A-D.

Appendix Table 2. Microsatellite markers used for MUP genotyping in the enclosure study (enc) only (Chapter 3).

| MUP Marker | Location | Forward Primer | Reverse Primer | Size | Repeat | Label | Study | F1* |
|---|---|---|---|---|---|---|---|---|
| **D4Nds6** | 4 B3 52.8 | CGGGGAAGGTTGTTTGTTTG | AGGCCAGCAATGTAGAAAGG | 240 | GT | NED | enc | |
| **D4Mit139** | 4 B3 55.2 | TCAAACTGGGAAGAGCCAAG | GCCGTAGAAGAGAAGTAATTTTTCC | 149 | GT | HEX | enc | D |
| **D4Mit241** | 4 C1 55.7 | TTTCCAGTGTTGTCCAGAGC | AAGGCAAATCACTAGGTGCTG | 219 | CA | HEX | enc | D |
| **D4Mit288** | 4 C1 56.8 | ACATTCAGCAAAGACTGAGCAC | TGCCATTTGTTATAGACCATGC | 168 | CA | NED | enc | A,B |
| **D4Mit164** | 4 C1 59.5 | AACACATATATACCAAGGCAGCAC | ATTTCCACCCTGTCCACTCC | 142 | CA | FAM | enc | A,B,C |
| **D4Mit243** | 4 C1 59.6 | AGCCCTACTGATTGCTCTCC | TGGAAAGTTGAAAACCACTGC | 168 | CA | FAM | enc | A,B,C,D |
| **D4Mit217** | 4 C1 59.7 | ACTCAATTAGGTTGTTCAGATAGCC | GGCACTTGCTGCCACATC | 246 | GT | HEX | enc | |
| **D4Mit17** | 4 C1 62.8 | GCCAACCTCTGTGCTTCC | CCTCTGACATCCACACACATC | 138 | GT | FAM | enc | A,B,C,D |

* After identifying MUP haplotypes in founders and their parents using all listed markers, only a subset were necessary to identify haplotypes in F1 offspring for each of populations A-D.

Appendix Table 3. Microsatellite markers used for parentage assignment in the enclosure study (3.3). The sub-set of markers used in heterozygosity study (3.4) are indicated.

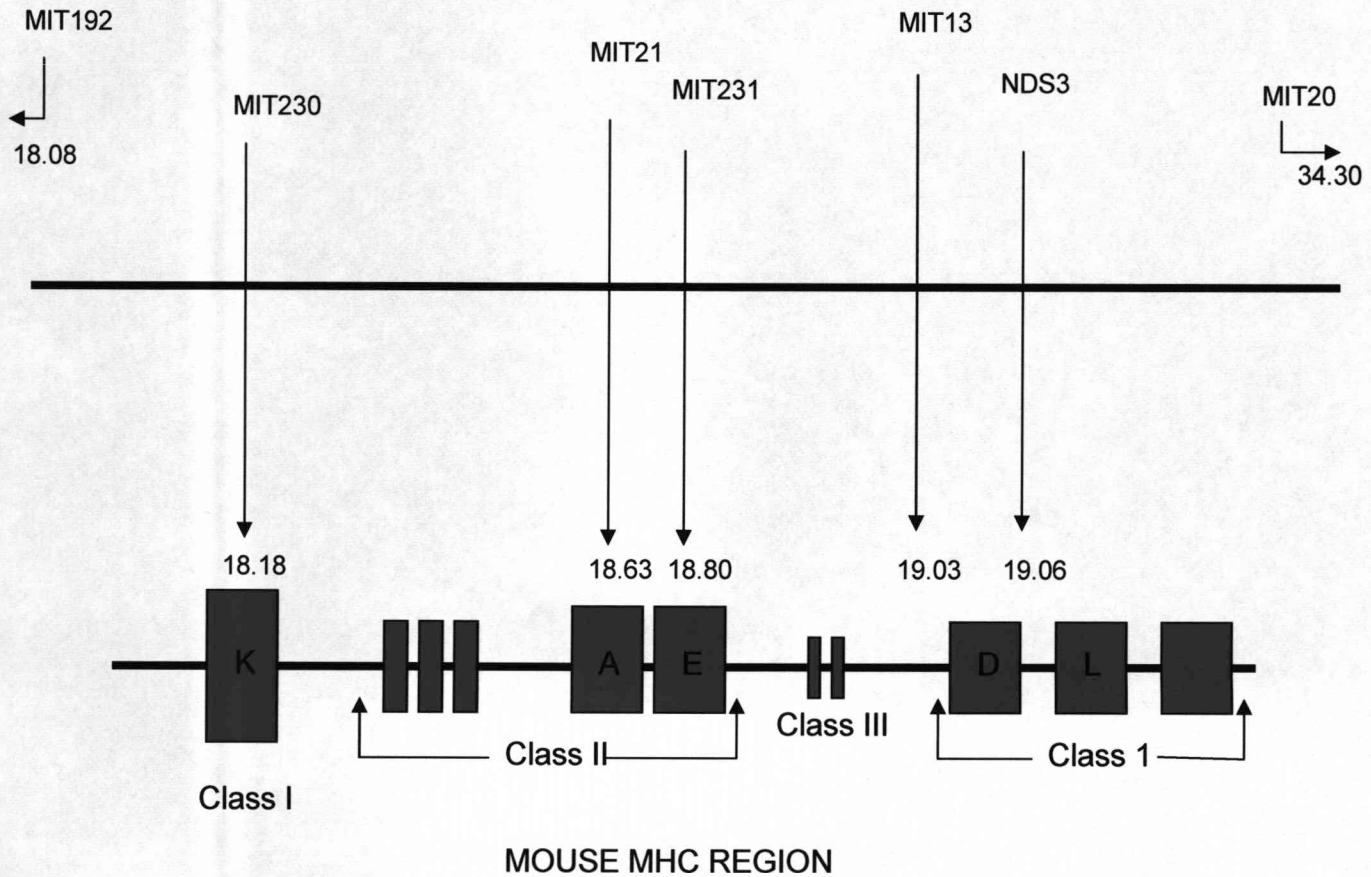
| arker | Position (cM) | Location (Mb) | Forward Primer | Reverse Primer | Size | Label | Study 3.4? | Balb/c Size | C75BL/ Size |
|---|---|---|---|---|---|---|---|---|---|
| 1Mit58 | 1 8.3 | 9.7 | GGACTGGCAATCCTCTTGTC | GCACGTTAGAGAGTGGGCTC | 254 | VIC | ✓ | 274 | 260 |
| 1Mit155 | 1 112.0 | 196.1 | ATGCATGCATGCACACGT | ACCGTGAAATGTTCACCCAT | 252 | NED | | | |
| 2Mit405 | 2 68.9 | 149.7 | TGATTATATCTTGGAATACACGTGTG | CTGTGTAGCAAAACAGTTTATGGC | 84 | 6-FAM | ✓ | 92 | 86 |
| 3Mit1 | 3 11.2 | 0.3 | TGTGCACAGGGGTACATACA | TCATTTTCTTCCTCCCCCTC | 143 | VIC | ✓ | 122 | 120 |
| 3Mit163 | 3 87.6 | 157.3 | TGGATACATACATATACATGGAAATGC | TTTCTCCAGACCCATGAACC | 143 | 6-FAM | | | |
| 4Mit171 | 4 6.3 | 22.6 | CAGGTGTAATAATGGTTTTTTGACC | CATATTAAATAAACACAGCAGCACG | 318 | PET | | | |
| 4Mit310 | 4 71.0 | 147.5 | TCTCCACGTGTGTGCCTTAG | TGAAAGCACTCTGCAGACTCA | 117 | PET | ✓ | 127 | 117 |
| 5Mit25 | 5 61.0 | 111.8 | AACACACCTCCATACTGGTCG | GGCTAACTGAAATTGTTTTGTGC | 234 | NED | ✓ | 228 | 234 |
| 6Mit105 | 6 45.5 | 108.6 | CTGTCTCCACTACTTCTATTCCTGG | CAAAAGCCTTATATATTACACCTCACC | 237 | 6-FAM | | | |
| 7Mit253 | 7 55.0 | 115.3 | TGTGGGTGCAACCAAATG | TTTGGTGATATAGATACTAGGTGTGTG | 89 | NED | | | |
| 8Mit29 | 8 33.0 | 70.7 | CCCTAGTGTATACATAGAGGGGTG | TCTTTGTGTTGTGATGTGTTGTAA | 109 | VIC | | | |
| 9Mit12 | 9 55.0 | 99.4 | ATTCAAGGGGCAGTACACAT | TGGTCCTGGTAAAACTGCCT | 96 | 6-FAM | | | |
| 10Mit80 | 10 4.0 | 11.5 | CAAAAAAAACCCTGATTCTACCA | GTGTGCATATGGCAGTAACTTTG | 154 | NED | ✓ | 144 | 158 |
| 10Mit98 | 10 59.0 | 102.0 | TCAGGCATCTAGTGAGATGATCC | CCCATAGATGCAGGGGTG | 150 | VIC | ✓ | 144 | 146 |
| 11Mit63 | 11 2.0 | 17.1 | GCCCACAACTTTGTGTCCTT | TTGACCATGCTCCTCATCAG | 139 | PET | ✓ | 146 | 144 |
| 11Mit300 | 11 68.0 | 111.3 | TTTGGCTGTGATAAAACAAAACA | TTGAGTTTTGATTTTGTATGTGGG | 145 | 6-FAM | ✓ | 138 | 150 |

| arker | Position (cM) | Location (Mb) | Forward Primer | Reverse Primer | Size | Label | Study 3.4? | Balb/c Size | C75BL/ Size |
|---|---|---|---|---|---|---|---|---|---|
| 12Mit4 | 12 34.0 | 79.9 | ACATCCCCAGCTCTTGTTTG | AAACCAAACCAAAGAAGCTTAGG | 201 | PET | ✓ | 194 | 200 |
| 13Mit77 | 13 73.0 | 118.0 | TCTTTGAAGTCCCTTTCAAAGC | ATAGCACTGCACTCATGCTCA | 275 | VIC | | | |
| 14Mit212 | 14 13.5 | 36.2 | AACATGTGCACTGGAACAATG | TCATTTATCAATTTACTTTGGTGAGG | 102 | PET | | | |
| 15Mit161 | 15 69.2 | 97.9 | TCTGTTTTGTTTGTTCGTTTGC | TAAAATCTCCCTGTATACAAGTCTGTG | 99 | NED | ✓ | 102 | 108 |
| 16Mit71 | 16 70.5 | 98.0 | TAGAAAATCTTCAAATAGGATCTGTTC | GAGCATTTCCCTTTTACCTGG | 154 | PET | | | |
| 17Mit94 | 17 45.9 | 75.8 | TGGAGAGGCATCCAACTCTC | TTCCTCTTAGTCCACCTTTTGC | 146 | NED | | | |
| 18Mit33 | 18 44.0 | 70.0 | GCATGTCGTATCCATAAACATACG | ATGCGGGCTTGACTTCTG | 140 | VIC | | | |
| 19Mit70 | 19 51.0 | 50.8 | AAAATATCAGGGCATGGTGC | GGGTTATTAGGAAAATTTATGTTGTG | 195 | 6-FAM | ✓ | 225 | 185 |

Appendix Table 4. Summary of founders and F1 offspring in each population of the enclosure study (3.3).

| | Population A | Population B | Population C | Population D | Total |
|---|---|---|---|---|---|
| Founders | 6 female, 7 male (3 litters) | 12 female, 18 male (5 litters) | 7 female, 12 male (3 litters) | 8 female, 11 male (3 litters) | 33 female, 48 male (14 litters) |
| MHC haplotypes | 6 | 9* | 5 | 6* | 16 |
| MUP haplotypes | 7* | 9 | 7 | 7* | 20 |
| Number of Female:Male Dyads | | | | | |
| Full sib | 18 | 36 | 14 | 28 | 96 |
| Half sib | 24 | 180 | 70 | 60 | 334 |
| MHC Haplotype Sharing between Founders (% Full-Sib Dyads/% Half-Sib Dyads) | | | | | |
| Both | 27.8/0 | 38.9/0.6 | 57.1/4.3 | 25.0/6.7 | 35.4/2.4 |
| One | 50.0/62.5 | 41.7/59.4 | 42.9/95.7 | 53.6/45.0 | 46.9/64.7 |
| None | 22.2/37.5 | 19.4/40.0 | 0/0 | 21.4/48.3 | 17.7/32.9 |
| MUP Haplotype Sharing between Founders (% Full-Sib Dyads/% Half-Sib Dyads) | | | | | |
| Both | 27.8/0 | 22.2/3.3 | 35.7/2.9 | 42.9/13.3 | 31.3/4.8 |
| One | 44.4/54.2 | 52.8/59.4 | 50.0/48.6 | 32.1/46.7 | 44.8/54.5 |
| None | 27.8/45.8 | 25.0/37.2 | 14.3/48.6 | 25.0/40.0 | 24.0/40.7 |
| F1 offspring genotyped | 99 | 192 | 101 | 91 | 483 |
| Number of Matings Observed (Expected) According to MHC Haplotypes Shared | | | | | |
| Both | 2 (3.4) | 9 (5.7) | 1 (5.3) | 5 (4.6) | 17 (19.1) |
| One | 20 (16.6) | 50 (45.1) | 41 (36.7) | 19 (19.0) | 130 (117.3) |
| None | 8 (10.0) | 22 (30.2) | 0 (0) | 16 (16.4) | 46 (56.5) |
| Number of Matings Observed (Expected) According to MUP Haplotypes Shared | | | | | |
| Both | 0 (2.9) | 3 (5.9) | 0 (3.2) | 3 (8.8) | 6 (20.7) |
| One | 22 (15.4) | 51 (48.1) | 19 (20.7) | 22 (16.4) | 114 (100.5) |
| None | 8 (11.7) | 27 (27.1) | 23 (18.2) | 15 (14.8) | 73 (71.8) |

"*" Indicates that founders inherited a crossover haplotype, combining alleles from two haplotypes. For analysis, this was regarded as a match for either of the original haplotypes.




Appendix Figure 1. Location of microsatellite markers used for typing at the MHC in relation to genes. Positions given are in cM.

# Appendix 2: Sherborne et al. (2007)

Current Biology 17, 2061–2066, December 4, 2007 ©2007 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2007.10.041

Report

# The Genetic Basis of Inbreeding Avoidance in House Mice

Amy L. Sherborne,[1] Michael D. Thom,[2] Steve Paterson,[1] Francine Jury,[3] William E.R. Ollier,[3] Paula Stockley,[2] Robert J. Beynon,[4] and Jane L. Hurst[2,*]
[1]Population and Evolutionary Biology Group
School of Biological Sciences
University of Liverpool
Liverpool L69 7ZB
United Kingdom
[2]Mammalian Behaviour and Evolution Group
Department of Veterinary Preclinical Science
University of Liverpool
Leahurst
Neston CH64 7TE
United Kingdom
[3]Centre for Integrated Genomic Medical Research
University of Manchester
Manchester M13 9PT
United Kingdom
[4]Proteomics and Functional Genomics Group
Department of Veterinary Preclinical Science
University of Liverpool
Crown Street and Brownlow Hill
Liverpool L69 7ZJ
United Kingdom

## Summary

**Animals might be able to use highly polymorphic genetic markers to recognize very close relatives and avoid inbreeding [1, 2]. The major histocompatibility complex (MHC) is thought to provide such a marker [1, 3–6] because it influences individual scent in a broad range of vertebrates [6–10]. However, direct evidence is very limited [1, 6, 10, 11]. In house mice (*Mus musculus domesticus*), the major urinary protein (MUP) gene cluster provides another highly polymorphic scent signal of genetic identity [8, 12–15] that could underlie kin recognition. We demonstrate that wild mice breeding freely in seminatural enclosures show no avoidance of mates with the same MHC genotype when genome-wide similarity is controlled. Instead, inbreeding avoidance is fully explained by a strong deficit in successful matings between mice sharing both MUP haplotypes. Single haplotype sharing is not a good guide to the identification of full sibs, and there was no evidence of behavioral imprinting on maternal MHC or MUP haplotypes. This study, the first to examine wild animals with normal variation in MHC, MUP, and genetic background, demonstrates that mice use self-referent matching of a species-specific [16, 17] polymorphic signal to avoid inbreeding. Recognition of close kin as unsuitable mates might be more variable across species than a generic vertebrate-wide ability to avoid inbreeding based on MHC.**

*Correspondence: jane.hurst@liv.ac.uk

## Results

Our experimental design met a set of stringent requirements to establish whether mice use major histocompatibility complex (MHC) and/or major urinary protein (MUP) to avoid inbreeding. First, genetic recognition and mate preferences need to be demonstrated against normally variable genetic and environmental backgrounds typical of natural populations. We therefore used wild house mice (*Mus musculus domesticus*) rather than genetically homogeneous laboratory mice that are hybrids of three *Mus* subspecies derived from an extremely small pool of founders [18, 19] and have been subject to strong artificial selection for ease of breeding in the laboratory [4, 20]. Second, previous studies examining kin recognition and inbreeding avoidance in mice have used only MHC types derived from laboratory strains (e.g., [3–5, 21, 22]). The role of MUPs in this context has not been examined previously, probably because of the lack of variation in MUP patterns within and between laboratory strains ([23], J.L.H. and R.J.B., unpublished data, and S.A. Cheetham, personal communication). This is in dramatic contrast to wild house mice, in which individual variation in the pattern of urinary MUPs is the main basis for individual recognition through scent [12, 13, 15, 24]. To properly reflect natural variation at these highly polymorphic gene complexes, we used founder mice that carried a large range of MHC and MUP haplotypes derived directly from wild animals. Third, in natural populations, animals that share the same genotype at a highly polymorphic marker such as MHC or MUP will be closely related and share many other alleles across the genome. In order to test whether inbreeding avoidance is driven by the sharing of MHC and/or MUP haplotypes (inherited independently) on naturally variable genetic backgrounds, we controlled for background relatedness in our analysis and design and thus the sharing of other alleles across the genome. Finally, to ensure natural behavior and mate choice, we allowed mice to breed freely in very large seminatural enclosures without intervention or disturbance.

In house mouse populations, dominant males defend small territories in which they monopolize mating opportunities [25], and extraterritorial matings by females with neighbor territory owners are frequent [4, 26]. Animals that remain in the same local population are thus likely to encounter unfamiliar half sibs that share the same father but different mothers in addition to full sibs that share both father and mother. To model this, we released four replicate populations of half-sib and full-sib outbred house mice (each derived from one father and several unrelated mothers with full sibs reared only by their mothers; see the Supplemental Experimental Procedures in the Supplemental Data available online) into four very large seminatural enclosures (each 250 $m^2$

Table 1. Multinomial Logistic Models for Frequency of Mating

| | β | Log Likelihood | Likelihood Ratio (LR) Statistic[a] | df | P[b] |
|---|---|---|---|---|---|
| **Model 1: Full-Sib Avoidance** | | | | | |
| full sib | −0.36 | −491.07 | 3.42 | 1 | 0.08 (one tailed) |
| **Model 2: MHC Sharing** | | | | | |
| one MHC haplotype | 0.36 | −490.63 | 4.31 | 2 | 0.27 |
| both MHC haplotypes | 0.12 | | | | |
| **Model 3: MUP Sharing** | | | | | |
| one MUP haplotype | 0.15 | −483.60 | 18.36 | 2 | 0.005 |
| both MUP haplotypes | −1.36 | | | | |
| **Model 4: Full MUP Sharing** | | | | | |
| both MUP haplotypes | −1.44 | −484.04 | 17.48 | 1 | 0.002 |
| **Model 5: Relatedness and Full MUP Sharing** | | | | | |
| full sib | −0.12 | −483.86 | 17.84 | 2 | 0.65[c] |
| both MUP haplotypes | −1.38 | | | | |
| **Model 6: Maternal MHC Imprinting and Full MUP Sharing** | | | | | |
| one MHC haplotype match between male and female's mother | −0.33 | −482.61 | 20.34 | 2 | 0.18[c] |
| both MUP haplotypes | −1.31 | | | | |
| **Model 7: Maternal MUP Imprinting and Full MUP Sharing** | | | | | |
| one MUP haplotype match between male and female's mother | −0.11 | −483.76 | 18.04 | 2 | 0.54[c] |
| both MUP haplotypes | −1.35 | | | | |

Summary of observed and expected frequencies for each category shown in Table S2.
[a] Compared to null model (log likelihood = −492.78).
[b] Probabilities calculated by random permutation of data (n = 10,000).
[c] Comparison to model 4.

with abundant cover and food). Tissue samples for genotyping were obtained from all F1 founders (33 female, 48 male) prior to release and from their parents, allowing us to establish the separate MHC and MUP haplotypes of each heterozygous founder. The combinations of males and females available as potential mating partners covered the full range of possible MHC and MUP haplotype sharing (Table S1). After 15 weeks, when founding females could have reared up to three litters to independence, animals were captured from each population and F2 offspring of the original founders (n = 483) were genotyped with 40 microsatellite markers so that parentage and MHC and MUP haplotypes could be independently established. From these data, we determined for each F1 founder female the minimum number of successful matings with each available male (based on recovered independent offspring) and the number of offspring captured from each female-male pair.

Consistent with inbreeding avoidance, we found a deficit in the overall frequency of successful matings between full sibs (Table S2), although the effect was too weak to differ significantly from random mating per female (Table 1: multinomial logistic model 1, p = 0.08). We found no evidence for disassortative mating on the basis of MHC haplotype sharing (Table 1: model 2), i.e., there was no evidence for fewer successful matings between mice that shared either one or both MHC haplotypes relative to those that shared no haplotypes (Table S2; Figure S1). By contrast, MUP sharing had a strong and highly significant effect on the likelihood of successful mating (Table 1: model 3, p = 0.005; Figure S1). Specifically, there was no deficit when only one MUP haplotype was shared, but there were many fewer matings between mice that shared both MUP haplotypes (complete match) than expected under random mating conditions (Table 1: model 4, $p < 0.002$). The same trend was evident across all four populations (Table S1). We confirmed that this effect was due to MUP sharing between males and females rather than inherent differences in male quality because there was no difference in the overall mating success of males represented in full MUP sharing dyads (mean ± standard error of the mean [SEM] matings per male with any female: 4.4 ± 0.6, n = 23) compared to the success of other males (3.8 ± 0.5 matings per male, n = 25; $t_{46} = -0.70$, p = 0.49). The strong deficit in successful matings between mice of the same MUP genotype also accounts for the weak deficit in matings between full sibs (Table 1: model 5 versus model 4, p = 0.6) because full sibs were much more likely than half sibs to share both MUP haplotypes (Table S1).

It has been hypothesized that familial (behavioral) imprinting on maternal haplotypes would allow animals to avoid inbreeding with a greater proportion of kin than the use of self-haplotype sharing alone [21]. Behavioral imprinting on maternal haplotypes (MHC, MUP, or other genetic loci) should result in a general avoidance of full sibs compared to paternal half sibs because all full sibs carry a maternal haplotype at each potential genetic marker. However, the inclusion of overall relatedness (full versus paternal half sib) as well as MUP genotype sharing in the model provides no better explanation of the likelihood of mating than does MUP genotype alone

Table 2. Multinomial Logistic Models for Number of Offspring per Mating

| | β | Log Likelihood | LR Statistic[a] | df | P[b] |
|---|---|---|---|---|---|
| Model 1: Full-Sib Avoidance | | | | | |
| full sib | −0.44 | −497.90 | 7.69 | 1 | 0.02[c] (one tailed) |
| Model 2: MHC Sharing | | | | | |
| one MHC haplotype | −0.12 | −499.24 | 5.02 | 2 | 0.25 |
| both MHC haplotypes | −0.49 | | | | |
| Model 3: MUP Sharing | | | | | |
| one MUP haplotype | 0.11 | −500.59 | 2.33 | 2 | 0.52 |
| both MUP haplotypes | −0.34 | | | | |

Summary of observed and expected frequencies for each category shown in Table S2.
[a] Compared to null model (log likelihood = −501.75).
[b] Probabilities calculated by random permutation of data (n = 10,000).
[c] Model 1 would also be significant at $p < 0.05$ in a two-tailed test.

(Table 1: model 5 versus model 4, $p = 0.6$). Specific analysis of female behavioral imprinting on maternal MHC haplotypes (Table 1: model 6) or maternal MUP haplotypes (Table 1: model 7) also failed to significantly improve explanation of the likelihood of mating above the direct sharing of MUP genotype alone (see the Supplemental Experimental Procedures for further discussion).

There were fewer offspring per successful mating with full sibs than random expectation (Table 2: model 1, $p = 0.02$), consistent with the lower viability of offspring due to inbreeding depression [2, 27] or with postcopulatory sperm or embryo selection by females [28]. However, there was no evidence that MHC or MUP sharing between parents per se reduced the viability of offspring from successful matings. Thus, it is unlikely that the strong deficit in successful matings between mice of the same MUP genotype (above) was due to lower viability of these offspring. The small deficit in offspring from matings between animals sharing both MHC haplotypes or both MUP haplotypes (Table S2) was nonsignificant (Table 2) and was due to the fact that most dyads in these two categories were full sibs. Full sibs had significantly fewer offspring per mating regardless of MHC or MUP sharing. Similarly, the MHC and MUP genotypes of all F2 offspring followed the expected pattern of Mendelian inheritance (Table S2), indicating that MHC and MUP homozygotes and heterozygotes had equal viability with no evidence for differential postcopulatory selection.

Why should mice avoid mates that share both MUP haplotypes but not those that share one haplotype? In our experiment, the sharing of both MUP haplotypes was a good predictor of whether a potential mate was a full sib or not: 31% of full-sib dyads shared both MUP haplotypes versus only 5% of half-sib dyads (Table S1). However, because 45% of full sibs and 54% of half sibs shared one MUP haplotype, this was a poor guide to whether a potential mate was a full sib. A simulation of populations containing different numbers of haplotypes shows this to be a general rule and not just a consequence of the mix of founders used in the current experiment (Figure S2). Even when there are many haplotypes in the population, the sharing of one or zero haplotypes provides very little information on whether a potential mate is likely to be a full sib or a nonrelative, whereas the sharing of both haplotypes considerably increases the likelihood that animals are full sibs. Mice thus avoid mating when shared MUP type reliably indicates very close relatedness.

## Discussion

The use of MUP alone was sufficient to explain inbreeding avoidance in this study. Although the deficit in mating between mice of the same MUP type was very strong, there was no evidence that animals used genetic markers other than MUP to improve their level of inbreeding avoidance. Further, there was no evidence to support the previously untested hypothesis that animals could increase the range of relatives avoided by imprinting on the separate haplotypes carried by their mother [21]. Such a strategy would also result in the rejection of any unrelated animals that happen to share one maternal haplotype. Notably, animals should only avoid related partners when there is a high inbreeding load [29, 30], which is only likely to occur between very close relatives within outbred populations. As we have shown, the sharing of both haplotypes (but not one) is a reliable indicator of very close relatedness. Nonetheless, in fully outbred populations, most full sibs will not share both MUP haplotypes (only 31% shared both MUP haplotypes in the current study), a situation that applies to any highly polymorphic loci, including MHC. Although this does not exclude a large proportion of close relatives, avoidance between mice sharing both MUP haplotypes might be sufficient to drive widespread avoidance between very close relatives under natural conditions. Theoretical modeling predicts that female discrimination against related males in polygynous mating systems will drive male-biased dispersal from natal areas even when discrimination is relatively weak [29]. Consistent with this, the high level of dispersal among young males in house mouse populations [31] will have a general effect of separating close relatives of the opposite sex, reducing inbreeding even between pairs that do not share MUP type.

The use of MUP as a genetic marker for inbreeding avoidance (whether a pre- or postcopulatory mechanism) will promote genome-wide heterozygosity that includes MHC. Despite widely held assumptions in the literature that MHC-based scents are used by females to avoid inbreeding and promote MHC heterozygosity [1, 5–9, 22, 32], direct evidence is surprisingly limited.

Correlations between MHC dissimilarity and mate choice within natural populations [11] might arise through the use of other non-MHC cues because of the normal correlation between MHC and genome-wide similarity. Some congenic laboratory mice that differ only at MHC (though see [7]) show disassortative mating preferences, but others do not [3, 5]. Moreover, the relevance of such studies to the behavior of normal mice is highly questionable [4, 20]. Genetic recognition also needs to be demonstrated against the variable genetic and environmental backgrounds typical of natural populations [15, 33]. Studies of hybrid laboratory mice crossed with wild mice to derive subjects with laboratory-derived MHC haplotypes but with 50% of their genomes from wild mice [4, 21, 34] reveal a deficit of MHC homozygous offspring. However, the design and interpretation of these studies remains controversial [1, 5, 35, 36]. Offspring were typed only for MHC, so parentage could not be assigned and, crucially, parental differences in overall relatedness and other genes that might contribute to inbreeding avoidance could not be assessed. Because the inbred strains used to derive MHC and half of the genome also had two different MUP types, the deficit in MHC homozygous offspring could have arisen from a correlation between MHC and MUP types in the derivation of the founder lines. Genetic variation in these hybrids was also likely to be substantially reduced compared to wild mice. Further research is needed for us to understand whether familial imprinting on MHC, MUP, or other genes is important in much more inbred populations when potential mates might share a complete match to maternal genotypes or perhaps in extreme situations when only MUP identical mates are available. However, such situations could be very unusual because MUP variation persists even in an isolated population with low overall genetic variation that regularly undergoes genetic bottlenecks [37]. Our findings further highlight the limitations of using laboratory strains in mate-selection studies when mice are unable to use their normally variable MUP signals for either kin or individual recognition. Not only will these key signals be unavailable in tests of preference, but the lack of exposure to natural variation in individual scents during the rearing of inbred animals might strongly impact the development of normal recognition processes.

Two broad outcomes derive from this study. First, it challenges the widely held assumption that MHC scents provide a general mechanism across vertebrates to avoid inbreeding and promote MHC heterozygosity. This idea has arisen largely from studies of laboratory mice under extremely abnormal genetic, social, and environmental conditions. Instead, normal wild mice use a set of species-specific urinary proteins to avoid inbreeding that has evolved to provide optimized characteristics for effective signaling through their urine scent-marking system. Given the importance of reliable identity information for mate selection and reproductive success, we should expect animals to evolve signals that are most appropriate for reliable communication in that species. The only known function of MUPs is in scent communication, and they are produced at very high abundance in urine by mice of both sexes but with particularly strong investment by adult males [14]. These proteins provide direct information about genetic identity at several different levels in mice (species, sex, individual, and kinship) ([12, 15, 16, 38], present study) in addition to integrating identity and status information through bound low-molecular-weight volatile pheromones [13, 14, 39]. Individual MUP signatures can easily be identified regardless of other genetic variation between wild mice [12, 15], confirmed by the disruption of recognition when a recombinant MUP is added to wild mouse urine [12]. The fact that mice failed to use MHC scents in addition to MUP when this could have improved inbreeding avoidance suggests that MHC scents might not be easily recognized under naturalistic conditions, despite the ability of mice to discriminate MHC scents in laboratory tests in which background variation is suppressed [7, 8, 40]. Although MUPs are produced in large quantity and are very resistant to degradation, providing a persistent signal that can be left in the environment as well as for direct communication [13], the short peptide ligands bound by MHC [8] lack the structural features for proteolytic resistance and are therefore highly susceptible to both endoproteolytic and exoproteolytic attack. This would limit their reliability and persistence in scent signals. Although this does not mean that MHC has no role in communication through scent, this might be more limited than previously assumed. MUP-like orthologs are likely to be present in other species that use scent communication because proteins can provide direct information on genetic identity, but as yet there has been little investigation of alternative polymorphic markers beyond MHC. Initial studies indicate that MUP-like lipocalins are expressed by many rodents with a high degree of species specificity and diversity of expression [17]. However, species that rely on visual, acoustic, or other forms of communication might encode genetic identity information through alternative polymorphic mechanisms.

The second general implication is that the ability to recognize kin as a mechanism to avoid inbreeding might be more variable across species than previously considered. Mate-selection mechanisms to avoid the deleterious consequences of inbreeding are only likely to be important where kin encounter each other as adults [2]. Because house mice often live at high density in family-based social groups to exploit locally abundant resources [25], the need to recognize kin to avoid inbreeding will be relatively strong. However, the extent of polymorphism in MUP-like urinary proteins varies considerably between *Mus* species [17]. Notably, MUPs are not polymorphic in *Mus macedonicus* [16], a closely related grassland species that lives at much lower densities where individuals are much more widely dispersed than house mice, and thus the need to recognize kin to avoid inbreeding is likely to be considerably less. The ability to avoid inbreeding through kin recognition between adults might be restricted to those species that have evolved specific polymorphic communication signals to achieve this rather than a general ability across vertebrates. Some species might also rely on familiarity of signals learnt during rearing rather than self-referent phenotype matching if relatives reliably interact in the absence of nonkin [41]. The understanding of these issues promises to provide considerable insight into the importance of kin recognition for the avoidance of inbreeding and maintenance of genetic heterozygosity in different social

systems. It is timely to extend the scope of research into such polymorphic signaling systems beyond the MHC.

### Supplemental Data

Experimental Procedures, four figures, and four tables are available at http://www.current-biology.com/cgi/content/full/17/23/2061/DC1/.

### Acknowledgments

We thank Richard Humphries, John Waters, Felicity Fair, Linda Burgess, and Steve Ramm for practical help and Jane Rees for comments on an earlier draft of the manuscript. The work was supported by the Biotechnology and Biological Science Research Council (grant S19816), the Natural Environment Research Council (studentship NER/S/A/2004/12816), and the Wellcome Trust (064691 and 071257/Z/03).

Received: September 5, 2007
Revised: October 17, 2007
Accepted: October 18, 2007
Published online: November 8, 2007

### References

1. Brown, J.L., and Eklund, A. (1994). Kin recognition and the major histocompatibility complex - an integrative review. Am. Nat. *143*, 435–461.
2. Pusey, A., and Wolf, M. (1996). Inbreeding avoidance in animals. Trends Ecol. Evol. *11*, 201–206.
3. Yamazaki, K., Boyse, E.A., Mike, V., Thaler, H.T., Mathieson, B.J., Abbott, J., Boyse, J., Zayas, Z.A., and Thomas, L. (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. J. Exp. Med. *144*, 1324–1335.
4. Potts, W.K., Manning, C.J., and Wakeland, E.K. (1991). Mating patterns in seminatural populations of mice influenced By MHC genotype. Nature *352*, 619–621.
5. Jordan, W.C., and Bruford, M.W. (1998). New perspectives on mate choice and the MHC. Heredity *81*, 239–245.
6. Penn, D.J. (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. Ethology *108*, 1–21.
7. Carroll, L.S., Penn, D.J., and Potts, W.K. (2002). Discrimination of MHC-derived odors by untrained mice is consistent with divergence in peptide-binding region residues. Proc. Natl. Acad. Sci. USA *99*, 2187–2192.
8. Brennan, P.A., and Kendrick, K.M. (2006). Mammalian social odours: Attraction and individual recognition. Philos. Trans. R. Soc. B Biol. Sci. *361*, 2061–2078.
9. Slev, P.R., Nelson, A.C., and Potts, W.K. (2006). Sensory neurons with MHC-like peptide binding properties: disease consequences. Curr. Opin. Immunol. *18*, 608–616.
10. Thom, M.D., Stockley, P., Beynon, R.J., and Hurst, J.L. (2007). Scent, mate choice and genetic heterozygosity. In Chemical Signals in Vertebrates, *Volume 11*, J.L. Hurst, R.J. Beynon, S.C. Roberts, T. Wyatt, and N. Yord, eds. (New York: Springer), pp. 291–301.
11. Piertney, S.B., and Oliver, M.K. (2006). The evolutionary ecology of the major histocompatibility complex. Heredity *96*, 7–21.
12. Hurst, J.L., Payne, C.E., Nevison, C.M., Marie, A.D., Humphries, R.E., Robertson, D.H., Cavaggioni, A., and Beynon, R.J. (2001). Individual recognition in mice mediated by major urinary proteins. Nature *414*, 631–634.
13. Hurst, J.L., and Beynon, R.J. (2004). Scent wars: The chemobiology of competitive signalling in mice. Bioessays *26*, 1288–1298.
14. Beynon, R.J., and Hurst, J.L. (2004). Urinary proteins and the modulation of chemical scents in mice and rats. Peptides *25*, 1553–1563.
15. Cheetham, S.A., Thom, M.D., Jury, F., Ollier, W.E.R., Beynon, R.J., and Hurst, J.L. (2007). The genetic basis of individual-recognition signals in the mouse. Curr. Biol. *17*, 1771–1777.
16. Robertson, D.H.L., Hurst, J.L., Searle, J.B., Gunduz, I., and Beynon, R.J. (2007). Characterization and comparison of major urinary proteins from the house mouse, *Mus musculus domesticus*, and the aboriginal mouse, *Mus macedonicus*. J. Chem. Ecol. *33*, 613–630.
17. Beynon, R.J., Turton, M.J., Robertson, D.H.L., Armstrong, S.D., MacNicoll, A., Humphries, R.E., and Hurst, J.L. (2007). Urinary lipocalins in Rodenta: Is there a generic model? In Chemical Signals in Vertebrates, *Volume 11*, J.L. Hurst, R.J. Beynon, S.C. Roberts, and T. Wyatt, eds. (New York: Springer), pp. 37–50.
18. Ferris, S.D., Sage, R.D., and Wilson, A.C. (1982). Evidence from mtDNA sequences that common laboratory strains of inbred mice are descended from a single female. Nature *295*, 163–165.
19. Yang, H., Bell, T.A., Churchill, G.A., and Pardo-Manuel de Villena, F. (2007). On the subspecific origin of the laboratory mouse. Nat. Genet. *39*, 1100–1107.
20. Manning, C.J., Potts, W.K., Wakeland, E.K., and Dewsbury, D.A. (1992). What's wrong with MHC mate choice experiments? In Chemical Signals in Vertebrates, R.L. Doty and D. Muller-Schwarze, eds. (New York: Plenum), pp. 229–235.
21. Penn, D., and Potts, W. (1998). MHC-disassortative mating preferences reversed by cross-fostering. Proc. R. Soc. Lond. Ser. B Biol. Sci. *265*, 1299–1306.
22. Roberts, S.C., and Gosling, L.M. (2003). Genetic similarity and quality interact in mate choice decisions by female mice. Nat. Genet. *35*, 103–106.
23. Robertson, D.H., Cox, K.A., Gaskell, S.J., Evershed, R.P., and Beynon, R.J. (1996). Molecular heterogeneity in the Major Urinary Proteins of the house mouse *Mus musculus*. Biochem. J. *316*, 265–272.
24. Hurst, J.L., Thom, M.D., Nevison, C.M., Humphries, R.E., and Beynon, R.J. (2005). MHC odours are not required or sufficient for recognition of individual scent owners. Proc. R. Soc. B Biol. Sci. *272*, 715–724.
25. Barnard, C.J., Hurst, J.L., and Aldhous, P. (1991). Of mice and kin - the functional significance of kin bias in social behavior. Biological Reviews of the Cambridge Philosophical Society *66*, 379–430.
26. Dean, M.D., Ardlie, K.G., and Nachman, M.W. (2006). The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). Mol. Ecol. *15*, 4141–4151.
27. Keller, L.F., and Waller, D.M. (2002). Inbreeding effects in wild populations. Trends Ecol. Evol. *17*, 230–241.
28. Zeh, J.A., and Zeh, D.W. (1997). The evolution of polyandry. 2. Post-copulatory defences against genetic incompatibility. Proc. R. Soc. Lond. Ser. B Biol. Sci. *264*, 69–75.
29. Lehmann, L., and Perrin, N. (2003). Inbreeding avoidance through kin recognition: Choosy females boost male dispersal. Am. Nat. *162*, 638–652.
30. Kokko, H., and Ots, I. (2006). When not to avoid inbreeding. Evolution Int. J. Org. Evolution *60*, 467–475.
31. Pocock, M.J.O., Hauffe, H.C., and Searle, J.B. (2005). Dispersal in house mice. Biological Journal of the Linnean Society *84*, 565–583.
32. Tregenza, T., and Wedell, N. (2000). Genetic compatibility, mate choice and patterns of parentage: Invited review. Mol. Ecol. *9*, 1013–1027.
33. Thom, M.D., and Hurst, J.L. (2004). Individual recognition by scent. Ann. Zool. Fenn. *41*, 765–787.
34. Potts, W.K., Manning, C.J., and Wakeland, E.K. (1994). The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. Philos. Trans. R. Soc. Lond. B Biol. Sci. *346*, 369–378.
35. Hughes, A.L., and Hughes, M.K. (1995). Natural selection on the peptide-binding regions of major histocompatibility complex-molecules. Immunogenetics *42*, 233–243.
36. Pomiankowski, A., and Pagel, M. (1992). Sexual selection and MHC genes. Nature *356*, 293–294.
37. Beynon, R.J., Veggerby, C., Payne, C.E., Robertson, D.H., Gaskell, S.J., Humphries, R.E., and Hurst, J.L. (2002). Polymorphism in major urinary proteins: Molecular heterogeneity in a wild mouse population. J. Chem. Ecol. *28*, 1429–1446.
38. Armstrong, S.D., Robertson, D.H.L., Cheetham, S.A., Hurst, J.L., and Beynon, R.J. (2005). Structural and functional differences in isoforms of mouse major urinary proteins: A male-specific

protein that preferentially binds a male pheromone. Biochem. J. *391*, 343–350.

39. Novotny, M.V. (2003). Pheromones, binding proteins and receptor responses in rodents. Biochem. Soc. Trans. *31*, 117–122.
40. Schaefer, M.L., Young, D.A., and Restrepo, D. (2001). Olfactory fingerprints for major histocompatibility complex-determined body odors. J. Neurosci. *21*, 2481–2487.
41. Mateo, J.M. (2004). Recognition systems and biological organization: The perception component of social recognition. Ann. Zool. Fenn. *41*, 729–745.

# Supplemental Data

# The Genetic Basis of Inbreeding Avoidance in House Mice



**Amy L. Sherborne, Michael D. Thom, Steve Paterson, Francine Jury, William E.R. Ollier, Paula Stockley, Robert J. Beynon, and Jane L. Hurst**

## Supplemental Experimental Procedures

### Subjects

The breeding plan simulated a naturalistic situation in which a dominant male sires offspring from several females, resulting in a local population of full-sib and paternal half-sib mice. Thus, animals had a choice of mating with full or half siblings and with individuals with whom they shared zero, one, or two MHC or MUP haplotypes (see Figure S3 for a schematic illustrating the sharing of haplotypes between full and half sibs). This design was used because inbreeding depression is only a high risk between very close (first order) relatives in outbred populations, and inbreeding depression must exceed a substantial threshold before any avoidance behavior is expected to evolve [S1]. Because outbred full sibs are twice as closely related as half sibs, their offspring have twice the risk of inheriting homozygous deleterious recessive alleles, as well as having lower heterozygosity. Factors other than inbreeding depression influence the selection of genetically compatible mates among less closely related mates [S2], and unrelated or more distant relatives are unlikely to share MHC or MUP types. For each of four separate populations, we bred founders by housing a single male house mouse (*Mus musculus domesticus*) with two sets of three unrelated females (each for 14 days) in breeding cages (40 × 23.5 × 12.5 cm). Mice were captive-bred F1–F3 unrelated animals from an outbred captive colony derived from wild ancestors captured from five different populations in the northwest of England, UK. Females were housed singly prior to parturition (cages 40 × 23.5 × 12.5 cm) so that full-sib mice were familiar during rearing but half sibs met only after release into enclosures as adults. For each population, three to five females produced F1 offspring sired by the same male (see Table S1). Offspring were separated into single sex and sibship groups at weaning (4 weeks) and released into enclosures when they were 48–65 days old. Prior to release, all founders were given a subcutaneous radiofrequency identification (RFID) tag for individual identification, and a small tissue sample was taken from the tip of the tail (1–2 mm) under general anesthesia (halothane) for genotyping. Genetic samples were also taken from the founders' parents to allow the individual MHC and MUP haplotypes carried by each founder to be identified. There was no attempt to eliminate endemic parasites or viruses from our stock colony, which was regularly supplemented with wild-caught animals, with the exception of a screening program for the elimination of lymphocytic chorionic meningitis virus. The animal procedures used in this study were approved by the UK Home Office and the Animal Welfare Committee of the University of Liverpool, UK.

### Population Enclosures

Founders sharing the same sire were released simultaneously into one of four very large (25 × 10 m) outdoor enclosures, each containing substantial long grass ground cover. The very large size of the enclosures ensured that all of the founders had the opportunity to establish territories. Vegetative cover was supplemented with 30 nest boxes spread throughout each enclosure, although mice generally preferred to nest in the grass; ten concrete block shelters (45 × 45 × 35 cm) were added after 12 weeks when rain became heavy. Ten food and water stations, spaced evenly around the outer walls of each enclosure, provided ad libitum food (Lab Diet 5002 Certified Rodent Diet) and water. Sheet-aluzinc walls prevented escape or contact between populations (1.3 m high with concrete foundations for the prevention of climbing or burrowing), and wire mesh upper walls and roof prevented predation. Mice were allowed to breed undisturbed before being live trapped 15–19 weeks after founder release. Each founder female could have reared up to three F2 litters to independence over this period that could be clearly distinguished as F2 offspring (see below). Upon capture, founder mice were identified from their RFID tags and housed in captivity for use in further studies. All founders were recaptured and so were available as possible mates throughout the experiment. Sex, weight, and age class (adult, subadult, juvenile) were recorded for all animals, and a urine sample was obtained for further studies on MUP phenotyping. Nonfounders were then culled humanely under halothane anesthetic, and tail snips were taken for genotyping. Blood and gut samples were also taken for other studies.

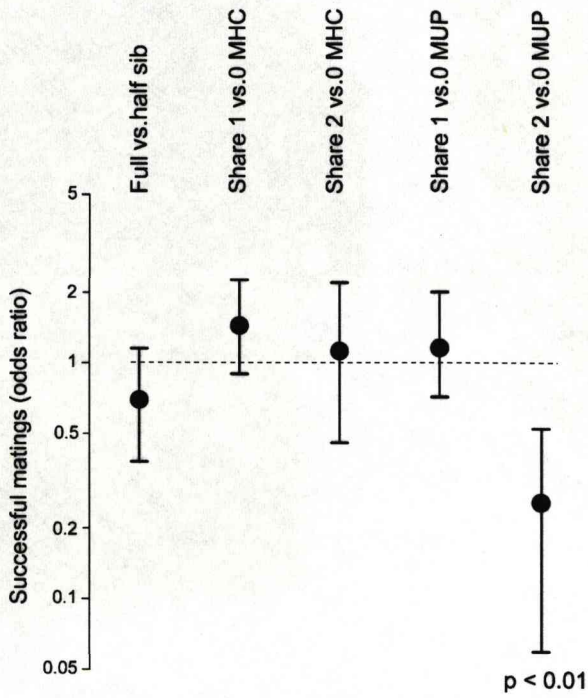



Figure S1. Odds Ratio of Successful Matings between Different Dyads

Successful matings per female with a full sib versus a half sib and when sharing one or two MHC or MUP haplotypes versus sharing no haplotypes, estimated from a multinomial logistic model. An odds ratio of 1 indicates no mating bias (pre- or postcopulatory), and an odds ratio of less than 1 indicates avoidance. Ninety-five percent confidence intervals generated by bootstrapping.

### MHC and MUP Genotyping

We established the MHC and MUP haplotypes of the founders for each population by using eight microsatellite markers across the MHC region on chromosome 17 and eight microsatellite markers surrounding the MUP region on chromosome 4 (Table S3). Full details of DNA extraction, polymerase chain reaction (PCR) amplification, and product analysis are given in reference [S3]. The patterns of alleles present in the founders were compared to their parents so that linked alleles in the same haplotype and any crossover events could be identified. All F1 founders were heterozygous for MHC and MUP except for four females in population B, which were MUP homozygous.

To genotype the F2 offspring of the F1 founders, we selected three to four of the MHC markers and three to four MUP markers that

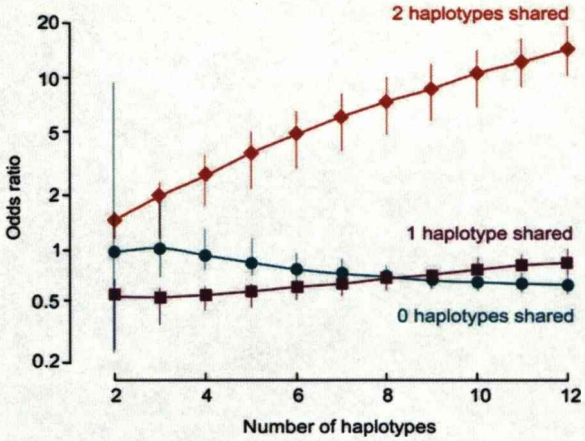



Figure S2. Simulation Model

The odds in favor of a male and female being full sibs given the number of haplotypes that they share at a polymorphic complex. This is equivalent to the Bayes factor, i.e., the extent to which genotype information modifies a prior odds ratio to give a posterior odds ratio. Circles indicate no haplotypes shared, squares indicate one haplotype shared, and diamonds indicate two haplotypes shared. Ninety-five percent confidence intervals are shown. An odds ratio of 1 indicates that the posterior odds ratio is unaffected by the genotype information. The odds in favor of a male and female being full sibs when two haplotypes are shared is analogous to the concept of genotype matching described by Grafen [S13], and here we extend this to the case for sharing only one haplotype.

reliably discriminated each haplotype within a population (different sets of markers were used for each population, see Table S3). To check that MHC homozygotes were correctly identified and were not due to poor primer amplification (leading to only one of two heterozygote alleles being detected by mistake), we ran an additional marker for any animals initially classified as MHC homozygotes. This confirmed homozygosity in all cases.

**Parentage Assignment**

Because parentage could only be assigned reliably to the F2 offspring of the founders, we excluded all animals that might have been F3 s on the basis of their weight, sex, and date of capture. If founders mated immediately after introduction to the enclosures, F2 offspring from the first litters could have started breeding 9 weeks after founder release (wild mice in enclosures rarely breed before 6 weeks old, plus a 3 week gestation period) and produced offspring up to 3 weeks old when trapping began. We therefore excluded all mice that were 3 weeks old or less at week 15, according to sex-specific growth curves for F2 mice subsequently bred in captivity from the same half-sib founders (n = 19 litters weighed at weekly intervals from age 3 weeks). This left 497 offspring across the four enclosures. DNA was extracted from tail snips with 96-well AGOWA mag DNA Isolation kits and a Hamilton Microlab STAR robot. We selected 24 microsatellite markers from the Mouse Genome Informatics site (MGI 3.51) spread across the genome and not linked (>50 cM) to either each other or the MUP or MHC regions (Table S4). The forward primer for each marker was 5′-fluorescently labeled with 6-FAM, NED, PET, or VIC among one of three size groups such that 12 markers could be pooled into a single run. PCR amplification was conducted in 10 μl reactions containing 10 ng of DNA, 0.5 μM of each primer, and 5.0 μl of 2X BioMix Red reaction mix (Bioline [London, UK]) and incubated at 95°C for 2 min, and 30 cycles of 95°C for 30 s, 52°C–58°C (depending on the primer) for 2 min, then 72°C for 30 s followed, with a final extension at 72°C for 10 min. The PCR reactions were then diluted 50- to100-fold and multiplexed in formamide with GeneScan LIZ500 size standard (Applied Biosystems), and size was determined with an ABI PRISM 3100 DNA analyzer and GeneMapper v3.0 software (Applied Biosystems). Parentage analysis was carried out with CERVUS v3.0 [S4] so that individuals

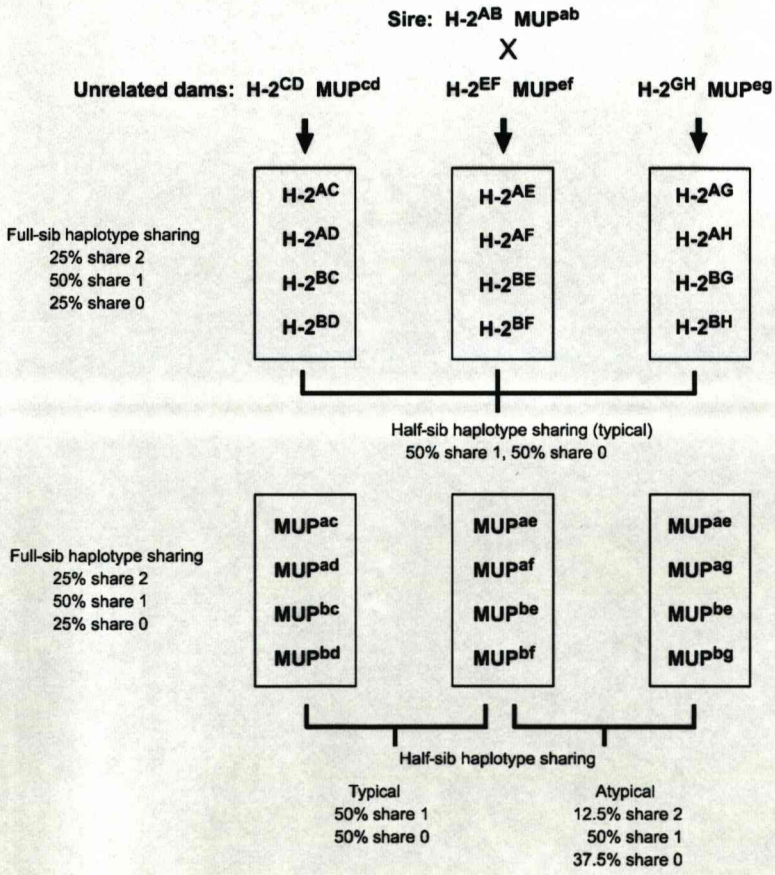



Figure S3. Expected Haplotype Sharing among Founders for MHC or MUP

Full-sib mice derived from unrelated heterozygous parents with different haplotypes for MHC (termed H-2 in mice) and MUP on average will share two haplotypes (25% of full-sib dyads), one haplotype (50%), or no haplotypes (25%). Among half sibs, 50% inherit the same haplotype from their father, whereas 50% share no haplotype. Occasionally, unrelated parents share a haplotype through common descent from the same populations ("MUP$^{e}$" in the example shown), resulting in a small proportion of half sibs that share two haplotypes. Haplotypes, denoted in uppercase letters for H-2 and lowercase letters for MUP, were derived from wild mice and do not refer to known haplotypes from laboratory strains.

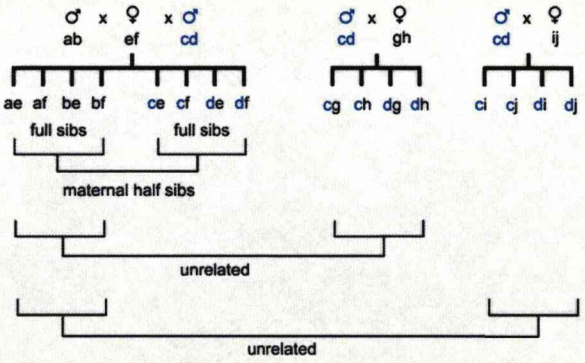



Figure S4. Behavioral Imprinting on Parental Haplotypes

Schematic illustrating haplotype sharing when a female ("ef") mates with more than one male ("ab," "cd"), such as two unrelated territory owners (letters represent MHC or MUP haplotypes, or alleles of other highly polymorphic genes, and show possible combinations among offspring). Familial imprinting on all haplotypes carried by littermates would lead to the incorrect recognition of many unrelated animals as relatives. In the example, offspring of female ef would recognize any animals carrying haplotypes a–f as relatives (depending on haplotypes represented in the litter). This would result in offspring of male ab × female ef incorrectly recognizing male cd and all his offspring as relatives. In house mice, where related females nest communally and each mate with local male territory owners, familial imprinting on littermates could mean that most of the local dominant males together with their offspring are avoided as mates, even when they are not relatives. Familial imprinting only on maternal haplotypes would allow the recognition of all full sibs and maternal half sibs as relatives [S7].

could be assigned to a parent pair with maximum-likelihood methods. All assignments were carried out blind to MHC and MUP type. We then checked for any incompatibilities between parentage and MHC and MUP genotypes. We found mismatches in only 2.8% of cases (14 out of 497 offspring typed). All of these were below the median weight for the youngest mice classified as F2 and, because these were probably F3 offspring, they were excluded from the data set.

### Mating Assignment

Because individual offspring are not independent data points, we determined the minimum number of successful matings that must have occurred to explain the offspring captured. For each founder female, we plotted offspring weight against capture date and used the sex-specific growth curves from captive F2 together with paternity to assign offspring to separate litters (females could have had a maximum of three litters that were over 3 weeks old by the start of trapping). We took the most conservative approach and only assigned offspring from the same father to separate matings if they were very unlikely to have come from the same litter (based on their weights and a maximum litter size of nine for wild house mice). This resulted in the assignment of 193 successful matings across the four enclosures, with all females producing at least two litters. Notably, 67% of litters were sired by more than one male. Because matings were assigned blind with respect to MHC or MUP genotypes, any errors in assignment could not cause any bias in our analysis.

### Statistical Analysis

We used the statistics package R v2.3.1 (www.r-project.org) to fit logistic multinomial models to the data with likelihood methods [S5]. Two response variables were analyzed: (1) the number of successful matings made by a female from each of a set of available males and (2) the number of offspring produced from each of a set of observed matings. The log likelihood can be given by

$$\left.\begin{aligned} \ln L &= \sum_i \sum_j y_{ij} \ln P_{ij} \\ P_{ij} &= \frac{\exp\left((\mathbf{x}_{ij} - \mathbf{x}_{i1})\beta\right)}{\sum_j \exp\left((\mathbf{x}_{ij} - \mathbf{x}_{i1})\beta\right)} \end{aligned}\right\}, \quad \text{(S1)}$$

in which $y_{ij}$ is the observed number of matings or offspring from sire $j$ and dam $i$, and $\mathbf{x}_{ij}$ is a vector of explanatory variables that describes sire $j$ with respect to dam $i$ (i.e., whether the sire is a half or full sib of the dam and whether the dyad shares one, two, or no MHC or MUP haplotypes, respectively). $\beta$ is the vector of coefficients, fitted by the maximization of the log likelihood with numerical optimization. In the case of two-tailed tests, the significance of explanatory variables was determined by the comparison of LR statistics (i.e., twice the log likelihood ratio for a pair of nested models) to a distribution generated by a permutation procedure in which the genotype or relatedness of the dam was randomly reordered with respect to the set of available males. This permutation approach is more conservative than the comparison of LR statistics against a chi-square distribution because it controls for potentially inflated Type I errors that might arise from the repeated preference of a dam to a particular sire, unconnected to the relatedness or genotype of the sire with respect to the dam. In the case of one-tailed tests, the significance was determined by the comparison of the coefficient fitted to the observed data with the corresponding distribution generated by the permutation procedure.

### Familial Imprinting

Animals might avoid inbreeding not only by the comparison of their own genotype to that of potential mates but also by behavioral imprinting on the scents of close relatives during rearing. Previous evidence for MHC effects on sexual preferences suggest that mice prefer mates of different MHC type from the parental odors experienced in the nest rather than those differing from self [S5–S8], and it has been hypothesized that negative imprinting on MHC-determined odors would allow females to avoid inbreeding with a greater proportion of kin than self inspection alone [S7]. Because extra-pair matings and multiple paternity occur frequently in house mice [S9, S10], and related females often rear their offspring communally [S11], indiscriminate imprinting on all nest mate haplotypes would be very error prone and result in avoidance of many unrelated mates (Figure S4). However, imprinting only on maternal haplotypes would allow the recognition and avoidance of all full sibs and maternal half sibs [S7]. To test this, we compared the likelihood of mating according to whether or not males had an MHC or MUP haplotype that matched one of those carried by a female's mother. Note that all males necessarily shared one MHC and one MUP haplotype with the mother of full sib sisters, whereas some also shared one MHC or MUP haplotype with mothers of paternal half sibs through common ancestry from the same populations. Because these were outbred populations, no males shared both MHC haplotypes with a female's mother, and only a very small number of dyadic combinations shared both MUP haplotypes (2.8%, all full sibs). We also checked for any evidence of negative imprinting on maternal haplotypes among males [S8] but found no effects.

### Simulation Modeling

We investigated, across a range of possible populations, whether a male sharing 0, 1, or 2 haplotypes with a female was a good guide as to whether he was a brother or not. We modeled the odds in favor of a male and female being full sibs compared to being unrelated given the number of haplotypes that they share at a polymorphic locus,

$$\frac{\Pr(\text{sibs}|x)}{\Pr(\text{unrelated}|x)} = \frac{r}{1-r} \cdot \frac{p(x|\text{sibs})}{p(x|\text{unrelated})}, \quad \text{(S2)}$$

i.e., the posterior odds ratio of a male and female being full sibs or not is the product of the prior odds ratio (in the simplest case derived from the proportion, $r$, of full sibs that a female encounters) and the Bayes factor [S12], which is the odds ratio of the probabilities that a male and a female will share $x$ haplotypes given that they are or are not sibs. The Bayes factor, therefore, is the quantity of interest because it is a measure of whether the extra genotypic information gained from odor cues have increased or decreased the relative odds of a potential mate being a brother.

In order to estimate the odds ratio $p(x|\text{sibs})/p(x|\text{unrelated})$, we performed separate simulations for different numbers of alleles at a locus, ranging from two alleles up to 12 (Figure S2). For each simulation, we drew 1000 allele frequency distributions at random. From each allele frequency distribution, we estimated $p(x|\text{sibs})$

Table S1. Summary of F1 Founders and F2 Offspring in Each Population

| | Population | | | | |
|---|---|---|---|---|---|
| | A | B | C | D | Total |
| F1 Founders | 6 female, 7 male (3 litters) | 12 female, 18 male (5 litters) | 7 female, 12 male (3 litters) | 8 female, 11 male (3 litters) | 33 female, 48 male (14 litters) |
| MHC haplotypes | 6 | 9* | 5 | 6* | 16 |
| MUP haplotypes | 7* | 9 | 7 | 7* | 20 |
| Number of Female:Male Dyads | | | | | |
| Full sib | 18 | 36 | 14 | 28 | 96 |
| Half sib | 24 | 180 | 70 | 60 | 334 |
| MHC Haplotype Sharing between Founders (% Full-Sib Dyads/% Half-Sib Dyads) | | | | | |
| both | 27.8/0 | 38.9/0.6 | 57.1/4.3 | 25.0/6.7 | 35.4/2.4 |
| one | 50.0/62.5 | 41.7/59.4 | 42.9/95.7 | 53.6/45.0 | 46.9/64.7 |
| none | 22.2/37.5 | 19.4/40.0 | 0/0 | 21.4/48.3 | 17.7/32.9 |
| MUP Haplotype Sharing between Founders (% Full-Sib Dyads/% Half-Sib Dyads) | | | | | |
| both | 27.8/0 | 22.2/3.3 | 35.7/2.9 | 42.9/13.3 | 31.3/4.8 |
| one | 44.4/54.2 | 52.8/59.4 | 50.0/48.6 | 32.1/46.7 | 44.8/54.5 |
| none | 27.8/45.8 | 25.0/37.2 | 14.3/48.6 | 25.0/40.0 | 24.0/40.7 |
| F2 offspring genotyped | 99 | 192 | 101 | 91 | 483 |
| Number of Matings Observed (Expected) According to MHC Haplotypes Shared | | | | | |
| both | 2 (3.4) | 9 (5.7) | 1 (5.3) | 5 (4.6) | 17 (19.1) |
| one | 20 (16.6) | 50 (45.1) | 41 (36.7) | 19 (19.0) | 130 (117.3) |
| none | 8 (10.0) | 22 (30.2) | 0 (0) | 16 (16.4) | 46 (56.5) |
| Number of Matings Observed (Expected) According to MUP Haplotypes Shared | | | | | |
| both | 0 (2.9) | 3 (5.9) | 0 (3.2) | 3 (8.8) | 6 (20.7) |
| one | 22 (15.4) | 51 (48.1) | 19 (20.7) | 22 (16.4) | 114 (100.5) |
| none | 8 (11.7) | 27 (27.1) | 23 (18.2) | 15 (14.8) | 73 (71.8) |

"*" Indicates that one founder inherited a crossover haplotype, combining alleles from two haplotypes. For analysis, this was regarded as a match for either of the original haplotypes.

and $p(x|\text{unrelated})$ from 250 samples in which we picked two genotypes randomly according to Hardy-Weinberg expectations, set these as parents, and then generated a daughter from these parents and determined (1) the proportion of her brothers sharing 0, 1, or 2 alleles and (2) the proportion of unrelated males in the population sharing 0, 1, or 2 alleles with her, which we approximated from Hardy-Weinberg expectations.

### Supplemental References


S1. Kokko, H., and Ots, I. (2006). When not to avoid inbreeding. Evolution Int. J. Org. Evolution *60*, 467–475.

S2. Bateson, P. (1983). Optimal outbreeding. In Mate Choice, P. Bateson, ed. (Cambridge: Cambridge University Press), pp. 257–278.

S3. Cheetham, S.A., Thom, M.D., Jury, F., Ollier, W.E.R., Beynon, R.J., and Hurst, J.L. (2007). The genetic basis of individual-recognition signals in the mouse. Curr. Biol. *17*, 1771–1777.

S4. Kalinowski, S.T., Taper, M.L., and Marshall, T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. *16*, 1099–1106.

S5. Jobson, J.D. (1992). Categorical and Multivariate Methods, *Volume II*, Applied Multivariate Data Analysis. (New York: Springer).

S6. Penn, D.J. (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. Ethology *108*, 1–21.

S7. Penn, D., and Potts, W. (1998). MHC-disassortative mating preferences reversed by cross-fostering. Proc. R. Soc. Lond. Ser. B-Biol. Sci. *265*, 1299–1306.

S8. Yamazaki, K., Beauchamp, G.K., Kupniewski, D., Bard, J., Thomas, L., and Boyse, E.A. (1988). Familial imprinting determines H-2 selective mating preferences. Science *240*, 1331–1332.

S9. Potts, W.K., Manning, C.J., and Wakeland, E.K. (1991). Mating patterns in seminatural populations of mice influenced By MHC genotype. Nature *352*, 619–621.

S10. Dean, M.D., Ardlie, K.G., and Nachman, M.W. (2006). The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). Mol. Ecol. *15*, 4141–4151.

S11. Manning, C.J., Wakeland, E.K., and Potts, W.K. (1992). Communal nesting patterns in mice implicate MHC genes in kin recognition. Nature *360*, 581–583.

S12. Bernardo, J.M., and Smith, A.F.M. (2000). Bayesian Theory (Chichester, United Kingdom: Wiley).

S13. Grafen, A. (1990). Do animals really recognize kin? Anim. Behav. *39*, 42–54.

Table S2. Summaries of Observed and Expected Frequencies among F2 Offspring

| Number of Matings | | | |
|---|---|---|---|
| Relatedness | Half Sib | Full Sib | |
| Observed | 159 | 34 | |
| Expected[a] | 149.16 | 43.84 | |
| MHC Sharing | None | One Haplotype | Both Haplotypes |
| Observed | 46 | 130 | 17 |
| Expected[a] | 56.53 | 117.35 | 19.12 |
| MUP Sharing | None | One Haplotype | Both Haplotypes |
| Observed | 73 | 114 | 6 |
| Expected[a] | 71.76 | 100.51 | 20.73 |
| Number of Offspring | | | |
| Relatedness | Half Sib | Full Sib | |
| Observed | 414 | 69 | |
| Expected[b] | 395.92 | 87.08 | |
| MHC Sharing | None | One Haplotype | Both Haplotypes |
| Observed | 135 | 311 | 37 |
| Expected[b] | 124.78 | 310.57 | 47.65 |
| MUP Sharing | None | One Haplotype | Both Haplotypes |
| Observed | 159 | 314 | 10 |
| Expected[b] | 164.48 | 304.67 | 13.85 |
| Offspring Genotype | | | |
| MHC | Homozygous | Heterozygous | |
| Observed | 94 | 389 | |
| Expected[c] | 96.25 | 386.75 | |
| MUP | | | |
| Observed | 77 | 406 | |
| Expected[c] | 83.50 | 399.50 | |

[a] Expected on the basis of the proportion of dyads of each type available for each female.
[b] Expected on the basis of matings of each type per female.
[c] Expected on the basis of parental genotypes and number of offspring per mating.

Table S3. Microsatellite Markers Used for MHC and MUP Genotyping

| MHC Marker | Location | Forward Primer | Reverse Primer | Size | Repeat | Label | Comments | F2* |
|---|---|---|---|---|---|---|---|---|
| D17Mit230 | 17 b2 32.3 Mb | 5′-TGACATAAACCTC TGGCTTCC-3′ | 5′-CCAGCCCATCTAAA GTCATTTC-3′ | 288 | CA | ned | duplicated in all mice | A, B, C, D |
| D17Mit22 | 17 B2 33.9 Mb | 5′-GCATTAGATAGAGA GTAGATGGGTTG-3′ | 5′-ATGGATGGCGAGAA TGAGAC-3′ | 216 | GT | hex | used | A, B, D |
| D17Mit231 | 17 B2 34.1 Mb | 5′-GCCTCAGCAAGAC CCTAAAC-3′ | 5′-ACTCCTCCTTTTCCC TCTCC-3′ | 285 | GT | hex | used | |
| D17MIT13 | 17 B2 34.6 Mb | 5′-TGCAGGCAAGATC CAAGAAG-3′ | 5′-GAAAGAGGGTGTCG ATGCTC-3′ | 239 | GT | hex | used | |
| D17Nds3 | 17 B3 34.8 Mb | 5′-CAGCCTTAATGGG TCTGGTC-3′ | 5′-ACAGAGGGGAAGAG GAAAGC-3′ | 222 | CT | ned | used | B |
| D17Mit47 | 17 B3 35.8 Mb | 5′-CTGAGACCAGTGCA GTGGAA-3′ | 5′-TTTTTCAATATGTGAG CATGTGC-3′ | 238 | CA | fam | used | C |
| D17Mit24 | 17 B3 37.0 Mb | 5′-ACCTCTCACCTCTC TCTGTG-3′ | 5′-GCAAGTTTAGGGATCT TTTCTCC-3′ | 131 | CA | fam | used | A, B, C, D |
| MUP Marker | Location | Forward Primer | Reverse Primer | Size | Repeat | Dye | Comments | F2* |
| D4NDS6 | 4 B3 52.8 Mb | 5′-CGGGGAAGGTTGT TTGTTTG-3′ | 5′-AGGCCAGCAATGTAG AAAGG-3′ | 240 | GT | ned | used | |
| D4Mit139 | 4 B3 55.2 Mb | 5′-TCAAACTGGGAAG AGCCAAG-3′ | 5′-GCCGTAGAAGAGAAG TAATTTTTCC-3′ | 149 | GT | hex | used | D |
| D4Mit241 | 4 C1 55.7 Mb | 5′-TTTCCAGTGTTGT CCAGAGC-3′ | 5′-AAGGCAAATCACTAG GTGCTG-3′ | 219 | CA | hex | used | D |
| D4Mit288 | 4 C1 56.8 Mb | 5′-ACATTCAGCAAA GACTGAGCAC-3′ | 5′-TGCCATTTGTTATAGA CCATGC-3′ | 168 | CA | ned | used | A, B |
| D4 mit164 | 4 C1 59.5 Mb | 5′-AACACATATATACC AAGGCAGCAC-3′ | 5′-ATTTCCACCCTGTCCA CTCC-3′ | 142 | CA | fam | used | A, B, C |
| D4Mit243 | 4 C1 59.6 Mb | 5′-AGCCCTACTGATT GCTCTCC-3′ | 5′-TGGAAAGTTGAAAAC CACTGC-3′ | 168 | CA | fam | used | A, B, C, D |
| D4Mit217 | 4 C1 59.7 Mb | 5′-ACTCAATTAGGTT GTTCAGATAGCC-3′ | 5′-GGCACTTGCTGCCA CATC-3′ | 246 | GT | hex | used | |
| D4Mit17 | 4 C1 62.8 Mb | 5′-GCCAACCTCTGTG CTTCC-3′ | 5′-CCTCTGACATCCAC ACACATC-3′ | 138 | GT | fam | used | A, B, C, D |

"*" Indicates that after MHC and MUP haplotypes in F1 founders and their parents were identified with all listed markers, only a subset were necessary for the identification of haplotypes in F2 offspring for each of populations A–D.

Table S4. Microsatellite Markers Used for Parentage Assignment

| Marker | Position (cM) | Location (Mb) | Forward Primer | Reverse Primer | Size | Label |
|---|---|---|---|---|---|---|
| D1Mit58 | 1 8.3 | 9.7 | 5′-GGACTGGCAATCCTCTTGTC-3′ | 5′-GCACGTTAGAGAGTGGGCTC-3′ | 254 | VIC |
| D1Mit155 | 1 112.0 | 196.1 | 5′-ATGCATGCATGCACACGT-3′ | 5′-ACCGTGAAATGTTCACCCAT-3′ | 252 | NED |
| D2Mit405 | 2 68.9 | 149.7 | 5′-TGATTATATCTTGGAATACACGTGTG-3′ | 5′-CTGTGTAGCAAAACAGTTTATGGC-3′ | 84 | 6-FAM |
| D3Mit1 | 3 11.2 | 0.3 | 5′-TGTGCACAGGGGTACATACA-3′ | 5′-TCATTTTCTTCCTCCCCCTC-3′ | 143 | VIC |
| D3Mit163 | 3 87.6 | 157.3 | 5′-TGGATACATACATATACATGGAAATGC-3′ | 5′-TTTCTCCAGACCCATGAACC-3′ | 143 | 6-FAM |
| D4Mit171 | 4 6.3 | 22.6 | 5′-CAGGTGTAATAATGGTTTTTTGACC-3′ | 5′-CATATTAAATAAACACAGCAGCACG-3′ | 318 | PET |
| D4Mit310 | 4 71.0 | 147.5 | 5′-TCTCCACGTGTGTGCCTTAG-3′ | 5′-TGAAAGCACTCTGCAGACTCA-3′ | 117 | PET |
| D5Mit25 | 5 61.0 | 111.8 | 5′-AACACACCTCCATACTGGTCG-3′ | 5′-GGCTAACTGAAATTGTTTTGTGC-3′ | 234 | NED |
| D6Mit105 | 6 45.5 | 108.6 | 5′-CTGTCTCCACTACTTCTATTCCTGG-3′ | 5′-CAAAAGCCTTATATATTACACCTCACC-3′ | 237 | 6-FAM |
| D7Mit253 | 7 55.0 | 115.3 | 5′-TGTGGGTGCAACCAAATG-3′ | 5′-TTTGGTGATATAGATACTAGGTGTGTG-3′ | 89 | NED |
| D8Mit29 | 8 33.0 | 70.7 | 5′-CCCTAGTGTATACATAGAGGGGTG-3′ | 5′-TCTTTGTGTTGTGATGTGTTGTAA-3′ | 109 | VIC |
| D9Mit12 | 9 55.0 | 99.4 | 5′-ATTCAAGGGGCAGTACACAT-3′ | 5′-TGGTCCTGGTAAAACTGCCT-3′ | 96 | 6-FAM |
| D10Mit80 | 10 4.0 | 11.5 | 5′-CAAAAAAAACCCTGATTCTACCA-3′ | 5′-GTGTGCATATGGCAGTAACTTTG-3′ | 154 | NED |
| D10Mit98 | 10 59.0 | 102.0 | 5′-TCAGGCATCTAGTGAGATGATCC-3′ | 5′-CCCATAGATGCAGGGGTG-3′ | 150 | VIC |
| D11Mit63 | 11 2.0 | 17.1 | 5′-GCCCACAACTTTGTGTCCTT-3′ | 5′-TTGACCATGCTCCTCATCAG-3′ | 139 | PET |
| D11Mit300 | 11 68.0 | 111.3 | 5′-TTTGGCTGTGATAAAACAAAACA-3′ | 5′-TTGAGTTTTGATTTTGTATGTGGG-3′ | 145 | 6-FAM |
| D12Mit4 | 12 34.0 | 79.9 | 5′-ACATCCCCAGCTCTTGTTTG-3′ | 5′-AAACCAAACCAAAGAAGCTTAGG-3′ | 201 | PET |
| D13Mit77 | 13 73.0 | 118.0 | 5′-TCTTTGAAGTCCCTTTCAAAGC-3′ | 5′-ATAGCACTGCACTCATGCTCA-3′ | 275 | VIC |
| D14Mit212 | 14 13.5 | 36.2 | 5′-AACATGTGCACTGGAACAATG-3′ | 5′-TCATTTATCAATTTACTTTGGTGAGG-3′ | 102 | PET |
| D15Mit161 | 15 69.2 | 97.9 | 5′-TCTGTTTTGTTTGTTCGTTTGC-3′ | 5′-TAAAATCTCCCTGTATACAAGTCTGTG-3′ | 99 | NED |
| D16Mit71 | 16 70.5 | 98.0 | 5′-TAGAAAATCTTCAAATAGGATCTGTTC-3′ | 5′-GAGCATTTCCCTTTTACCTGG-3′ | 154 | PET |
| D17Mit94 | 17 45.9 | 75.8 | 5′-TGGAGAGGCATCCAACTCTC-3′ | 5′-TTCCTCTTAGTCCACCTTTTGC-3′ | 146 | NED |
| D18Mit33 | 18 44.0 | 70.0 | 5′-GCATGTCGTATCCATAAACATACG-3′ | 5′-ATGCGGGCTTGACTTCTG-3′ | 140 | VIC |
| D19Mit70 | 19 51.0 | 50.8 | 5′-AAAATATCAGGGCATGGTGC-3′ | 5′-GGGTTATTAGGAAAATTTATGTTGTG-3′ | 195 | 6-FAM |