

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Genetics of house mice (*Mus domesticus*) on the Isle of May, Scotland

Michael Fenton BSc (Hons)

Thesis submitted to the University of Nottingham for the degree of Masters by Research

April 2018

Supervisor: Jan Bradley

Abstract

The Isle of May hosts a unique population of house mice, *Mus domesticus*, following an artificial introduction of Eday mice onto the island 35 years previously. DNA from the uniparentally inherited Sry gene on the Y chromosome and D-loop region in the mitochondrial genome was sequenced from mice from the Isle of May, Eday, nearby Inchkeith and mainland Tollerton, in order to identify SNPs for use as molecular markers. These would be used to distinguish mice of different parental origin, for the purpose of investigating associations between different haplotypes in the Isle of May mice and variation in immune function, and to confirm earlier phylogenetic work on the population using updated techniques. It was suspected that mice of different parental origin had the potential to be exhibiting behavioural and fitness differences such that they may be segregating to different parts of the island whereupon differential selection pressures could act to shape their immune variation. Ruling out such possibilities enables the visible impact of other environmental factors to be magnified and their effect on immune function understood. The Sry gene failed to yield a significant amount of polymorphism between mice of different locations, and as such, investigations were instead focused on detecting population structuring in the population. This was done via analysis of a panel of immune SNPs which could be considered effectively neutral where genetic drift outweighs selection, as Is often the case in small populations such as the Isle of May. Population genetics analysis yielded conflicting results from the Bayesian clustering program STRUCTURE, and an isolation by distance plot produced a negative correlation between genetic distance (1/(1-Fst)) and linear geographic distance, the opposite way to expected. There was not enough sufficient evidence to confirm that the population may be sub-structured, and reject a null hypothesis of panmixia. However, the genetic dynamics of this population are far from clear. Such results may be circumvented through principal components analysis, which may more reliably detect subtle structuring in this feral mouse population.

Acknowledgements

I would like to extend my thanks to the following for their support in this project, without whom this would not have been possible:

- Jan Bradley
- John Brookfield
- Angus Davison
- Benoit Poulin
- Ann Lowe
- Jonathan Fenn
- Stuart Young
- Sarah Goertz
- Christopher Taylor

Table of Contents

Abstract…… i

Acknowledgments……ii

Table of Contents…… iii

1. Introduction…… 1

1.1 Overview of the thesis…… 1

1.2 Background on *Mus domesticus*…… 1

1.3 Introduction to the Isle of May, Scotland…… 3

1.4 Introduction to Eday, Orkney…… 4

1.5 Study of house mice…… 4

1.6 *Mus domesticus* on the Isle of May…… 5

1.7 Introduction of mice from Eday onto the Isle of May…… 7

1.8 Phylogenetic analysis of post-introduction house mice on the Isle of May…… 9

1.9 Population genetics analysis of house mice on the Isle of May…… 16

1.10 Sampling areas on the Isle of May…… 17

1.11 Molecular tools for phylogenetic and population genetics study…… 19

1.12 Use of immune SNPs for population genetics analysis…… 20

2. Aims and Objectives…… 22

3. Materials and Methods…… 23

3.1 Trapping of mice…… 23

3.2 DNA extraction…… 25

3.3 Ctla4 PCR…… 25

3.4 Sry PCR…… 27

3.5 D-loop PCR…… 29

3.6 Gel electrophoresis…… 30

3.7 Ethanol precipitation…… 30

3.8 Gel extraction..... 31
3.9 Sequencing..... 32
3.10 Analysis of D-loop data..... 32
3.11 Analysis of immune SNPs..... 33
4. Results..... 37
4.1 Sequencing of the Sry region..... 37
4.2 Analysis of D-loop data..... 43
4.3 Immune SNP analysis..... 53
4.3.1 Linkage disequilibrium plot..... 53
4.3.2 STRUCTURE analysis..... 55
4.3.3 Hardy-Weinberg calculations..... 56
4.3.4 Isolation by distance analysis..... 57
4.3.5 Phylogeny of individuals..... 59
5. Discussion and Concluding Remarks..... 60
5.1 Phylogenetics..... 60
5.2 Population structure..... 62
5.3 Concluding remarks..... 66
References..... 67

1. Introduction

1.1 Overview of the thesis

The usefulness of house mice as a model organism in the study of biology is unbounded, and though most research has gained from manipulations of laboratory populations, wild populations provide a whole different level of insight into the fields of ecology, immunology, and population genetics. The study of Berry et al. (1991) on house mice (*Mus domesticus*), particularly, has been fundamental in the development of understanding of isolated natural populations. In his study, house mice were introduced from Eday, an island in the Orkney island group to an isolated natural population of feral house mice on the Isle of May in the Firth of Forth, and successfully, and unexpectedly, introgressed. Behavioural differences between the current mice on the island may be occurring as a result of this hybridisation event. This thesis presents an attempt to employ modern molecular techniques to provide answers to the genetic questions that have arisen out of this introduction. Determining the dynamics of this population will allow for investigations of natural immune variation in the wild Isle of May mice to go forward, with important implications for understanding of the natural functioning of the immune system in vertebrates.

1.2 Background on *Mus domesticus*

The house mouse, *Mus musculus*, has successfully spread throughout the world to all continents with the exception of Antarctica and survives in a diverse range of environments, from grasslands and forests to salt marshes and deserts. The evolutionary and natural history of the house mouse is of relevance to its usefulness as a model organism. The most recent branch of the genus Mus, it is thought to have originated in southwestern Asia (Suzuki et al., 2013, Suzuki et al., 2015). From there it radiated out into several subspecies lineages between 0.37-0.47 million years ago, namely *M. musculus*, found in Asia and Eastern Europe, *M. castaneus*, exclusively in Asia, and the subject of this study, *M. domesticus*, which has a range that covers Western Europe, Australasia, Africa, the Americas and a number of oceanic islands (Suzuki et al. 2013). Analysis of mitochondrial DNA (mtDNA) has typically been

employed to investigate the phylogeography of the house mouse at a genetic level, due to the substantial availability of D-loop sequence data for house mice throughout their distribution, and based on the success of studies utilising mtDNA for this purpose, particularly regarding its use to indicate initial colonisation events of an area (Boursot et al. 1996; Gabriel et al., 2010; Gabriel et al., 2011). Analysis of nuclear DNA has also assisted in the understanding of the phylogeography of sub-species of the house mouse, usually in concordance with mitochondrial DNA analysis (Din et al., 1996; Yakimenko et al., 2015).

The speed of radiation and success of the *Mus* species group has been largely due to its association with people, which is so significant that the colonisation history of house mice can be used to infer human migration patterns (Jones et al., 2013). In particular, the sub-species *Mus domesticus* experienced several waves of dispersal associated with its commensalism with sedentary humans, a relationship that dates back to southern Levantine hunter-gatherer settlements in 15,000 B.P., prior to Neolithic farming (Weissbrod et al., 2017). Aggregation of cereal crops into food stores, increased waste accumulation and construction of artificial human shelters constructed a new niche into which house mice were able to settle. From there they followed human migration patterns first into the Eastern Mediterranean Basin around 7,800 B.C., then further into the Mediterranean and Western Europe much later during the Iron Age, with the delay in colonisation indicative of important potential factors such as the degree of anthropization of Western areas compared to the Eastern Mediterranean and competition with other anthropophilous rodent species (Cucchi et al., 2005). More recent range expansions for *Mus domesticus* are attributable to passive transport by humans, such as those in sub-Antarctic islands (Pocock et al., 2005; Hardouin et al., 2010).

House mice are primarily commensal throughout their distribution, though in some areas where conditions are appropriate they are able to revert to a non-commensal, “feral” state, where they are not dependent on human beings for their survival (Silver, 1995a). Numerous examples of feral populations of house mice are described, such as throughout New Zealand and Australia, where they are known to be detrimental to native species (Fitzgerald et al., 1981; Parkes and Murphy, 2003; Brown 2007). On the British mainland, house mice are mostly unable to depart from commensal lifestyles, with competitor species such as the wood mouse (*Apodemus sylvaticus*) excluding them from feral niches (Tattersall et al. 1997). As such, feral

populations tend to be confined to offshore islands devoid of competition and where the risk of predation is either low or completely absent, such as Skokholm and the Isle of May (Berry, 1968; Triggs, 1991). Feral populations are also observed to be able to "switch" back to a commensal lifestyle where such conditions allow them to (Silver, 1995a).

1.3 Introduction to the Isle of May, Scotland

The Isle of May sits approximately 8km east of the Scottish coast in the outer Firth of Forth, reachable by ferry from nearby Anstruther. Approximately 1.6km long and 0.5km wide, the 57-hectare island plays host to a remarkable diversity of animal life, with over 285 bird species recorded on the island; as such it was declared a National Nature Reserve in 1956 - one of over 50 currently operating in Scotland ("The Story of the Isle of May National Nature Reserve", 2014). In addition to its rich bird life, the Isle of May supports several mammalian species, namely rabbits (*Oryctolagus cuniculus*), grey seals (*Halichoerus grypus*) and house mice (*Mus domesticus*). House mice have been present on the island for an indeterminate amount of time, first mentioned by Muir (1885) who quotes an undated letter from a lighthouse keeper on the island "We have no rats but legions of mice, and most impudent mice they are". As will be discussed in section 2.1.3, this longstanding population of house mice remained relatively unchanged until a number of mice from the island of Eday in the Orkney archipelago were introduced to the island in 1982 in a study of gene flow and hybridisation by Berry and colleagues (1991). This introduction fundamentally changed the genetics of mice on the Isle of May and is central to this study.

For its size, the island also boasts an impressive cultural history. A priory was founded on the island in the 12th century, dedicated to St. Adrian of May, killed there by raiding Vikings in 669. The site became a centre of pilgrimage in the middle ages, with James IV said to have made frequent visits ("St. Adrian's Priory", 2012). Earlier chapels existed on this site dating back to the 6th century, with recent archaeological excavation uncovering flint arrowheads and a piece of stone axe that brings back the date of human habitation of the island to at least 2000BC ("The Story of the Isle of May National Nature Reserve", 2014). A village existed on the island for a brief period in the 17th and 18th centuries, with the last villager, John Wishart, dying in

1730. The current lighthouse on the island was constructed in 1816, though its original lighthouse was built in 1636, the first to appear in Scotland ("The Story of the Isle of May National Nature Reserve", 2014). Mice may have been introduced to the island at any stage along with humans settling the island.

1.4 Introduction to Eday, Orkney

The island of Eday is situated in the Orkney archipelago, situated off the north-eastern Scottish coast. It is inhabited by 130 people as of 2013, though supports a much larger bird and wildlife population. The house mouse, *Mus domesticus*, is included amongst these animals. The phylogeography of Orkney mice has previously been investigated, and they constitute a mitochondrial DNA lineage which has a distribution across, and is restricted to, the Northern and Western edges of the British Isles and Norway, consistent with the Norwegian Viking "sphere of influence" (Searle et al., 2009). Mice were captured from Eday and released into the resident Isle of May population, which until that point had remained relatively unchanged, in 1982 (Berry et al., 1991). At the time of introduction, mice on Eday were genetically distinct from the Isle of May mouse population (Berry and Peters, 1977; Berry, 1981; Scriven 1992). This study, for analytical purposes, also includes mice from Inchkeith, a neighbouring island to the Isle of May, and Tollerton, a rural pig farm in Nottinghamshire which would act as a mainland reference.

1.5 Study of house mice

House mice are one of the most studied organisms on the planet. This is owed to their usefulness as models for studying mammalian biology, particularly regarding immunity, development and neurological disorders (Morse, 2007). Their usefulness in the study of genetics goes back to Mendel, who is suspected to have bred mice before he bred pea plants (Henig, 2000). Short generation times, relatively large litters and the ease with which they are reared are all reasons they make for excellent organisms of study in both the lab and the wild. Wild house mice are particularly informative due to the extent to which they have ecologically and genetically differentiated in their colonisation of the globe, highlighted with the

arrival of modern molecular mapping techniques (Guenet and Bonhomme, 2003), and are able to shed light on areas of biology laboratory mice cannot.

Isolated populations that exist on small islands such as the Isle of May prove ideal for furthering the study of ecology, due to the simplicity of the habitat they occupy. This project intends to further understanding in the field of ecological immunology through study of the population genetics of the Isle of May mice. Study of ecological immunology aims to learn how the immune system functions in its natural setting, with the ecology and population dynamics that influence variation in immune response to infection taken into consideration (Schulenburg, 2009). The Isle of May house mouse population, as opposed to mainland populations, experiences no migration and likely negligible predation (Triggs, 1991). With these factors removed, the effect of other ecological processes on the immune variation within the population is amplified to the degree that their impact can be more accurately inferred (White and Searle, 2007). The Isle of May, as will be discussed, hosts a unique, isolated feral population that accommodated a successful introgression of commensal mice following an artificial introduction event (Berry et al., 1991), and, as such, provides a unique scenario where population sub-division may be occurring due to behavioural and/or physiological differences between descendants of different origin populations. This project aims to uncover any correlation between these mice and the variation in their immune response, while considering their ecological context.

1.6 *Mus domesticus* on the Isle of May

House mice on the Isle of May have been the subject of scientific study for decades. Southern (1938) described them as of “a comparatively large race” in a survey of vertebrate fauna on the island, noting that their morphology was typical of island mice found around the British Isles. Similar observations have been made of insular populations of small rodents and other mammals compared with their mainland counterparts (Foster, 1963; Foster, 1964), an effect Lomolino (1985) explains as possibly the result of reduced competition within some island communities which in smaller species such as the house mouse is likely to select for individuals of a larger body size.

May mice have an analogous pattern of breeding and survival to those on British islands with a similar ecology such as Skokholm in Wales, with climate having a crucial effect on natality and mortality. Low spring numbers increase to a peak in the middle of winter, following which breeding is minimal and population numbers decrease. May mice experience this seasonal mortality in January/February in contrast to mice on Skokholm for whom it is October/November (Berry, 1968). Density-dependent regulatory factors are implied by the degree to which the population increases in size relative to the numbers of overwintered mice from the previous year, though these are not known (Triggs, 1991). Generally, the mice on the island keep within their ranges, which remain relatively stable and overlap with each other. However, up to 15% of the mice that survive the winter have been observed to move from the area which they occupied in autumn to a new location on the island, such that the population is regularly "churned" to the degree that it can considered a panmictic unit (Triggs, 1991).

Owing to their geographic isolation the population is genetically constrained, with no migration possible. A sampling of 93 mice from the original population on the island prior to the Eday introduction found the mice to be completely monomorphic in 17 loci examined in a study involving 22 island mouse populations; the only other populations producing such a phenomenon being the nearby island of Inchkeith and Fugløy in the Faroe Islands (Berry and Peters, 1977). A similar result was obtained when 71 loci were examined in a sample of 40 May mice, where only seven variants were found at two loci, again prior to the introduction of mice from Eday to the island in 1982, after which monomorphism of the population was no longer the case (Berry et al., 1991). Further to this, the Isle of May mice, prior to the introduction, were seen to be monomorphic at loci of the H-2, the major histocompatibility complex (MHC) in mice (Figueroa et al., 1986). The lack of polymorphism seen at these various loci was most likely a consequence of genetic drift acting on the population, a result of a genetic bottleneck such as a small number of founder mice, or both (Berry, 1986; White and Searle, 2007). The prior monomorphic state of the May mice appeared not to have impacted their ability to survive in their environment, and they were seen, by virtue of several long-term studies (Triggs, 1991, and the references therein), to be biologically comparable to mice on the ecologically similar island of Skokholm, which were polymorphic at 18% of their loci (Berry, 1981).

Widely scattered resources and a lack of predation have also been implicated for a lack of aggressive behaviour observed in the Isle of May mouse population (Gray and Hurst, 1998), in addition to the generally reduced aggressiveness seen in island populations of rodents (Adler and Levins, 1994). Predation risk in open habitats normally limits rodents to a small number of protective nest sites; in its absence, such sites are harder to defend from intruders taking advantage of the free use of open areas surrounding the sites, while resources, when not concentrated, are more difficult to aggregate and thus defend effectively, contributing towards a reduced requirement for aggression (Jensen et al., 2005). Such explanations may account for the overlap seen in the ranges of the May mice. However, the dynamics of predation upon the Isle of May house mice and the impact it may have on their behaviour are not fully understood. Great black-backed gulls (*Larus marinus*), for example, are known to be permanent residents of the island and are considered its top predator. These birds are opportunistic and are known to feed on terrestrial mammals such as mice, with 72 nesting pairs recorded in 2016, mainly around Rona on the north of the island (Outram, 2016). These numbers are known to increase in the early months of winter, drawn by the grey seal pupping season. Transient birds may also be of significance in the predation dynamics on the island. Short-eared owls are known to feed mainly on rodents and winter on the Isle of May, likely subsisting on the house mice (Steel, 2018).

1.7 Introduction of mice from Eday onto the Isle of May

In April 1982, 77 house mice, 42 males and 35 females, captured from Eday in the Orkney Islands were introduced into the local Isle of May population, estimated at the time to be composed of around 1000 individuals, in an effort to further understanding of gene flow and hybridisation (Berry et al., 1991). Released at the southern end of the island, they were observed to have undergone rapid dispersal, and 6 months after the introduction hybrid mice were trapped on all areas of the island, indicated through the possession of introduced Eday alleles at any of 6 segregating allozymic loci detected from electrophoresis of blood samples. Following a period of rapid increase, frequencies of introduced alleles were seen to have stabilised, and by 1985 no significant heterogeneity was observed on the island (Berry et al., 1991).

This manner of rapid introgression of foreign alleles into the local population was unexpected. Prior to the introduction there was a general consensus that immigrant mice into a resident population will generally be unsuccessful. This conclusion was supported by studies such as that of Anderson, Dunn and Beasley (1964), in which mice heterozygous for a deleterious t-allele with a 95% transmission rate by males were introduced into an island population of mice in which the allele was absent. While in theory an allele with such a high transmission rate might spread rapidly and increase in frequency, it only managed to become established in the local area into which it was introduced, after which it dispersed slowly with no overall increase in frequency within the island population; following this the allele disappeared from the population (Myers, 1974b). Such a result highlighted the probable importance of social and ecological factors in introgression of foreign high-transmission-rate alleles, such as sub-structuring of the population into separate mating groups, though this did not account for the picture as a whole as it was later discovered that heterozygosity for the t-allele did incur some fitness deficit (Berry, 1986). Studies continue to draw similar conclusions; an investigation into *Mus domesticus* in the Kerguelen archipelago demonstrated that following the initial colonisation of an island by mice, further colonisation attempts are likely to be unsuccessful, due to the local mice being better adapted to their environment (Hardouin et al., 2010).

The rapid increase in frequencies of introduced alleles in the May population suggested that there was a lack of territorial division in the population, unlike the rigid demic structuring seen in prior studies of house mice (Klein, 1975, for a review), though it is not the first such island study to exhibit a diversion from this paradigm (Berry and Jakobson, 1974). Indeed, mice captured on the island in 1995 were shown to have significantly reduced aggression in comparison to commensal mice living on the British mainland (Gray and Hurst, 1998). This finding was unexpected, though may be explained by the different selective pressures operating in the contrasting environments in which commensal and feral house mice are found (Frynta et al. 2005). Food for example, typically limits feral populations which consequentially must form large home-ranges (Krasnov and Chochlova, 1994; Choquenot and Ruscoe, 2000), while it is available in excessive quantities to commensal populations in human-built environments, which, as a result, reach comparatively high densities

whereby a hierarchal system of dominant males aggressively defending exclusive territories prevails (Gray et al., 2000).

A follow-up study investigating differences in gene flow between the sexes based on frequencies of introduced Y-chromosome and mitochondrial haplotypes on the island determined that the success of the introgression was largely male-determined, with the agonistic and territorial behaviour of the previously commensal Eday mice likely enabling them to outcompete the less aggressive, native May males (Jones et al., 1995). Due to the high levels of inbreeding on the island prior to the introduction, success of the introduced males may also be explained by female mate choice driven by inbreeding avoidance, with May females preferentially mating with males that were least related to them (Jones et al., 1995). It is further plausible that introduced alleles were able to spread quickly due to some extent to the effects of heterosis, whereby hybrids were at a fitness (generally defined as lifetime reproductive success) advantage relative to resident mice depressed by inbreeding (Ingvarsson and Whitlock, 2000). Heterosis was indicated by body size and skeletal traits of the hybrids (Scriven and Bauchau, 1992), and likely contributed to the success of the introduced alleles, as they may have been swept to higher frequencies on the back of the competitively successful hybrids, though the extent to which this is the case is unknown.

1.8 Phylogenetic analysis of post-introduction house mice on the Isle of May

As mentioned, Jones and colleagues (1995) conducted a follow-up study of the post-introduction mouse population on the Isle of May several years after the artificial introduction event. This study investigated the phylogenetics of the mouse population on the island, and utilised genetic markers from maternally inherited mitochondrial DNA and paternally inherited Y-chromosomal DNA to distinguish between mice of different parental origin from the ancestral Eday and pre-introduction May populations.

Using these markers, they were able to illuminate how gene flow was operating within the population immediately following the introduction of the commensal mice from Eday to the feral mouse population on the Isle of May. Their analysis showed that after the introduction, the most predominant genotype for

males on the island was of an individual with May mitochondria and Eday Y chromosomes, such that the introduced Eday Y-chromosome had largely replaced the original May Y-chromosome within the population. The May Y-chromosome was seen to persist in the population, though only in a small proportion of the males sampled, while the May mitochondria were at a much higher frequency in the population than the introduced Eday mitochondria, indicative of no advantage of Eday females over May females in breeding success. This indicated that differential genetic success was being experienced between the sexes of introduced mice, with Eday males more successful than native May males. This in turn affected the rates and patterning of gene flow within the population (Jones et al., 1995).

It was not known what definitively caused the successful spread of the immigrant Y-chromosome into the Isle of May population at the expense of the native Y-chromosome, though several suggestions were put forward by the study's authors. Sexual selection, which in most mammals involves both intrasexual competition among males and female mate choice (Darwin, 1871), may have been acting on the males of a specific genotype. Female mate choice driven by inbreeding avoidance may be responsible to an extent, whereby male immigrant mice were chosen by the native females on the basis of their genetic dissimilarity to the native males. This can be manifested via odour specific cues involving major urinary proteins (MUPs) present in urine, that house mice may use to distinguish between mates on the basis of genetic compatibility (Sherborne et al., 2007). Furthermore, as has been discussed, there are behavioural differences between feral and commensal mice involving levels of aggression, which are heightened in commensal populations where resources are more abundant and ranges are smaller and more stable (Gray and Hurst, 1998; Gray et al., 2000). This may have served to benefit the reproductive success of the Eday males on the Isle of May, who were potentially able to drive away less aggressive native competitors and achieve more matings. It should further be considered, should they be the underlying cause for the success of the introduced males, that these components of sexual selection may have acted in tandem rather than independently (Wong and Candolin, 2005).

Genes located on the paternally inherited non-recombinant regions of the Y-chromosome and on the maternally inherited cytoplasmic genes may be implicated in behavioural and fitness differences between animals of different parental inheritance, such as the mice on the Isle of May. Jones et al. (1995) postulated the

idea that the differences in inter-male agonistic behaviour between the males of different ancestral origin on the Isle of May could have a genetic basis in the Y-chromosome. The connection between Y-chromosome variants and differential degrees of aggressive behaviour has been previously investigated in strains of inbred wild and inbred lab mice (Carlier et al., 1990; Roubertoux et al. 1994; Van Oortmerssen and Sluyter, 1994). Multiple genes, including Sry (sex-determining region Y), on the non-recombining region of the Y chromosome, and Sts (steroid sulfatase), on the pseudoautosomal region of the Y chromosome, have been associated in variance in aggression (Guillot et al., 1995; Mortaud et al., 2010; Dow et al., 2011), as has interaction of offspring with the maternal postnatal environment (Carlier et al., 1991). Sry is likely involved in strain differences in aggressive behaviour (Maxson 1996a; Maxson 1996b), and, as it doesn't undergo recombination, could lead to continued differential aggression between mice currently on the Isle of May of different paternal origin, though the degree to which it solely affects intermale aggression is not defined.

This relies on the assumption, however, that the introduced Eday Y chromosome has not completely replaced the original May Y chromosome, which is probable given various other factors which may have been responsible for the success of its initial spread. The amount of time that has elapsed since the phylogenetics of the population were last investigated must also be taken into account.

Gray and Hurst (1998) concluded that male mice on the Isle of May after the introduction displayed generally reduced aggression with respect to commensal house mice, so it may be unlikely that mice on the island are exhibiting aggressive behaviours towards each other. A lack of scarring was found on the captured males in their study, implying a lack of aggressive encounters that resulted in injury. In addition, there was a lack of response, either startled or defensive, from male mice encountering a different male, with the implication that they did not expect aggression from other males (Gray and Hurst, 1998). The extent to which these observations are representative of the behaviour of all the mice on the island is not certain however, nor is the extent to which their experimental parameters accurately reflect the natural situation on the Isle of May. However, a lack of aggression does seem to be a reasonable assessment of the position.

Taking such observations into account, the intricacies of competitive behaviour in feral house mice are by no means fully understood, and there is a limited amount of literature that explores it. Dominance behaviour prevails in mice where there are limited safe nesting sites that provide protection from predation and turbulent weather, given that the small size of most rodents leaves them vulnerable to such pressures (Hurst et al., 1996). Weather conditions on the Isle of May are typically harsh such that the original May population when investigated showed signs of adaptation to cold conditions such as huddling behaviour and relevant physiological traits (Triggs, 1991). Gray and Hurst (1998) attributed the lack of aggressive behaviour observed to a lack of significant predation. Predation risk for the mice on the island may vary temporally, however, with seagulls reaching record nesting numbers on the island in recent years, causing concern for the local puffin population (Steel, 2017).

Such factors may interplay in the dynamics of dominance behaviour in this population, in which variable levels of aggression may be important in determining spatial structuring in a landscape where available nesting sites differ in quality of protection from both predators and the elements. It could be the case, for example, that mice with an original May Y-chromosome are driven to sub-optimal habitat on the island by more aggressive mice possessing an Eday Y-chromosome.

Maternally inherited cytoplasmic genes such as those contained within mitochondrial DNA have classically been assumed to be selectively neutral (Ballard and Kreitman, 1995). However, some relatively recent work has focused on how interactions between the mitochondrial and nuclear genomes may govern fitness with respect to life-history traits and health phenotypes at both the intrapopulation and interpopulation levels in species (Wolff et al., 2014). Such work is predicated on the knowledge that one of the most basic and vital processes in functioning of eukaryotes, oxidative phosphorylation, is dependent on these mitochondrial-nuclear interactions (Ryan and Hoogenraad, 2007).

Due in part to high mutation rates and a lack of recombination in the mitochondrial genome, non-adaptive forces allow for the accumulation of deleterious mutations in functional regions of mitochondrial DNA, including those associated with oxidative phosphorylation, such that compensatory mutations in the nuclear genome that restore function are highly selected for (Hagström et al., 2014,

Wolff et al., 2014). As a result, polymorphic alleles in both mitochondrial and nuclear DNA involved in these interactions are likely under intense selective pressure to maintain function in mitochondrial activities, with disruption in this tight co-evolution seen for example to increase disease susceptibility (Bar-Yaacov et al., 2012; Wolff et al., 2014; Gershoni et al., 2014). Research in insects has found that particular nuclear and mitochondrial genotype pairings, relevant to several distinct population haplotypes from which they were obtained, yielded faster or slower development times in sand beetles depending on the temperature at which they were reared (Dowling et al., 2007). Some genotype pairings would do better at a higher temperature and worse at a lower temperature, and for other genotypes, vice versa. Another experiment of this sort in sand beetles showed similar findings with regards to metabolic rate (Arnqvist et al., 2010).

This is congruent with the notion that populations that are geographically isolated from each other are likely to evolve away from each other through mutations that more optimally adapt each population to its own specific environment. This includes co-adaptations in interacting mitochondrial and nuclear regions of DNA to temperature, with a shift in the optimal temperature range involving trade-offs through reduced function at different temperatures (Pörtner et al., 2007).

No experiments in this domain of investigation have yet focused on such questions in house mice. However, it is possible that there may be interpopulation variation in mitochondrial and nuclear genomes optimised for a higher temperature in the commensal Eday mouse population, than the historical May population. This is considering the fact that commensal mice on Eday had access to human structures which provided warmer temperatures than feral mice on the Isle of May would have been exposed to in winter months. Eday mice would likely have utilised these, given observations that house mice are infrequently found outside human dwellings in the UK, with increased frequencies of mice found in dwellings in colder months (Langton et al., 2001; Lambert et al., 2017). Hybrid offspring between introduced Eday and native May mice may have had a fitness advantage against parental May genotypes with regards to sexual selection, such that they were able to outcompete a less aggressive historical May male phenotype, or appear preferable to females through, for example, possessing more variable MHC genotypes. Whether such a reproductive benefit in the hybrid mice is then counterbalanced by a reduction in fitness at the

harsher temperatures that mice may be experiencing on some more exposed parts of the Isle of May, due to disrupted adaptive mitochondrial-nuclear genotypes remains to be seen. Should this be the case, mice closely resembling the original population prior to the introduction event may persist in these harsher areas, having been driven there by more aggressive mice, and being able to be more reproductively successful in harsher climate than hybrid mice. Even given the possibility that the original May Y chromosome has been completely replaced by the Eday Y-chromosome, mtDNA differences between mice of different maternal inheritance may still produce fitness differences related to metabolic function based on the temperature of the different sites occupied by the mice.

Moving from the previous molecular analysis (Jones et al., 1995), the phylogenetics of this population over 20 years later will be revisited using modern molecular techniques. Various intrinsic and extrinsic factors with regards to the current population of house mice on the Isle of May were being investigated by others in my lab in the University of Nottingham parasitology department. These included parasite load, diet, microbiome constitution and various health and immune system parameters as part of a wider study of wild house mice immunity on the Isle of May. Determining the phylogenetic properties of the current Isle of May population will assist in establishing whether the original May Y-chromosome is still present on the island, and whether the introduced Eday mitochondrial haplotype is persisting at similar frequencies to those observed by Jones et al. (1995). Such a result would be indicative of no differential fitness of this haplotype relative to the native May haplotype. Further to this, any relationship between mitochondrial and Y-chromosome genotype, phenotype, consisting of health and immune function, and various ecological considerations such as parasite load, diet and location can be investigated and assessed.

Such work builds on the relatively recent field of ecological immunology, which seeks to understand how ecological factors contribute to the variation in immunity in natural populations (Sheldon and Verhulst, 1996; Schulenberg, 2009)., Ecological immunology is a multidisciplinary approach which combines the study of immunology with ecological constraints such as evolutionary biology, population genetics and biotic and abiotic interactions. This enables the limitations of classical reductionist immunological research, in which natural variation is intentionally minimised using inbred strains of lab mice in a sterile environment (Viney, 2006;

Schulenburg et al., 2009) to be circumvented. It is vital to understand all these ecological interactions in order to further understanding of how the immune system functions in its natural state in the wild; particularly in the house mouse, a species upon which the majority of the body of immunological research is based.

One of the most important ecological pressures to consider with regard to immune variability is infectious disease. It is known that strong immune responses can increase fitness in areas of high infection, however, they incur a fitness cost in absence of selective pressure from parasites and other pathogens (Sheldon and Verhulst, 1996; Martin et al., 2008; Graham et al., 2010). As such, investment in immune defence has been seen to correlate with parasite prevalence (Lindström et al., 2004; Corby-Harris and Promislow, 2008; Mayer et al., 2016). Due to the fitness cost of both inducing too strong an immune response and one that isn't strong enough, host fitness is generally seen to be maximised through an intermediate response to parasite infection (Stjernman et al., 2008; Råberg, 2014). Various parasitic infections have been found in the Isle of May mice by my department including Babesia microti, pinworm (Syphacia oblevata) and whipworm (Trichuris muris). Parasite prevalence in the Isle of May has also been investigated previously within my department, establishing that parasites in the Isle of May mice are highly aggregated (Robertson, 2014). This correlates with other studies suggesting that parasites tend to aggregate within populations in a few highly infected individuals, with most of the population having either very few or no parasites (Shaw and Dobson, 1995; Woolhouse et al., 1997; Shaw et al., 1998; Baird et al., 2012).

Teasing apart the other ecological factors that tie into immune response in addition to parasite prevalence is difficult, as the extent of resistance to, or tolerance of, pathogens by the immune response is also proportional to the cost of initiating such a response in relation to the fitness of the organism as determined by its environment (Graham et al., 2010). Such factors that affect organismal fitness on an island such as the Isle of May include diet, in respect of the quantity and quality of food, weather, and predation (Bedhomme et al., 2004; Lazarro et al., 2008; Otti et al., 2012). The selection pressure each of these factors imposes on the fitness, and thus immunity, of house mice can be considered to vary spatially across the Isle of May.

The landscape on the Isle of May differs in the amount and quality of shelter available for house mice. So, there are areas on the Isle of May where mice can be more exposed to wind and lower temperatures, and, to an unknown extent, predators. The mice on the Isle of May have the potential to exhibit variance in aggressive behaviours and core fitness traits associated with their distinct paternal and maternal lineages. This may affect the distribution of, and gene flow between, different mitochondrial and Y-chromosome haplotype combinations across the island, where these groups could be exposed to different selective pressures. Separation of groups of mice based on their genotypes may also affect spatial distribution of parasites, whereby one group of mice harbour more of the highly infected individuals by chance, and thus exhibit higher transmission rates (Woolhouse et al., 1997). Differences in selective pressures between genotype groups of mice would reflect in immune variation, and as such, this study aims to either highlight or rule out such relationships through revisiting the phylogenetics of this unique island population, to further understanding in ecological immunology.

Jones et al. (1995) performed the original analysis of the phylogenetics of the Isle of May population using restriction fragment length polymorphisms (RFLPs) between Eday and native May mice. While these were able to distinguish between mtDNA and Y-chromosomes from Eday and from the Isle of May, they were limited in the resolution of polymorphism that they were able to provide (Silver, 1995b). Modern molecular techniques utilising DNA sequencing and SNPs as genetic markers will be able to provide more information on the specific variation between paternally and maternally-inherited genetics on this island, such as which SNPs differ between haplotypes on the island. Furthermore, the findings of Jones et al. (1995) relied on the fundamental assumption that the diagnostic molecular markers they used were reliable. Through identification of specific SNPs between the sequences of mice on the Isle of May, from Eday, from Inchkeith and from Tollerton, these findings will be able to be confirmed.

1.9 Population genetics analysis of house mice on the Isle of May

The relationship between maternal and paternal inheritance, ecological factors and immune variation was not investigated further in this project due to the unfortunate lack of polymorphism found in the Sry gene fragment examined. Due to time constraints, and due to the availability of a panel of immune SNPs in a sample of

Isle of May mice, the aims and objectives of this project were adjusted; both are described in section 2. Phylogenetic analysis of mitochondrial haplotypes present on the island would continue, to investigate whether the introduced Eday mtDNA was still present on the island, and to confirm the findings of Jones et al. (1995). However, rather than looking at genetic factors which may be responsible for population structuring within the Isle of May house mice population and looking for corresponding effects of differential selection pressures on sub-populations, population structure will be investigated directly. Acceptance of a null hypothesis of panmixia would allow for investigation of other genetic and environmental factors underlying immune variation in the May mice to be continued by my department, with the knowledge that the mice are either more, or less likely being differentially affected by selection pressures.

To understand the genetic diversity in populations, and thus understand variation in immune response, several key processes need to be understood. Gene flow, genetic drift and natural selection all play a part in shaping allele frequencies in populations, and are interdependent on one another (Andrews, 2010). If the Isle of May population does not show population structuring, then it can be considered to be panmictic, and gene flow is not restricted between groups of mice. Genetic drift can then be thought not to be driving immune variation between individuals on the island, through either stochastic loss of immune alleles or of parasites (White et al., 2012), and other causes identified.

1.10 Sampling Areas on the Isle of May

Traps were placed in a diverse variety of notable locations across the Isle of May where mice were likely to be caught. These locations correspond to named areas of the island such that they were constant across trapping visits to the island. Some of the locations where traps were laid were quite distinct in their features, which may have important implications for the mice that inhabit them. Trapping site names and locations are discussed in detail in Materials and Methods. Isolation by distance will be considered as a measure of population genetic structure, an effect whereby mice that are further away from each other are more genetically distinct due to geographically limited dispersal reducing gene flow (Wright 1943). The effect of geographic distance on genetic differences between groups of mice has previously

been demonstrated (Ryan et al., 1993; Dallas et al., 1995; Muñoz-Muñoz et al., 2011). The distances at which mice begin to differ genetically as a consequence of distance demonstrated in these studies however have been much greater than the length of the Isle of May, which is only 1.6km by 0.5km, so it might be expected that gene flow is not limited between mice on opposite ends of the island. Importantly, the Isle of May mice are feral in comparison to the commensal mice examined in these studies, so differences in dispersal are of relevance, as feral house mice generally disperse further and more frequently than commensals (Pocock et al., 2005). Study of the May house mice in particular has suggested that they move all over the whole of the island, such that geographic distance need not be considered a barrier to gene flow (Berry et al., 1991). With this considered, isolation by distance is a simple model relying on the assumption that linear distance is the only thing constraining gene flow, whereas in complex landscapes this may not be the case, and as such, more recent landscape genetics studies aim to investigate the effects of landscape on effective distances in such analyses (Van Strien et al., 2015). In line with this, the effect of the sampling sites used in this analysis should be considered when interpreting such results.

Sampling sites differed in the amount and type of shelter they offered from both the elements and potential predators such as owls and gulls. As such, mice may more readily disperse from some sites than others. Triggs (1991) observed that mice inhabiting open, more featureless areas tended to have larger ranges than those mice whose habitats offered cover. There are no human made structures in the Burnett's Leap sampling site for example, and it is the furthest North of all the sampling sites, and closest to Rona where the majority of the island's great black-backed gull population is known to nest (Outram, 2016). A limited number of rabbit and/or puffin burrows offering shelter in this area may therefore mean that mice move out of this area fairly frequently. On the other hand, the Main Light and Garden sampling sites are fairly central on the island, and consist of walled gardens which enclose shrubbery, all of which offer more protection to mice within these sites from both inclement weather and predators. Mice may therefore may be less inclined to disperse from these sites than other areas, though that is not to say that they may not be dispersing at all. Furthermore, these two sites are adjacent while being walled off from one another. Topology of sites may affect levels of gene flow relative to distance between them.

1.11 Molecular tools for phylogenetic and population genetics study

Molecular genetic markers, in recent decades, have been hugely significant in biological research, particularly regarding ecology, conservation and evolution (Morin, 2004), and they can be used for both study of parental inheritance and of population genetic structure. As has been discussed, paternal and maternal inheritance markers have previously been used in restriction fragment length polymorphism (RFLP) analysis in a study on the Isle of May mouse population to identify differences between sexes in gene flow following the Eday introduction event (Jones et al., 1995). Several decades later, molecular techniques in investigations of phylogeography have advanced, and use of RFLPs as a genetic marker in such analyses as those of Jones et al. (1995) has decreased, namely due to technical reasons keeping them from being utilised for wider range of applications (Schlötterer, 2004; Hoshino et al. 2012). Modern phylogenetic and phylogeographic studies typically involve use of microsatellites or SNPs as genetic markers. Microsatellites have previously been relied upon more heavily as they are highly polymorphic due to high mutation rates and can be informative for both parentage identification and population genetics (Jones et al., 2010; Guichoux et al., 2011). However, microsatellite-based population studies suffer from drawbacks not associated with SNP analysis; such as lack of reproducibility between labs and homoplasy, and appropriate mutation models can vary among loci (Piry et al., 1999; Seeb et al., 2011; Haasl and Payseur, 2011). As such, SNPs are fast becoming the molecular tool of choice in population studies (Putman and Carbone, 2014), and are reported to be preferable to microsatellites for hybrid and parentage analyses (Väli et al., 2010; Hauser et al., 2011), while being able to perform similarly to microsatellites in population structure inference even at low numbers of SNPs (Coates et al., 2009; Livingstone et al., 2010).

Free from meiotic recombination, the Y chromosome and mitochondrial genome are inherited uniparentally and are thus able to be used to trace paternal or maternal lineages in many animals. While maternal inheritance markers alone are powerful for such purposes as informing on initial colonisation events (Gabriel et al., 2010), when viewed in tandem with paternal inheritance markers they can inform on

the patterns of interaction between mice of different origin populations (Terashima et al., 2006). Recent successes in phylogeographic studies in mice include development of Y chromosome microsatellites as paternal inheritance markers (Hardouin et al., 2010; Jones and Searle, 2015). This study, in keeping with the advancement of SNPs as the molecular tool of choice for population study, aims to use a SNP typing approach to study the population genetics of the Isle of May house mice population. Building on a previous phylogenetic study of this population (Jones et al., 1995), any changes in phylogeography will be monitored which may be, for example, indicative of a short-term advantage of heterosis from admixture wearing off (Rius and Darling, 2014), and the extent of penetration of the Eday female genetics. Population genetic analysis testing a null hypothesis of panmixia will further reveal the consequences of the successful artificial introduction event decades later and could be informative towards the growing field of reintroduction biology, whereby potential consequences of experimentally restored immigration into an endangered inbred population could be better predicted.

The Y chromosome in mammals suffers from a general lack of variation and as such areas of polymorphism for identifying SNPs between individuals within a sub-species are sparse (Greminger et al., 2010). Sry, a gene on the non-homologous region of the Y chromosome encoding a testis-determining factor, was chosen as a candidate gene having had previous success in a phylogeographic study of house mice in Japan (Terashima et al., 2006). The DNA sequences of uniparentally inherited Sry regions will be obtained in order to identify SNPs with which to type mice by alignment of sequences of mice from the Isle of May, Eday, a nearby island to the Isle of May, Inchkeith, and a mainland reference, Tollerton. The mitochondrial D-loop region, in contrast, has previously been extensively used in phylogeographic studies (Yonekawa et al., 2003; Terashima et al., 2006; Searle et al., 2009). Sry sequences will be considered alongside information from pre-obtained mitochondrial D-loop sequences and autosomal bi-allelic genetic markers to investigate the ecological genetics of this unique island population as it exists today.

1.12 Use of Immune SNPs for population genetics analysis

Selectively neutral SNPs have traditionally been segregated for use in analysis of genetic structure and demographic history of populations, as loci under

balancing or directional selection will violate most assumptions of neutral genetics models, and may affect estimates of genetic differentiation (Nielsen et al., 2006). As such, measures are typically taken to identify outlier adaptive genetic markers so that they may be removed prior to inferences about neutral evolutionary processes (Beaumont and Nichols, 1996; Helyar et al., 2011). SNPs derived from functional genic regions associated with the immune response are likely to be affected by both positive directional selection for mutations that increase resistance to disease (Bernatchez and Landry, 2003; Barreiro and Quintana-Murci, 2010;) and balancing selection for maintenance of polymorphism at particular loci such as those of the major histocompatibility complex (MHC) (Aguilar et al., 2004; Spurgin and Richardson, 2010).

However, while such loci can be influenced by adaptive, in addition to neutral (Lynch et al., 2011), processes, in small populations, particularly those that are isolated, genetic drift appears to be the dominant evolutionary force shaping genetic variation rather than selection (Strand et al., 2012; Grueber et al., 2013; Gonzalez-Quevedo et al., 2015). In such populations, immune SNPs have been able to inform on population structure to a similar degree to neutral genetic markers (Elbers et al., 2017). It is not unreasonable to suspect that this may be the case with the mice on the Isle of May; the population that inhabited the island prior to the Eday introduction was previously found to be monomorphic at the H-2, the MHC in mice, usually highly polymorphic in mainland mouse populations (Figueroa et al., 1986). Such observations highlight the importance of founder effects, genetic drift, or both in insular populations, given their ability to overcome balancing selection.

It should still be taken into account, however, that there are inevitable caveats to the use of immune SNPs. While their use has previously yielded inferences that mostly correlated with those of microsatellites, which are assumedly neutral, differences have still been observed that may be attributable to the varying effects of selection between immune SNPs and microsatellites, such as the inferred number of clusters in STRUCTURE (Pritchard et al., 2000) in assessments of admixture (Elbers et al., 2017). The extent to which this is causally related to the use of immune SNPs rather than the differences between SNPs and microsatellites is, however, unknown, as it has been the case that non-genic SNPs and microsatellites yield different estimates of cluster number in STRUCTURE in other studies (Morin et al., 2012).

Furthermore, successful correlations with other marker types are dependent on the effective neutrality of the immune markers, which is by no means guaranteed.

However, given that the effects of selection on genetic markers derived from immune loci may be outweighed by the effects of genetic drift in a small isolated population such as the Isle of May, it was considered appropriate to attempt to derive possible population sub-structuring in the Isle of May mouse population using the immune SNPs available to me, though caution should be taken when interpreting such results.

2. Aims and Objectives

The original aims of the project were to investigate the following null hypotheses:

- There is no correlation between variation in immune response and maternal/paternal inheritance of the mice on the Isle of May
- There are no inconsistencies in the inferences of Y-chromosome and mitochondrial haplotype distribution of the Isle of May mice made by RFLPs and the inferences made by SNPs

Following the failure to find polymorphism in the Sry gene fragment however, aims were reconsidered given the amount of time remaining, and the following null hypotheses were investigated:

- There is no non-random mating in the Isle of May mouse population
- There are no inconsistencies in the inferences of mitochondrial haplotype distribution of the Isle of May mice made by RFLPs and inferences made by SNPs

3. Materials and Methods

3.1 Trapping of Mice

Mice were captured over the course of several days during mid-October 2016 on the Isle of May. Of relevance are the locations from which the mice that will be used in the isolation by distance analysis were sampled; these are detailed in Table 3, and are described below:

- Fluke Street: This a path in the middle of the island lined with buildings, including the human residences. Traps were laid outside these buildings and along the foot of the path that leads up to Main Light, near which much human food waste is discarded. Traps were also laid in the small garden that exists between the buildings and the Byres, and in the available bedrooms
- Byres: Traps were laid around the walled courtyard outside the Byres, which is towards the Kirkhaven harbour from Fluke Street
- Main Light: Traps were laid within a walled garden near the main lighthouse on the island along the walls. This garden is laden with shrubbery and contains a Heligoland trap used by ornithologists for capturing birds
- Garden: Traps were laid in the adjacent walled garden to that described in the Main Light sampling site, which is also encloses more shrubbery than the surrounding area
- High Tarn: Traps were laid just of the path in the High Tarn area. This location, along with Three Tarn Nick and Burnett's leap, is more exposed, being at the highest elevation on the island out of all the other sampling sites, and thus is subject to increased wind speeds due to reduced surface friction from structures such as buildings at lower elevations
- Three Tarn Nick: Located in the North Plateau. Traps were laid just off the path in this area
- Burnett's Leap: This sampling site is the furthest north on the island, and the closest to Rona. Traps were laid just off the path in this area
- Low Light: Traps were laid off McLeod's Path between Low Light and the arch that leads towards Low Light

- South Horn: Traps were laid near the South Horn on the southern end of the island
- Holyman's Road: Traps were laid just off the path at this location

Named locations mentioned in these descriptions can be seen on an NNR map of the island (National Nature Reserve, 2014). Longworth traps were filled with straw bedding and bird seed and left overnight in locations likely to be frequented by mice, such as along walls and exposed rock, and at the entrances to burrows. Traps were left overnight and checked the following mornings for success. 54 mice were captured in this way during the visit, which were subsequently killed, weighed, sexed and dissected. Mice were euthanised by carbon dioxide inhalation, following which death was confirmed by exsanguination in accordance with Schedule 1 to the Animals (Scientific Procedures) Act 1986. Dissection was carried out using sterilised scissors and forceps in a makeshift lab on the island, with dissection equipment sterilised throughout the process by flaming using ethanol and a Bunsen burner. Relevant testes and livers were removed, weighed and placed in respective labelled test tubes, then stored at -80^{o}C upon return from the island. Other organs such as stomach, caecum and eyes were also removed and stored to the wider benefit of the University of Nottingham parasitology department.

Mice captured on this trip formed part of a larger database of mice provided by the University of Nottingham parasitology department consisting of 366 house mice collected from the Isle of May over a three-year period at various time points since 2013; using the same methods of baiting and capture. These mice were trapped in varying climactic conditions in spring, summer, winter and autumn. Individuals from which tissue was taken for use in this study were randomly selected from the database and included both mice obtained on the October trip and those captured on previous trips to the May by the department.

Other mice used in this study included those from Eday in the Orkney archipelago and Inchkeith, another island in the Firth of Forth located approximately 39km from the Isle of May. Testes of these mice were provided by the University of Nottingham Parasitology department, dissected from mice caught in 2010 and since stored at -80^{o}C. DNA was also extracted from the liver tissue of mice caught from Tollerton, a rural pig farm located in Nottinghamshire, by the department for use as a reference of mice from the British mainland.

3.2 DNA extraction

Suitable DNA was extracted from mostly liver, but some testes tissue of mice from the Isle of May, from liver tissue of mice from Tollerton and from testes tissue of mice from Eday and Inchkeith using a two-component Extracta™ reagent kit (Quanta Biosciences). Tissue samples of an appropriate size (roughly $1mm^2$ for animal tissues) were cut out and fully submerged in 110µl of the extraction reagent, which was heated at 95°C for 30 minutes. Under these conditions mammalian tissues are known to break down, at a much faster rate than conventional proteinase-K extraction procedures. The samples were left to cool to room temperature following which an equal volume of the stabilisation buffer was added and solutions were mixed, then stored at -20°C. These would form the stock solutions of sample DNA from which all subsequent lab work would derive. DNA was also extracted from BALB/c and C57BLK/6 lab mouse tail tissue for use as positive controls in Ctla4 polymerase chain reactions (PCR), relevant to the next step of restriction enzyme digestion in Ctla4 typing experiments.

3.3 Ctla4 PCR

Quality of extracted DNA was periodically confirmed using a set of primers specific to the mouse Ctla4 gene, which encodes a protein receptor, CTLA-4, that downregulates immune responses, for use in a PCR. These primers were used to genotype newly caught mice from the Isle of May based on a single nucleotide polymorphism (SNP) at base 77 of exon 2 of the Ctla4 gene, prior to a restriction enzyme digestion reaction. I carried out a portion of this work myself, typing all the mice caught in the most recent October trip to the island, though this work is not relevant to this project and so is not shown. Having the quality of the DNA verified in this way was primarily useful for comparing efficacy of different Sry primer combinations in amplifying the Sry gene fragment, as 3 sets of forward and reverse primers for the Sry region were available in the lab at the start of my experiments, with previously reported variance in brightness of DNA bands on agarose gels visualised in UV light between different primer combinations. Knowing the quality of the DNA being amplified allowed primer combinations to be evaluated as the

dimness or lack of a band could be linked to the ability of the primers to amplify DNA. From this it was determined which primer set should be used going forward with subsequent PCR and sequencing. The Ctla4 primer set consisted of the following primers (Sigma-Aldrich):

- CTLA4 L (forward): 5'-GAGCTTGCAGGAGTTCATCC-3'
- CTLA4 R (reverse): 5'-GCAGAGGTACAGTCCCGTGT-3'

Reaction mixtures consisted of:

- 12.5µl AccuStart II GelTrack PCR SuperMix (2X) (Quanta BioSciences)
- 1µl forward primer (10µM) (final reaction concentration 400nM)
- 1µl reverse primer (10µM) (final reaction concentration 400nM)
- 8µl nuclease-free water
- 2.5µl extracted DNA

Cycling conditions were as such:

- 94 ^{O}C for 3 minutes
- 35 cycles of
  - 94 ^{O}C for 15 seconds
  - 60 ^{O}C for 15 seconds
  - 72 ^{O}C for 60 seconds
- 4 ^{O}C hold

Included in these PCR reactions were a negative control, which would indicate whether contamination was occurring, and positive controls relevant to the proceeding step of the Ctla4 typing experiment, which would be useful in the event of failure of the extraction procedure of sample DNA, though not necessary otherwise. Positive controls consisted of previously successfully extracted DNA from the tail tissue of BALB/c and C57BLK/6 laboratory mouse strains, kindly provided by my department. Negative controls consisted of reaction mixtures with nuclease-free water in place of template DNA (totalling 10.5µl water per reaction). These polymerase chain reactions, and all the ones proceeding them in this study, were performed in a GS1^{TM} thermal cycler (G-Storm).

3.4 Sry PCR

Sequences of the Sry gene (also described as Sry in the literature) located downstream of the 5' end of the MHG box on the Y chromosome were amplified using primers originally described in Albrecht and Eicher (1997), as well as a pair of primers designed by my department. Those from the Albrecht and Eicher paper (1997) are as follows:

- Sry-8653 (forward): 5'-GGAGTAGAGCTCCACACCTGTACTCG-3'
- Sry-9475 (reverse): 5'-CCAGTGTCATGAGACTGCCAACC-3'
- Sry-8212 (forward): 5'-TTGATTTTTAGTGTTCAGCCCTACAGCC-3'
- Sry-9791 (reverse): 5'-AGCTGTTTGCTGTCTTTGTGCTAGCC-3'

Those designed by my department are as such:

- Sry-7956 (forward): 5'-TGGACTAGGGAGGTCCTGAA-3'
- Sry-9821 (reverse): 5'-CATGGCATGCTGTATTGACC-3'

Reaction mixtures for these reactions were as for the Ctla4 PCR, as per the AccuStart II GelTrack PCR SuperMix (2X) protocol, and consisted of:

- 12.5µl AccuStart II GelTrack PCR SuperMix (2X) (Quanta BioSciences)
- 1µl forward primer (10µM) (final reaction concentration 400nM)
- 1µl reverse primer (10µM) (final reaction concentration 400nM)
- 8µl nuclease-free water
- 2.5µl extracted DNA

Cycling conditions varied depending on the primer combinations, and are visualised below in Table 1. The optimal primer combinations were determined as described in the previous paragraph. Determination of the ideal primer combination was done in two stages, as this project was started prior to the designing of the primers to the Sry gene by my department. Initially, Sry-8653, Sry-9475, Sry-8212 and Sry-9791 (Albrecht and Eicher, 1997) were tested in various combinations, and the project advanced using the ideal primer pair determined from this. Following difficulties with sequencing, primer combinations were re-tested with the new primers, and all further sequencing proceeded using the best determined combination from this.

Table 1 Different combinations of primers specific to the region of the Sry gene downstream of the 5' end of the MHG box. Cycling conditions used for each primer combination are included as are the size of the fragments amplified by them. Where combinations of primers contained primers with different cycling conditions, the conditions with the lower annealing temperature were chosen such that primers had the most chance for both to work. The primer combination that produced the brightest bands of amplified DNA on an agarose gel under UV light, indicating highest concentration, was selected for use in further PCR and sequencing.

Primer combination	Pairing	Initial denaturation	Cycling stage: denaturation	Cycling stage: annealing	Cycling stage: elongation	Final elongation	No. of cycles	Fragment length
Sry-8653 + Sry-9475	Original set	94 °C 4 mins	94 °C 30 secs	64 °C 30 secs	72 °C 90 secs	72 °C 10 mins	35	845 bp
Sry-8212 + Sry-9791	Original set	94 °C 4 mins	94 °C 30 secs	64 °C 30 secs	72 °C 90 secs	72 °C 10 mins	35	1605 bp
Sry-7956 + Sry-9821	Original set	94 °C 3 mins	94 °C 30 secs	58 °C 30 secs	72 °C 105 secs	72 °C 10 mins	35	1885 bp
Sry-8653 + Sry-9791	New combination	94 °C 4 mins	94 °C 30 secs	64 °C 30 secs	72 °C 90 secs	72 °C 10 mins	35	1164 bp
Sry-8653 + Sry-9821	New combination	94 °C 3 mins	94 °C 30 secs	58 °C 30 secs	72 °C 105 secs	72 °C 10 mins	35	1188 bp
Sry-8212 + Sry-9475	New combination	94 °C 4 mins	94 °C 30 secs	64 °C 30 secs	72 °C 90 secs	72 °C 10 mins	35	1286 bp
Sry-8212 + Sry-9821	New combination	94 °C 3 mins	94 °C 30 secs	58 °C 30 secs	72 °C 105 secs	72 °C 10 mins	35	1629 bp
Sry-7956 + Sry-9475	New combination	94 °C 3 mins	94 °C 30 secs	58 °C 30 secs	72 °C 105 secs	72 °C 10 mins	35	1542 bp
Sry-7956 + Sry-9791	New combination	94 °C 3 mins	94 °C 30 secs	58 °C 30 secs	72 °C 105 secs	72 °C 10 mins	35	1861 bp

3.5 D-Loop PCR

A set of primers specific to the mitochondrial D-loop region were used to generate a partial D-loop fragment as described by Searle et al. (2009) prior to sequencing. These amplified a fragment 1397bp in length. Primers were as follows:

- L831 (forward): 5'-ATACGCCATTCTACGCTCAA-3'
- H2228 (reverse): 5'-TTATAAGGCCAGGACCAAAC-3'

Reaction mixtures again consisted of:

- 12.5µl AccuStart II GelTrack PCR SuperMix (2X) (Quanta BioSciences)
- 1µl forward primer (10µM) (final reaction concentration 400nM)
- 1µl reverse primer (10µM) (final reaction concentration 400nM)
- 8µl nuclease-free water
- 2.5µl extracted DNA

Cycling conditions were as follows:

- 93 ^{O}C for 5 minutes
- 30 cycles of
  - 93 ^{O}C for 60 seconds
  - 50 ^{O}C for 60 seconds
  - 72 ^{O}C for 90 seconds
- 72 ^{O}C for 10 minutes
- 4 ^{O}C hold

This was primarily done to compare technically how well mitochondrial D-loop DNA would sequence against Y-chromosome DNA from the Sry region. With good quality sequences previously returned from the fragments amplified by these primers by others in my department, it would serve to familiarise myself with sequencing experiments. From this I would be better equipped to deal with problems arising from attempts to obtain Sry fragment sequences, as had been encountered prior to my experiments in my department with the Sry-8653 and Sry-

9475 primers. 18 mice were sequenced from the Isle of May, as were 4 from Eday, 6 from Inchkeith and 5 from Tollerton.

3.6 Gel electrophoresis

Following PCR amplification, DNA fragments were visualised by way of agarose gel electrophoresis. Tris-acetate-EDTA (TAE) buffer was necessary for electrophoresis reactions and for making gels. Concentrated 50X TAE buffer was made 1000ml at a time. This was done by dissolving 242g Tris-base (FW 121.14) in 750mL of deionized water, then adding 57.1mL acetic acid and 100mL of 0.5M ethylenediaminetetraacetic acid (EDTA) (pH 8.0). A working solution of 1X TAE buffer was made from this by diluting the stock solution by 50X in deionized water. This was typically done 1000mL at a time, adding 20mL of 50X solution to 980mL of water. Gels were made to be either 1% or 1.5%. These were made by placing 1.0g or 1.5g, respectively, of agarose in 100µl of 1X TAE electrophoresis buffer. This was heated in a microwave until boiling and 8µl of ethidium bromide were added, mixed by swirling then left to set and cool for 25-30 minutes prior to the electrophoresis procedure. Samples were viewed next to a 100bp ladder (New England BioLabs), with the exception of one gel that was run with a 1kb ladder (New England BioLabs) due to no 100bp ladder stock being available at the time. Gels were run between 110V and 120V for 30-40 minutes then viewed in a Gel Doc™ XR imager (BioRad).

3.7 Ethanol Precipitation

Ethanol precipitation was employed to concentrate DNA for sequencing where failure of Sry sequencing attempts was attributable to low concentration of PCR products. Reactions were pooled following PCR amplification with the number of replicates for each sample proportional to the degree to which they needed to be concentrated, as was learned with progressive experiments. This was determined through reference to the brightness of the 100bp DNA ladder, such that a DNA band with roughly half the brightness of the brightest band in the DNA ladder would be concentrated to twice its original concentration by pooling two replicates prior to ethanol precipitation. 10µl of 3M sodium acetate and 250µl of cold 100% ethanol

were proportionally required per 100µl of supernatant of pooled PCR product. Solutions were mixed by inversion 10 times, then kept in a -20°C freezer for 10 minutes prior to centrifugation, which was performed at 13,000rpm for 10 minutes at 4°C. Supernatant was removed, leaving a "pellet" of desired sample DNA, to which 100µl of 70% ethanol was added prior to a second centrifugation step, which consisted of another 10 minutes of 13,000rpm at 4°C. Supernatant was again removed, following which samples were left on ice with their tube lids kept open to be air dried, then re-suspended in an appropriate volume of nuclease-free water which would comprise the final solution to be sequenced. Samples were again checked on a 1% agarose gel to confirm that the DNA had been concentrated to the desired degree for sequencing, which was run at 100V for 30-40 minutes, using 5µl of the concentrated product mixed with roughly 3µl of purple gel loading dye (6X) (New England BioLabs).

3.8 Gel Extraction

Where samples were producing multiple bands, a gel extraction procedure was employed in order to isolate the fragment of desired length for sequencing. Problematic samples were first re-run on a 1% agarose gel at 100V for 30-40 minutes, with 20µl of (ethanol precipitated) PCR product added to 4µl loading dye, which was then placed in a UV transilluminator to visualise bands for excision. Clean, sharp, disposable scalpel blades were used to cut out the desired fragment(s) from any one gel, with length and intensity of exposure to UV light limited as much as was possible so as to minimise damage to DNA. Scalpel blades were changed between excisions, where multiple samples occupied a gel. Appropriate precautions were taken to ensure protection from UV light. A QIAquick™ Gel Extraction Kit (Qiagen) was then used to purify DNA fragments from the gel slices, following the associated protocol. Products of this procedure were then checked again on a 1% agarose gel, run at 100V for 30-40 minutes using 5 µl of gel extraction product mixed with 2.5 µl loading dye and visualised under UV light, to determine final concentration of samples, and to what degree to dilute them, if any, for sequencing.

3.9 Sequencing

Final reaction products of desired concentration were sent to Source Bioscience UK Limited who provide a Sanger sequencing service (www.sourcebioscience.com). This involved preparing the samples prior to submitting them. 5µl of PCR product with a DNA concentration of 10ng/µl was required per primer. 2 sequencing reactions, generating sequences from the forward primer and from the reverse primer, were needed per sample. Samples were first purified of excess primers and nucleotides by adding a proportionate volume of ExoSAP-IT™ PCR product cleanup reagent (ThermoFisher Scientific), with 2µl added per 5µl of PCR product. Samples were first incubated at 37^{o}C for 15 minutes then heated to 80^{o}C for 15 minutes to inactivate the cleanup enzyme. 5µl of each primer per reaction was required, with primers needing to be sent at 3.2pmol/ul. Diluted primers and purified samples were then sent to Source BioScience for sequencing. Forward and reverse sequences were spliced together based on where they overlapped to generate single contiguous sequences. Sequences were aligned and edited using BioEdit v7.2.5 (Hall, 1999).

3.10 Analysis of D-loop data

A total of 34 D-loop sequence alignments were available of mice from various locations consisting of the Isle of May, Eday, Inchkeith, and Tollerton, a pig farm on the British mainland, having been mostly obtained prior to this project by others in my department, and added to with the D-loop sequence I obtained and aligned using BioEdit. Mitochondrial sequences were varying lengths between positions 14971 and 16356, with reference to the published mouse mtDNA sequences of Bibb et al. (1981). These were all truncated to a standard length of 895bp between positions 15039 and 15934, for alignment purposes. SNPs between these sequences were distinguished, and these were used to draw a network of D-loop haplotypes based on SNP differences identified between the mice using NETWORK 5.0 (fluxus-engineering.com).

3.11 Analysis of immune SNPs

With the limited information gained from sequencing of the aforementioned Sry region, investigations were focused away from paternal inheritance and remaining time was allocated to determining population structure on the island. A dataset was available of mice from the Isle of May database genotyped for SNPs at 32 immune loci, kindly provided by others in my department. Each of these immune loci were characterised by di-allelic, polymorphic SNPs. The loci used and the SNPs associated with them are outlined in Table 2. At one of the loci, Ifngr1_1, all the mice in the sample were homozygous for the T-allele, and as such the alternative allele for this locus is not known.

The 32 immune loci are located on 16 different chromosomes. With the exception of 6 of the loci, each immune locus was on the same chromosome with at least one other locus. As such, pairwise linkage disequilibrium between SNPs was calculated and plotted using Haploview (Barrett et al., 2005). For simplicity, all SNPs were included in the same plot, though many of them would certainly not show linkage with one another.

Various statistical analyses were performed on these data to identify any sub-structuring within the Isle of May population. The final version of the dataset, where mice for which trapping location information was not available were excluded, consisted of 115 mice typed at 32 immune loci. The data file was initially converted for use in various population genetics software using CONVERT (Glaubitz, 2004). The dataset was converted into files for inputting into either STRUCTURE v2.3.4 (Pritchard et al., 2000), ARLEQUIN v3.5.2.2 (Excoffier and Lischer, 2010) or GENEPOP v4.6.9 (Raymond and Rousset, 1995; Rousset, 2008).

STRUCTURE was used to process the genetic data using Bayesian clustering to determine underlying population structuring. STRUCTURE was run with the parameters of a burn-in period of 10,000 iterations and 50,000 Markov chain Monte Carlo (MCMC) repetitions, testing K values between 1 and 10, where K is the optimal number of groups (sub-populations), with at least 50 runs for each value of K. The method of Evanno et al. (2005) was subsequently employed whereby the true number of populations was inferred via the second order rate of change of the

likelihood, given as delta K (ΔK), calculated from averaged results over 50 runs for each value of K.

Departure of each locus in the population from Hardy-Weinberg equilibrium was tested in GENEPOP by an estimation of exact P-values using the Markov chain method. Parameters of this testing consisted of a dememorization period of 10,000 generations, 100 batches and 100 iterations per batch.

For further analysis of isolation by distance (IBD), individuals from the dataset were considered as 10 populations (Table 3), based on the rough location on the island where they were trapped (Figure 1), in order to test for population substructuring. Coordinates were obtained for trapping sites by selecting appropriately determined points on an online map of the Isle of May (Google Maps, 2017). Extensive knowledge of the area from my time spent on it and of others in my department as well as reference to an NNR map of the island (National Nature Reserve, 2014) directed me to where the points should have been placed for the various trapping site locations. Populations containing less than 5 individuals were not included. IBD between populations was calculated to determine whether groups of mice separated by longer distances were more genetically differentiated from each other. Pairwise Fst values (Weir and Cockerham, 1984) were calculated between the populations using ARLEQUIN, and linearised according to Slatkin (1995) to give genetic distance values. A distance matrix was calculated based on the coordinates of the group locations. A graph representing IBD was created from the data using R v3.2.5 (R Core Development Team, 2008), which also gave a value for regression. A Mantel test of significance between the matrices was also performed using ARLEQUIN, with 9999 permutations. Finally, the genetic structure of the population was investigated by constructing a phylogenetic network of concatenated SNPs of the individuals; this would highlight any obvious and significant non-random mating among the mice. This was built with SplitsTree4 (Huson and Bryant, 2006), using the NeighbourNet algorithm (Bryant and Moulton, 2004).

Table 2 Di-allelic Immune loci in *Mus domesticus* used for population genetics analysis and their associated single nucleotide polymorphisms. The alternative allele for the SNP at the Ifngr1_1 locus is not known as all mice in the dataset were homozygous for the T allele at this locus

Immune locus	SNP
Il17a_1 (Interleukin 17A)	A/G
Cd27_1 (CD27 antigen)	C/T
Ctla4_1 (Cytotoxic T-lymphocyte-associated protein 4)	A/G
GATA3_2 (GATA binding protein 3)	C/T
Gmza_19 (Gene not known)	A/T
Ifnar2_2 (Interferon [alpha and beta] receptor 2)	A/G
Ifngr1_1 (Interferon gamma receptor 1)	T/-
Il10_1 (Interleukin 10)	A/G
Il12a_2 (Interleukin 12A)	A/G
Il12b_1 (Interleukin 12B)	C/T
Il16_3 (Interleukin 16)	C/T
Il1r1_1 (Interleukin 1 receptor, type I)	C/T
Il20ra_1 (Interleukin 20 receptor, alpha)	C/T
Il25_2 (Interleukin 25)	C/T
Il2ra_1 (Interleukin 2 receptor, alpha chain)	C/T
Il7r_1 (Interleukin 7 receptor)	A/C
Il9_10 (Interleukin 9)	C/T
Irf7_3 (Interferon regulatory factor 7)	C/T
Irf9_1 (Interferon regulatory factor 9)	C/T
Mr1_2 (Major histocompatibility complex, class I-related)	C/T
Myd88_1 (Myeloid differentiation primary response gene 88)	C/T
Nod2_3 (Nucleotide-binding oligomerization domain containing 2)	C/T
Rorc_1 (RAR-related orphan receptor gamma)	A/G
Socs5_1 (Suppressor of cytokine signalling 5)	A/T
Socs7_2 (Suppressor of cytokine signalling 7)	A/G
Tirap_1 (Toll-interleukin 1 receptor [TIR] domain-containing adaptor protein)	A/G
Tlr3_3 (Toll-like receptor 3)	C/T
Tlr4_1 (Toll-like receptor 4)	A/G
Tlr5_1 (Toll-like receptor 5)	C/T
Tnf_1 (Tumour necrosis factor)	C/T
Traf3_2 (TNF receptor-associated factor 3)	C/T
Tyk2_1 (Tyrosine kinase 2)	C/T

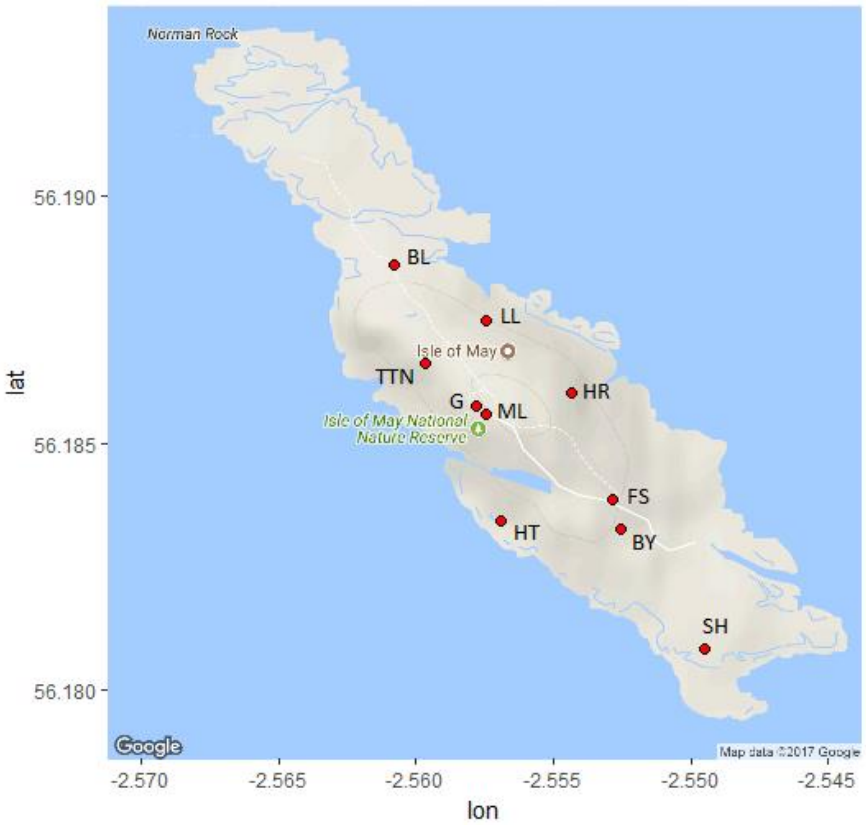


Figure 1 Trapping locations considered for isolation by distance analysis, indicated by red circles superimposed on a map of the Isle of May. Geographic distances between these coordinates were calculated and used for IBD analysis. Locations are a mark of the general areas of the island where groups of traps were laid in successive years. Though there were more areas where groups of traps had been set in visits to the island, these were not relevant to the mice in the immune SNP dataset

Table 3 Groups of mice from the immune SNP dataset considered as "populations" based on the areas where they were trapped. Areas in the dataset containing less than five mice were excluded

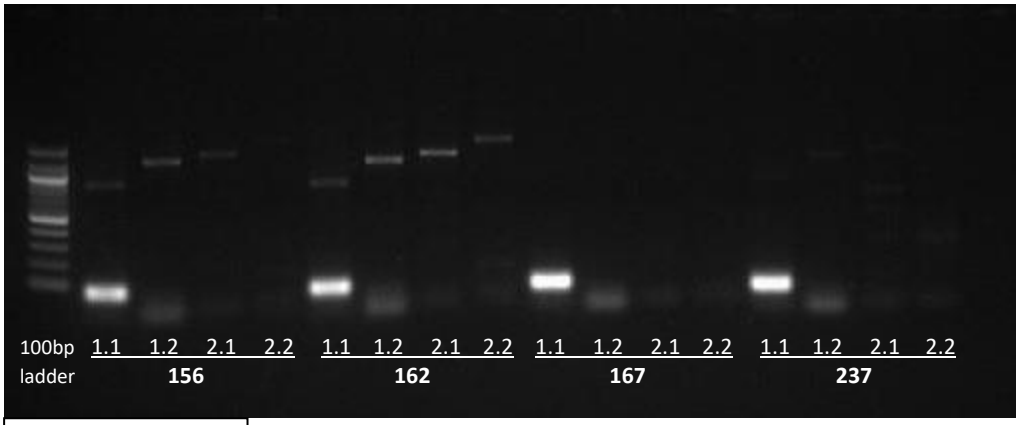
Group	Area Coordinates	No. of Mice
South Horn (SH)	56.180839, -2.549472	19
Byres (BY)	56.183279, -2.552562	5
High Tarn (HT)	56.183436, -2.556917	14
Fluke Street (FS)	56.183911, -2.552835	27
Main Light (ML)	56.185639, -2.557417	6
Garden (G)	56.185754, -2.557781	5
Holyman's Road (HR)	56.186068, -2.554319	5
Three Tarn Nick (TTN)	56.186638, -2.559667	13
Low Light (LL)	56.187463, -2.557451	6
Burnett's Leap (BL)	56.188614, -2.560814	15

4. Results

4.1 Sequencing of the *Sry* region

Comparison of primer combinations first determined that the ideal primer combination for amplifying the desired region of the Sry gene was the Sry-8653 forward primer and Sry-9791 reverse primer (Albrecht and Eicher, 1997) (Figure 2). These were selected for use in further PCR and sequencing procedures. Viability of extracted DNA was confirmed by a Ctla-4 PCR (Figure 3), determining that dimness of DNA bands, as visualised under UV light, was not due to the extraction procedure.

Further PCR reactions highlighted stark differences between the Sry and D-loop primers in their amplification of DNA from the Isle of May mice (Figure 4). In particular, D-loop primers amplified DNA of consistent quality across samples that was of relatively high concentration, while quality of Sry fragments appeared to be largely sample-dependent, with primers generally amplifying DNA of much lower concentration, as indicated by dimmer bands. Apparent lack of differences in consistency between D-loop samples may have been due to the fact that DNA bands on agarose gels are harder to distinguish at higher concentrations, however differences in concentration of amplified DNA between Sry and D-loop primers were irrefutable. Attempts to sequence PCR products of the samples producing the two brightest bands from the Sry PCR, as well as two of the D-loop samples from the same mice, returned an error email from Source Bioscience indicating that all sequencing reactions had failed due to low signal strength. Possible reasons given for failure included concentrations of submitted samples being either too low or too high. It was determined that D-loop PCR products were too concentrated while the Sry products were not concentrated enough, and



Key:

1.1: Sry-8653 + Sry-9475
1.2: Sry-8653 + Sry-9791
2.1: Sry-8212 + Sry-9475
2.2: Sry-8212 + Sry-9791

Figure 2 PCR of Sry primer combinations on a 1.5% agarose gel, with 15µl PCR product of each reaction. Combinations were tested on liver DNA from 3 male mice (156, 162, 237) and one female (167) from the Isle of May. Mice are numbered according to their position in the department database. The female mouse was included by accident; it was not known to be a female at the time. DNA bands are dim, and band brightness varies between individuals but combination 1.2 (key) is seen to produce the strongest bands overall.

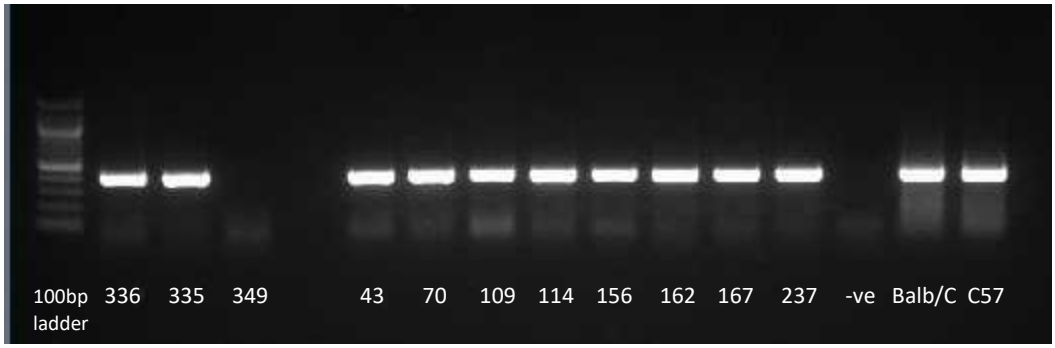


Figure 3 Ctla4 PCR of Isle of May mice on 1.5% agarose gel, with 10µl PCR product of each sample. DNA was extracted from mouse livers. Included are positive controls of previously extracted DNA from laboratory mouse strains (Balb/C and C57) for use in proceeding restriction reaction, from which newly caught mice (336, 335, 349) would be typed as either A or G. Balb/C and C57 mice act as reference as they type G and A upon restriction, respectively, producing bands of different length. A negative control of nuclease-free water ensures no contamination is occurring. Ctla4 fragments of the desired length are sufficiently amplified for most mice, as indicated by clear bands. One of the reactions (349) is seen to have failed, possibly due to a failure of the extraction procedure.

as such, ethanol precipitation was employed to concentrate Sry samples to twice their original concentration, while D-loop samples were diluted 1:2 in nuclease-free water. Electrophoresis elucidates the effect of ethanol precipitation on

concentration of amplified Sry fragments, represented by an increase in band brightness in Figure 5. Submission of diluted D-loop and concentrated Sry samples subsequently returned sequences of good quality, judged by numbers of misreads and length of forward and reverse sequences.

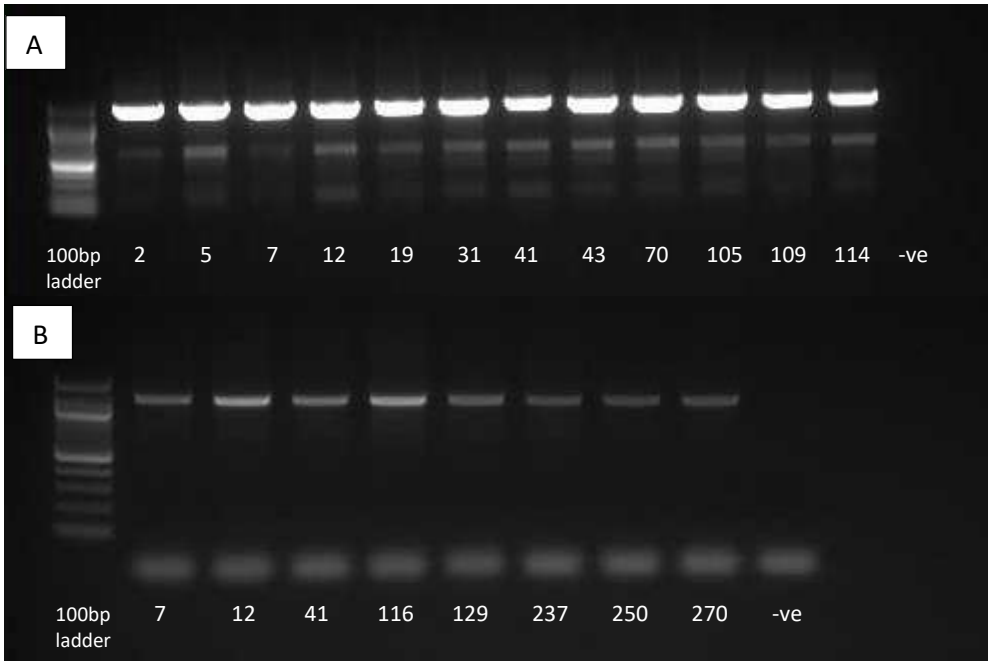


Figure 4 Sry and D-loop PCRs for a number of Isle of May mice, including a negative control of nuclease-free water. DNA fragments of the expected size are seen to have been successfully amplified in all samples. DNA was extracted from mouse livers. **A)** D-loop PCR on a 1.5% agarose gel with 10µl of PCR product of each sample. Amplified DNA for all samples is seen to be of high concentration as indicated by bright bands. **B)** Sry PCR on a 1.5% agarose gel using 15µl of PCR product of each sample. Bands are seen to be of varying brightness between samples.

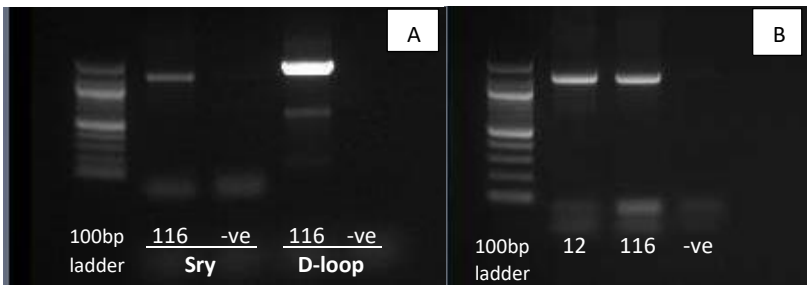


Figure 5 Sry PCR product of mouse 116 is seen before and after ethanol precipitation to twice its original concentration. **A)** PCR of DNA extracted from mouse 116 amplified by Sry and D-loop primers each with a negative control prior to concentration of the Sry PCR product, run on a 1% agarose gel. 12µl of Sry PCR product is used while 10µl of D-loop product is used. Differences between the primer sets are seen by way of band brightness. **B)** Sry PCR of samples 12 and 116 following ethanol precipitation, using 5µl PCR product of each sample on a 1.5% agarose gel.

This strategy was employed for 11 more Isle of May mice, with final samples diluted based on brightness of DNA bands post-ethanol precipitation to four times their original concentration on an agarose gel (Figure 5?), returning sequences of varying length and quality. Reverse sequences for two of the mice failed, and some other forward and reverse sequences were not long enough to overlap, giving incomplete sequences with many misreads. Polymorphism was not able to be confirmed as nucleotides that differed from the majority of other sequences would usually be flanked by stretches of misreads in their sequences, making them more likely than not to be misreads themselves. In a further PCR of 4 Eday and 6 Inchkeith, Eday mice revealed a similar pattern of varying, low concentration DNA bands, while Inchkeith mice showed either very dim bands or no bands at all (Figure 6). Concentration of Eday mice PCR products and visualisation on a gel revealed bright bands below the bands of expected size (Figure 7), and while these were initially taken to be primer dimers due to their position on the gel, treatment with ExoSAP-IT™ cleanup reagent did not eliminate the bands when viewed again on a gel (Figure 7). Furthermore, sequencing returned unfamiliar sequences, and indicated that mispriming was occurring. Gel extraction was employed to extract the fragments of desired length for sequencing, and this was successful for the single Eday mouse sample it was initially tested on, which returned a good sequence.

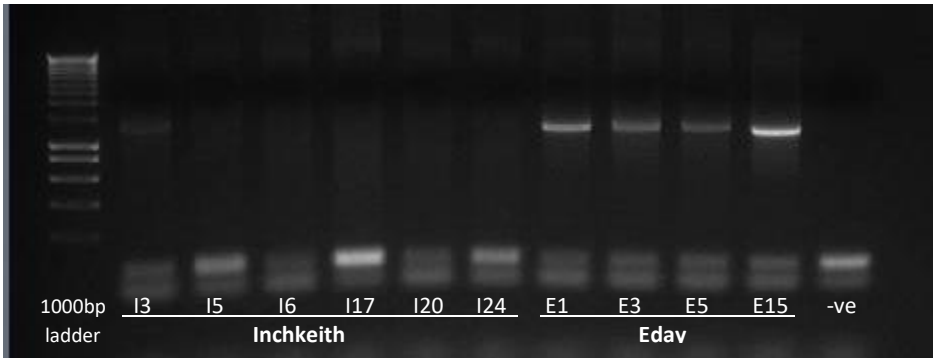


Figure 6 Sry PCR of Inchkeith and Eday samples. A faint band is seen at the correct size for one of the Inchkeith mice, no bands are seen for the other Inchkeith samples. Bands are seen for all the Eday samples though they vary in brightness, and thus concentration.

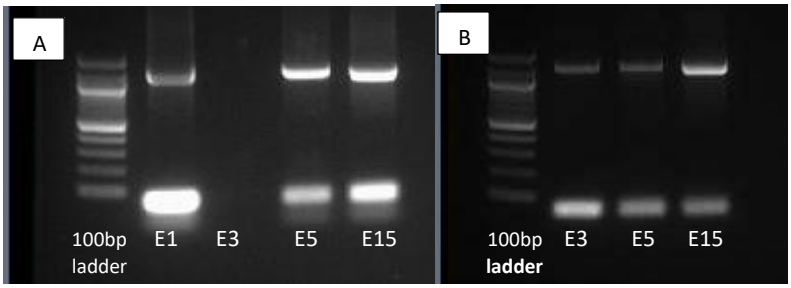


Figure 7 Sry PCR of ethanol precipitated Eday samples, at four times their initial concentration, before and after ExoSAP-IT™ treatment. **A)** Before treatment; bright bands are seen below the bands of expected size, one of which seems to be proportionally dimmer as a result. No band is seen for sample E3 as it failed to sink into the gel, due to not being mixed with enough loading dye **B)** After treatment; ExoSAP-IT™ reagent has failed to get rid of bands, indicating that they are due to mispriming, rather than excess primers. Submission of sample E15 after it had been re-concentrated and gel extracted returned a good sequence.

Less success was had for the other Eday mice, who returned sequences that were either too short or failed outright, a result that was initially determined to be an issue of concentration; sequencing of the Sry fragments appeared to be very sensitive to concentration. Extensive tweaking of concentration of the Eday samples in further experiments produced varied results, and ultimately did not produce complete sequences, mostly due to shortness of reverse sequences, such that concentration was no longer determined to be the primary issue. At this point primers were re-evaluated and a second round of testing of primer combinations was carried out, incorporating the new Sry primers, first tested on an Inchkeith sample, to determine which would be viable for obtaining Inchkeith sequences, and

then on an Eday sample, respectively (Figure 8). Comparison here highlighted the Sry-8653 forward primer and the Sry-9475 combination to be the most effective on both Inchkeith and Eday mice, based on band brightness, and thus experiments proceeded with this combination instead. These primers, amplifying a shorter fragment than the previous combination, successfully returned sequences of high quality with minimal misreads for all Eday and Inchkeith mice, including 6 additional Eday mice. They were also tested on Isle of May samples, and despite small areas of misreads at either end of the sequences, sequences were mostly free of misreads along their length, considerably better than for the previous primers, possibly as shorter fragments are easier to sequence. A total of 44 sequences were obtained with the final primer combination, consisting of 5 mice from Inchkeith, 10 from Eday, and 29 from the Isle of May. An attempt was also made to obtain mice sequences from Tollerton, a pig farm in Nottinghamshire, to represent a mainland mouse type, though DNA extracted from the livers of five of these mice failed to produce any visible bands in an initial PCR, and further attempts to obtain sequences from them would not have been feasible within the timeframe of this project.

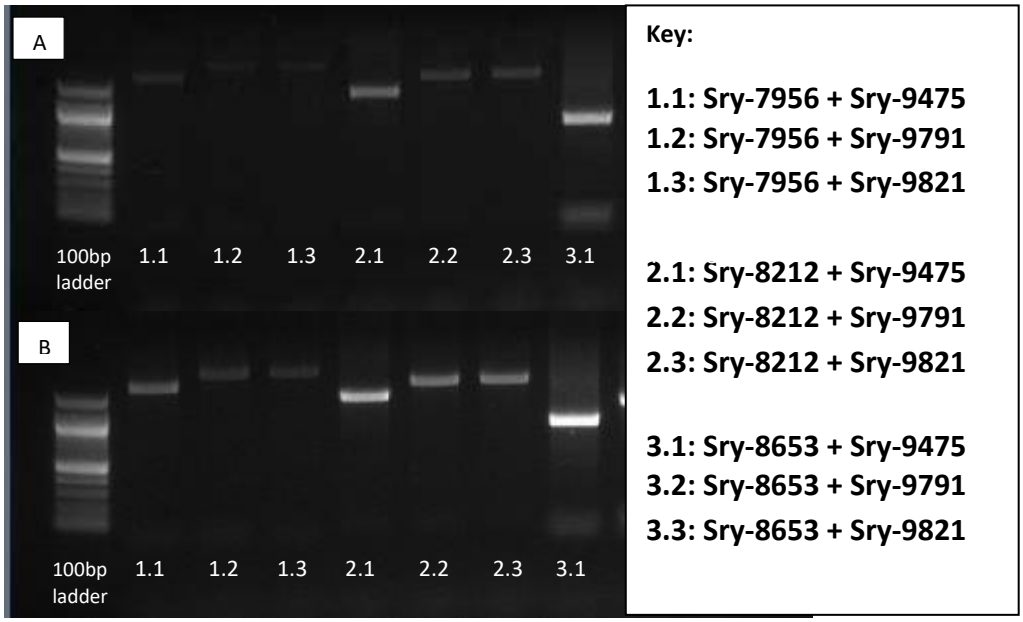


Figure 8 PCRs of Sry primer combinations, tested on an Inchkeith sample and an Eday sample. Viewed on 1% agarose gels. **A)** Primer combinations amplifying DNA from an Inchkeith mouse **B)** Primer combinations tested on an Eday sample

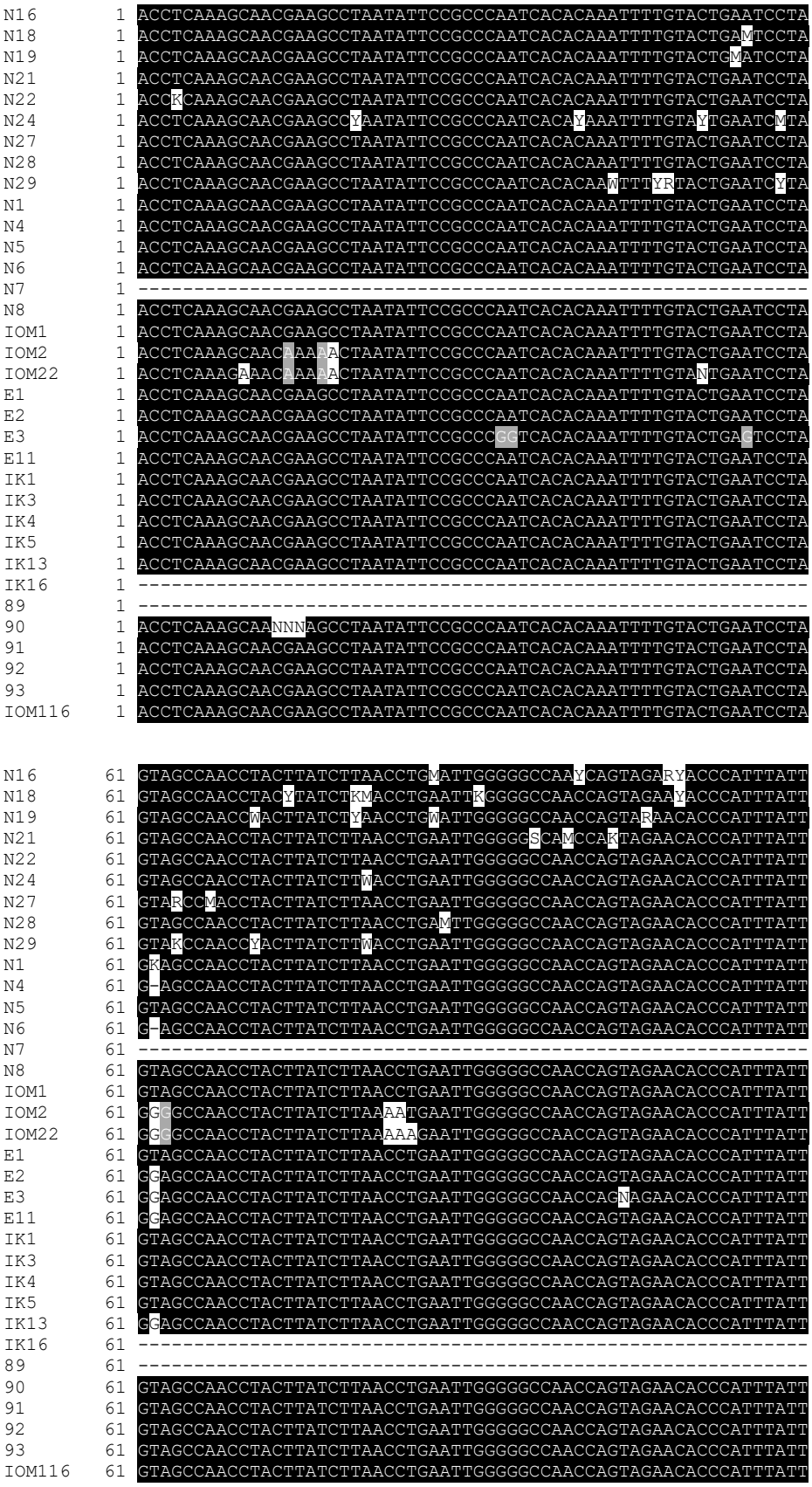


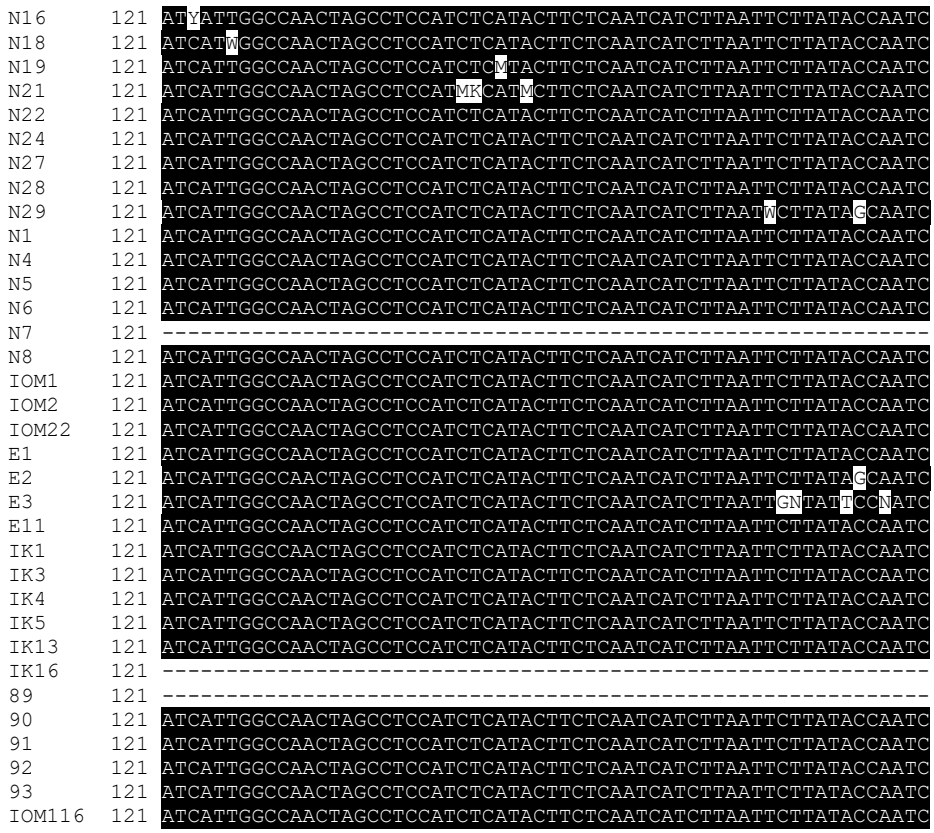
Figure 9 (CAG)n polymorphism identified between positions 8842 and 8846 on the Sry gene, the start of which is indicated by an arrow. Sequences of mice from the same locations were all monotypic, and so sequences are grouped by location.

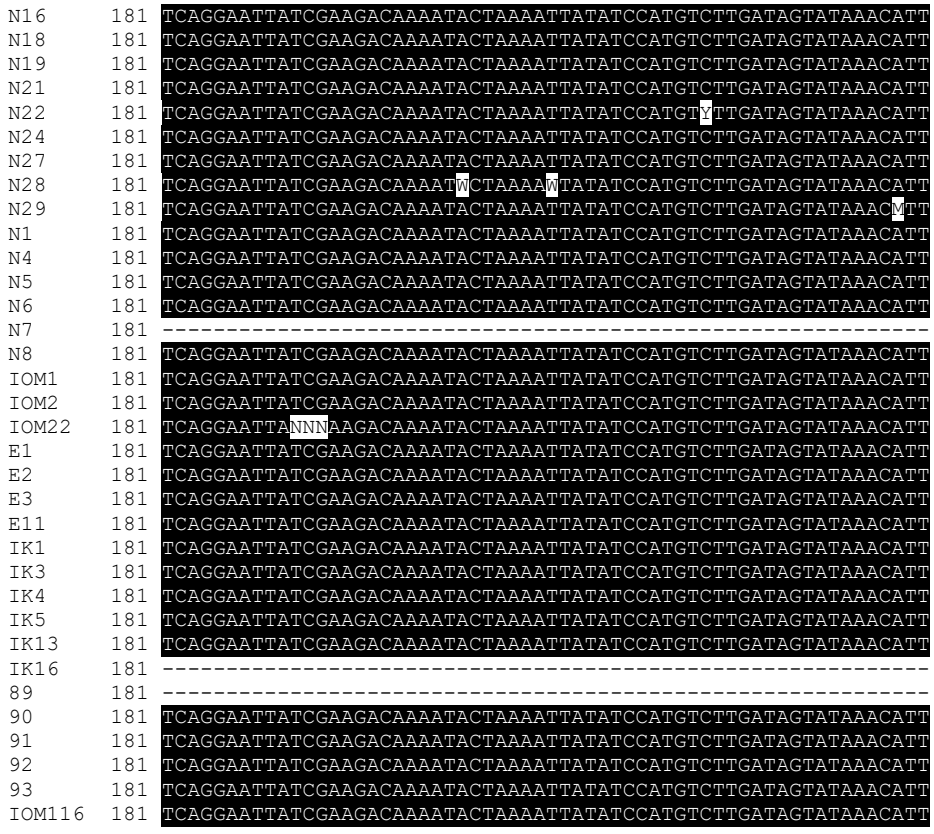
Alignment of sequences revealed just a single difference between the mice of different locations, which was a CAG trinucleotide repeat number variation starting at position 8843 (Figure 9). Specifically, Inchkeith mice had one less CAG than did the mice from Eday and the Isle of May. All sequences were otherwise identical.

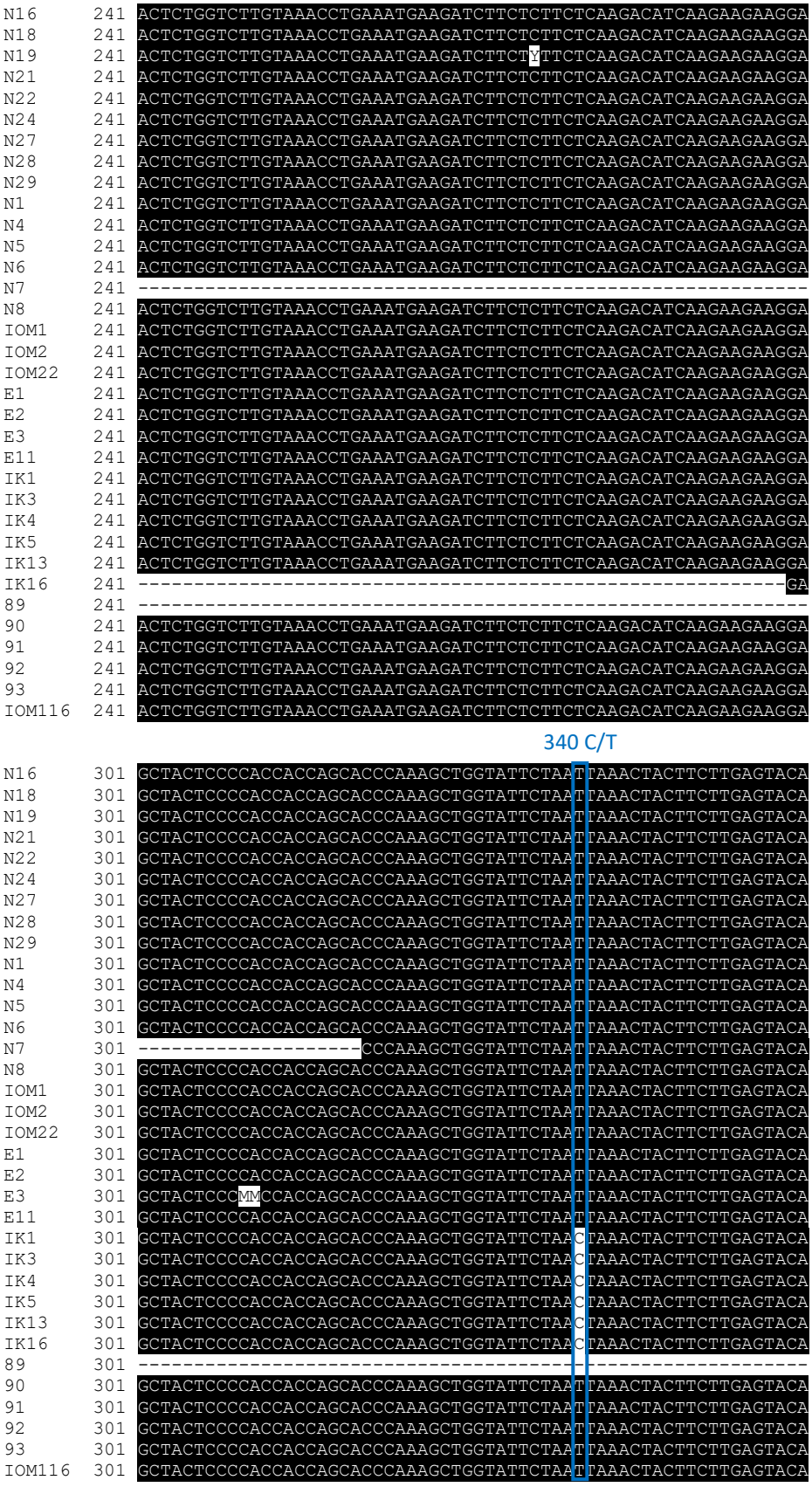
4.2 Analysis of D-loop data

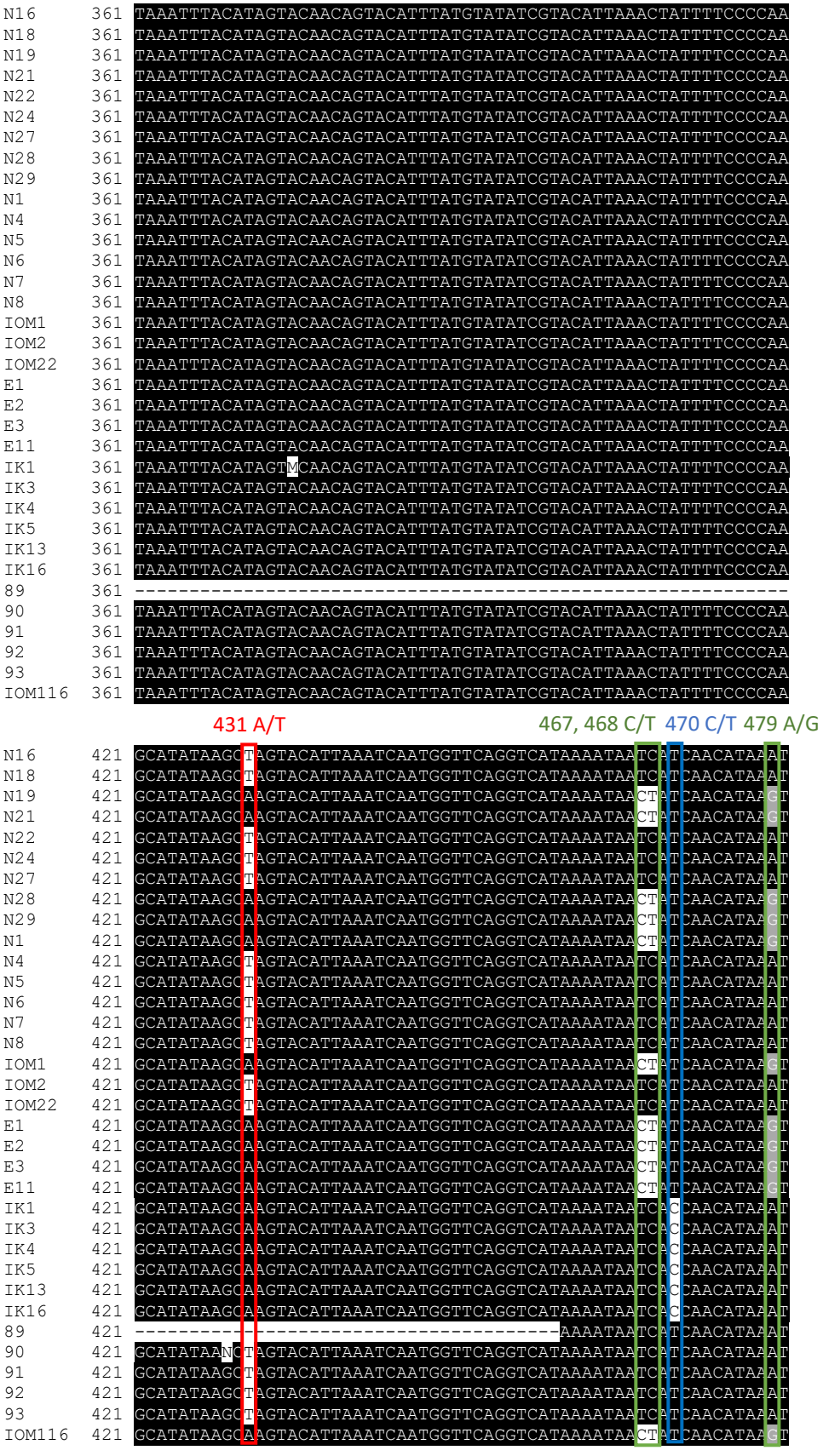
The number of SNPs identified between the D-loop sequences of the various location samples was sufficient to differentiate between mice of different maternal inheritance. 14 SNPs were identified, indicating 4 distinct haplotypes among the mice of all locations. The D-loop sequences and distinguishing SNPs characterising the different haplotypes are shown in Figure 10. Mice on the Isle of May divide into two main groups; those that have the same mitochondrial haplotype as Eday mice, and those of a separate, distinct, mitochondrial haplotype considered to be the historical May haplotype (Figure 11). Inchkeith mice form their own separate haplotype, as do Tollerton. The main Tollerton haplotype is separated from the May haplotype by a single mutational SNP, though one of the Tollerton mice is separated by two mutations. One of the May mice also diverges from the main historical May haplotype by a SNP. Inchkeith and Eday haplotypes are separated from the May haplotype by three shared SNPs, then diverge from each other and from the May by four SNPs each.

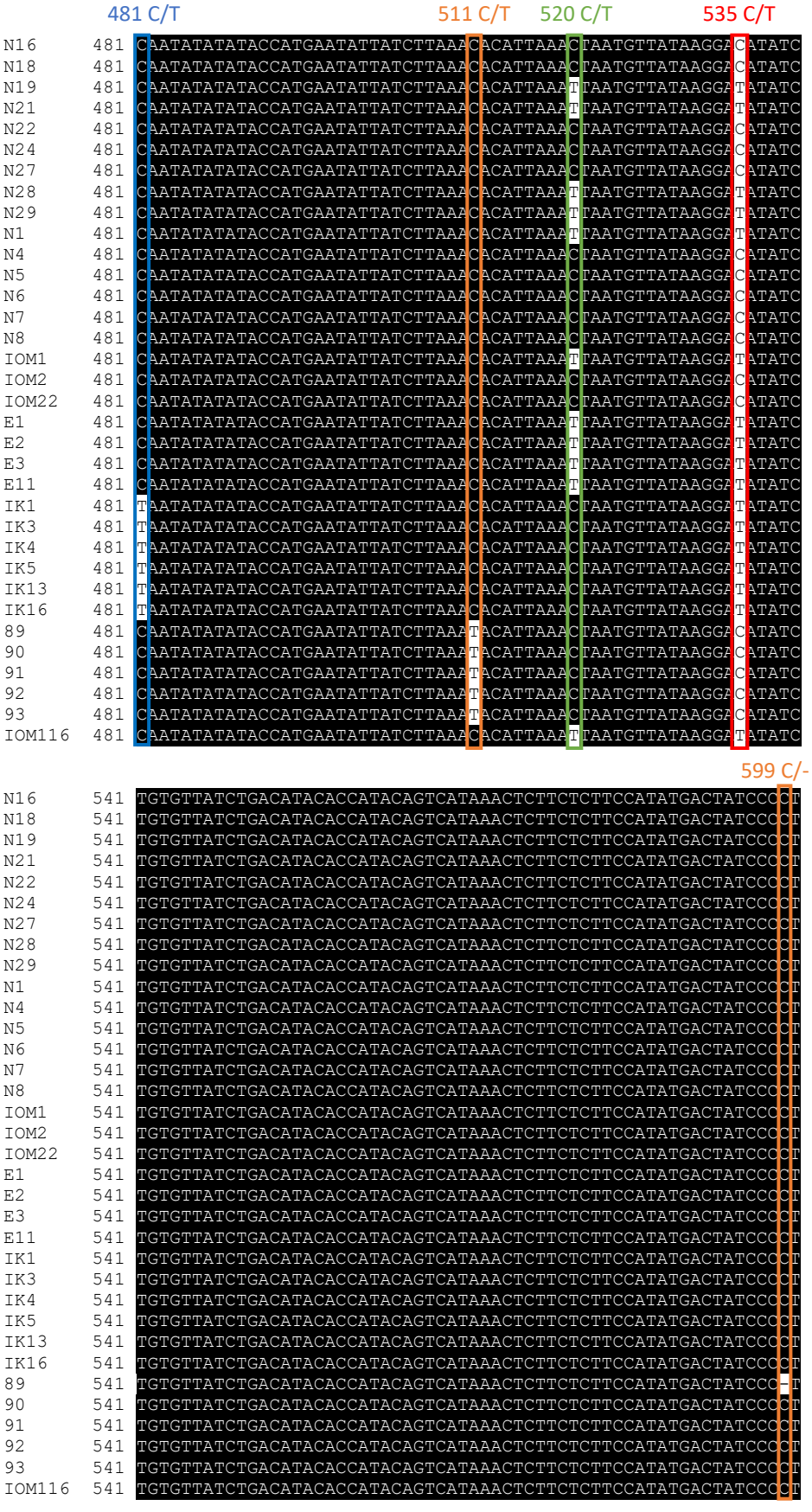

N16 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N18 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAMTCCTA
N19 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGMATCCTA
N21 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N22 1 ACCKCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N24 1 ACCTCAAAGCAACGAAGCCYAATATTCCGCCCAATCACAYAAATTTTGTAYTGAATCMTA
N27 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N28 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N29 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAWTTTYRTACTGAATCYTA
N1 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N4 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N5 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N6 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N7 1 ----------
N8 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IOM1 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IOM2 1 ACCTCAAAGCAACAAAAACTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IOM22 1 ACCTCAAAGAACAAAAACTAATATTCCGCCCAATCACACAAATTTTGTANTGAATCCTA
E1 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
E2 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
E3 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCGGTCACACAAATTTTGTACTGAGTCCTA
E11 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IK1 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IK3 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IK4 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IK5 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IK13 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IK16 1 ----------
89 1 ----------
90 1 ACCTCAAAGCAANNNAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
91 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
92 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
93 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
IOM116 1 ACCTCAAAGCAACGAAGCCTAATATTCCGCCCAATCACACAAATTTTGTACTGAATCCTA
N16 61 GTAGCCAACCTACTTATCTTAACCTGMATTGGGGGCCAAYCAGTAGARYACCCATTTATT
N18 61 GTAGCCAACCTACYTATCTKMACCTGAATTKGGGGCCAACCAGTAGAAYACCCATTTATT
N19 61 GTAGCCAACCWACTTATCTYAACCTGWATTGGGGGCCAACCAGTARAACACCCATTTATT
N21 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGSCAMCCAKTAGAACACCCATTTATT
N22 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N24 61 GTAGCCAACCTACTTATCTTWACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N27 61 GTARCCMACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N28 61 GTAGCCAACCTACTTATCTTAACCTGAMTTGGGGGCCAACCAGTAGAACACCCATTTATT
N29 61 GTAKCCAACCYACTTATCTTWACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N1 61 GKAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N4 61 G-AGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N5 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N6 61 G-AGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
N7 61 ----------
N8 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IOM1 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IOM2 61 GGGGCCAACCTACTTATCTTAAAATGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IOM22 61 GGGGCCAACCTACTTATCTTAAAAAGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
E1 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
E2 61 GGAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
E3 61 GGAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGNAGAACACCCATTTATT
E11 61 GGAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IK1 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IK3 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IK4 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IK5 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IK13 61 GGAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IK16 61 ----------
89 61 ----------
90 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
91 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
92 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
93 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT
IOM116 61 GTAGCCAACCTACTTATCTTAACCTGAATTGGGGGCCAACCAGTAGAACACCCATTTATT


N16 121 ATYATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N18 121 ATCATNGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N19 121 ATCATTGGCCAACTAGCCTCCATCTCMTACTTCTCAATCATCTTAATTCTTATACCAATC
N21 121 ATCATTGGCCAACTAGCCTCCATMKCATMCTTCTCAATCATCTTAATTCTTATACCAATC
N22 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N24 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N27 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N28 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N29 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATWCTTATAGCAATC
N1 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N4 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N5 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N6 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
N7 121 ----------
N8 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IOM1 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IOM2 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IOM22 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
E1 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
E2 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATAGCAATC
E3 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTGNTATTCCNATC
E11 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IK1 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IK3 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IK4 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IK5 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IK13 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IK16 121 ----------
89 121 ----------
90 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
91 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
92 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
93 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC
IOM116 121 ATCATTGGCCAACTAGCCTCCATCTCATACTTCTCAATCATCTTAATTCTTATACCAATC


N16 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N18 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N19 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N21 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N22 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTYTTGATAGTATAAACATT
N24 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N27 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N28 181 TCAGGAATTATCGAAGACAAAATWCTAAAAWTATATCCATGTCTTGATAGTATAAACATT
N29 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACMTT
N1 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N4 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N5 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N6 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
N7 181 ----------
N8 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IOM1 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IOM2 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IOM22 181 TCAGGAATTANNNAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
E1 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
E2 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
E3 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
E11 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IK1 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IK3 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IK4 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IK5 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IK13 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IK16 181 ----------
89 181 ----------
90 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
91 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
92 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
93 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT
IOM116 181 TCAGGAATTATCGAAGACAAAATACTAAAATTATATCCATGTCTTGATAGTATAAACATT

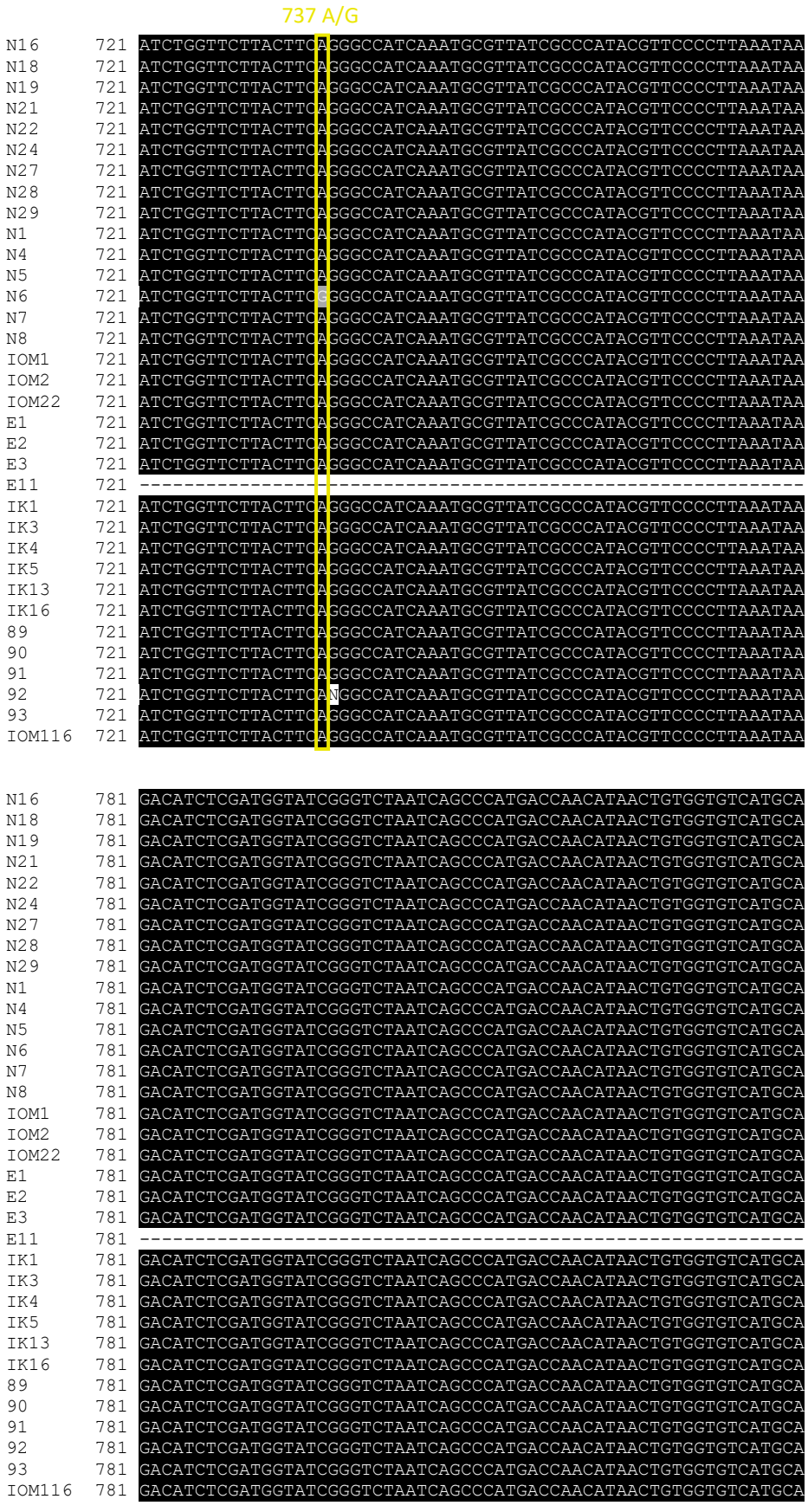

N16 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N18 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N19 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTYTTCTCAAGACATCAAGAAGAAGGA
N21 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N22 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N24 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N27 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N28 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N29 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N1 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N4 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N5 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N6 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
N7 241 ----------
N8 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IOM1 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IOM2 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IOM22 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
E1 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
E2 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
E3 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
E11 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IK1 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IK3 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IK4 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IK5 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IK13 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IK16 241 ----------GA
89 241 ----------
90 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
91 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
92 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
93 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
IOM116 241 ACTCTGGTCTTGTAAACCTGAAATGAAGATCTTCTCTTCTCAAGACATCAAGAAGAAGGA
340 C/T
N16 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N18 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N19 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N21 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N22 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N24 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N27 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N28 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N29 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N1 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N4 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N5 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N6 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N7 301 ----------CCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
N8 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
IOM1 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
IOM2 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
IOM22 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
E1 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
E2 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
E3 301 GCTACTCCCMMCCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
E11 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
IK1 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAACTAAACTACTTCTTGAGTACA
IK3 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAACTAAACTACTTCTTGAGTACA
IK4 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAACTAAACTACTTCTTGAGTACA
IK5 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAACTAAACTACTTCTTGAGTACA
IK13 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAACTAAACTACTTCTTGAGTACA
IK16 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAACTAAACTACTTCTTGAGTACA
89 301 ----------
90 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
91 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
92 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
93 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA
IOM116 301 GCTACTCCCCACCACCAGCACCCAAAGCTGGTATTCTAATTAAACTACTTCTTGAGTACA


431 A/T
467, 468 C/T
470 C/T
479 A/G


481 C/T
511 C/T
520 C/T
535 C/T
599 C/-
N16 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N18 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N19 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
N21 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
N22 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N24 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N27 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N28 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
N29 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
N1 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
N4 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N5 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N6 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N7 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
N8 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
IOM1 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
IOM2 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
IOM22 481 CAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGACATATC
E1 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
E2 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
E3 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
E11 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
IK1 481 TAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGATATATC
IK3 481 TAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGATATATC
IK4 481 TAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGATATATC
IK5 481 TAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGATATATC
IK13 481 TAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGATATATC
IK16 481 TAATATATATACCATGAATATTATCTTAAACACATTAAACTAATGTTATAAGGATATATC
89 481 CAATATATATACCATGAATATTATCTTAAATACATTAAACTAATGTTATAAGGACATATC
90 481 CAATATATATACCATGAATATTATCTTAAATACATTAAACTAATGTTATAAGGACATATC
91 481 CAATATATATACCATGAATATTATCTTAAATACATTAAACTAATGTTATAAGGACATATC
92 481 CAATATATATACCATGAATATTATCTTAAATACATTAAACTAATGTTATAAGGACATATC
93 481 CAATATATATACCATGAATATTATCTTAAATACATTAAACTAATGTTATAAGGACATATC
IOM116 481 CAATATATATACCATGAATATTATCTTAAACACATTAAATTAATGTTATAAGGATATATC
N16 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N18 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N19 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N21 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N22 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N24 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N27 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N28 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N29 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N1 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N4 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N5 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N6 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N7 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
N8 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IOM1 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IOM2 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IOM22 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
E1 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
E2 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
E3 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
E11 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IK1 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IK3 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IK4 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IK5 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IK13 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IK16 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
89 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCC-T
90 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
91 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
92 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
93 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT
IOM116 541 TGTGTTATCTGACATACACCATACAGTCATAAACTCTTCTCTTCCATATGACTATCCCCT

```
N16      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N18      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N19      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N21      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N22      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N24      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N27      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N28      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N29      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N1       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N4       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N5       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N6       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N7       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
N8       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IOM1     601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IOM2     601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IOM22    601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
E1       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
E2       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
E3       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
E11      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IK1      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IK3      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IK4      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IK5      601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IK13     601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IK16     601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
89       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
90       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
91       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
92       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
93       601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC
IOM116   601 TCCCCATTTGGTCTATTAATCTACCATCCTCCGTGAAACCAACAACCCGCCCACCAATGC

N16      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N18      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N19      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N21      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N22      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N24      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N27      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N28      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N29      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N1       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N4       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N5       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N6       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N7       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
N8       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IOM1     661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IOM2     661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IOM22    661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
E1       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
E2       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
E3       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
E11      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACT--------
IK1      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IK3      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IK4      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IK5      661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IK13     661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IK16     661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
89       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
90       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
91       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
92       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
93       661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
IOM116   661 CCCTCTTCTCGCTCCGGGCCCATTAAACTTGGGGGTAGCTAAACTGAAACTTTATCAGAC
```


737 A/G
N16 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N18 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N19 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N21 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N22 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N24 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N27 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N28 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N29 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N1 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N4 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N5 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N6 721 ATCTGGTTCTTACTTCGGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N7 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N8 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IOM1 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IOM2 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IOM22 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
E1 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
E2 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
E3 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
E11 721 ----------
IK1 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IK3 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IK4 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IK5 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IK13 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IK16 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
89 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
90 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
91 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
92 721 ATCTGGTTCTTACTTCANGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
93 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
IOM116 721 ATCTGGTTCTTACTTCAGGGCCATCAAATGCGTTATCGCCCATACGTTCCCCTTAAATAA
N16 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N18 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N19 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N21 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N22 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N24 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N27 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N28 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N29 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N1 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N4 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N5 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N6 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N7 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
N8 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IOM1 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IOM2 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IOM22 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
E1 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
E2 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
E3 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
E11 781 ----------
IK1 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IK3 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IK4 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IK5 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IK13 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IK16 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
89 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
90 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
91 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
92 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
93 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA
IOM116 781 GACATCTCGATGGTATCGGGTCTAATCAGCCCATGACCAACATAACTGTGGTGTCATGCA

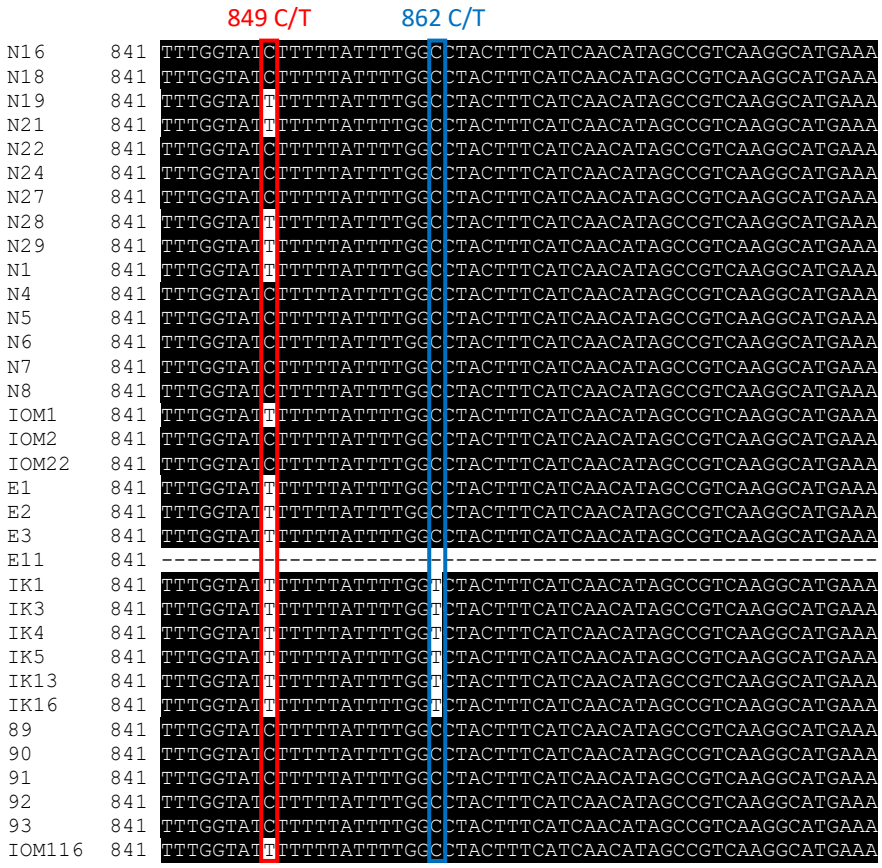


Figure 10 D-loop sequences alignments. Mice E1-E11 were captured from Eday; mice IK1-IK16 from Inchkeith; mice 89-93 from Tollerton; all others are from the Isle of May. Position 1 corresponds to position 15039 of Bibb et al. (1981). SNPs highlighted in red at positions 431, 535 and 849 indicate mutations shared by Inchkeith and Eday haplotypes that separate them from the historical May haplotype. SNPs highlighted in blue, 340, 470, 481, 862, separate Inchkeith mice from all other haplotypes. SNPs highlighted in green, 467, 468, 479, 520, separate Eday mice and May mice with Eday mtDNA from all other haplotypes. SNPs highlighted in orange, 511, 599, separate Tollerton mice from the May haplotype. Just the SNP at 511 separates most of the Tollerton mice from the Isle of May mice, while SNP 599 separates one Tollerton mouse from the rest and from the May type. The SNP highlighted in yellow, 737, separates one of the May mice from the main historical May haplotype, and from all other haplotypes

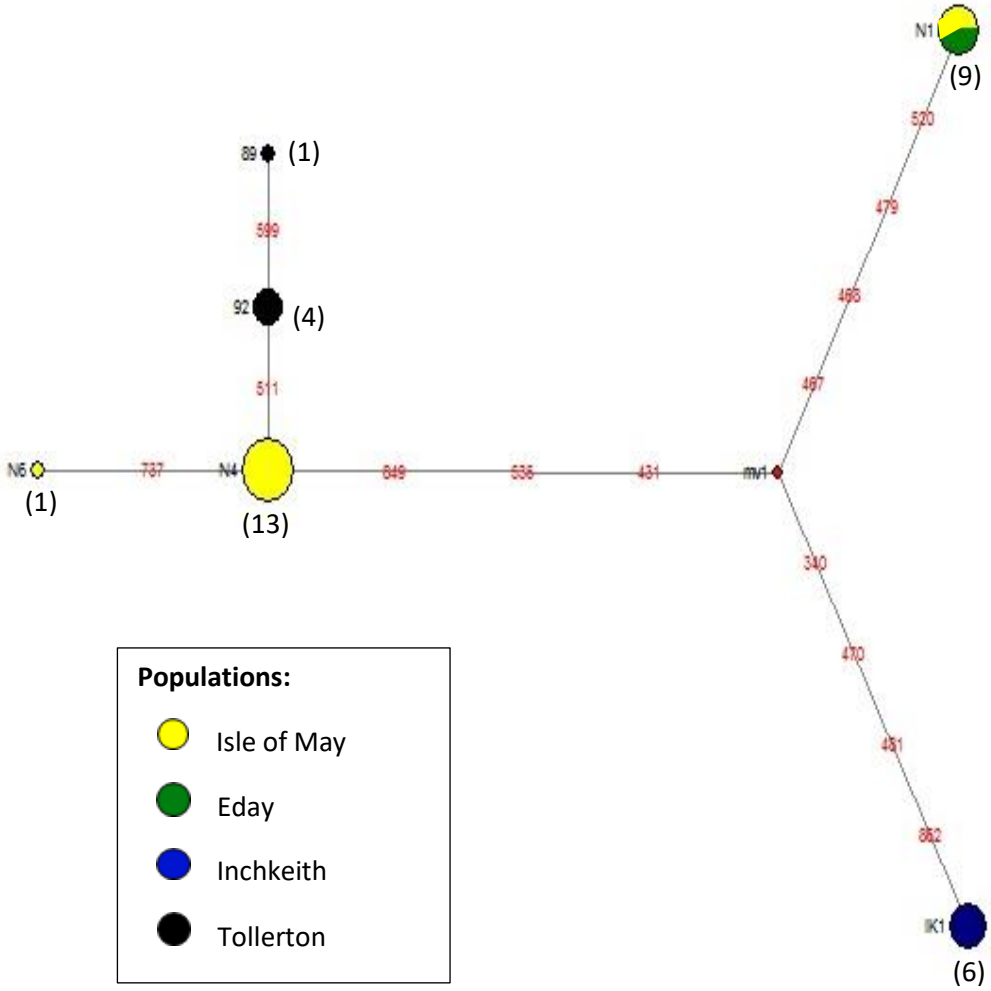


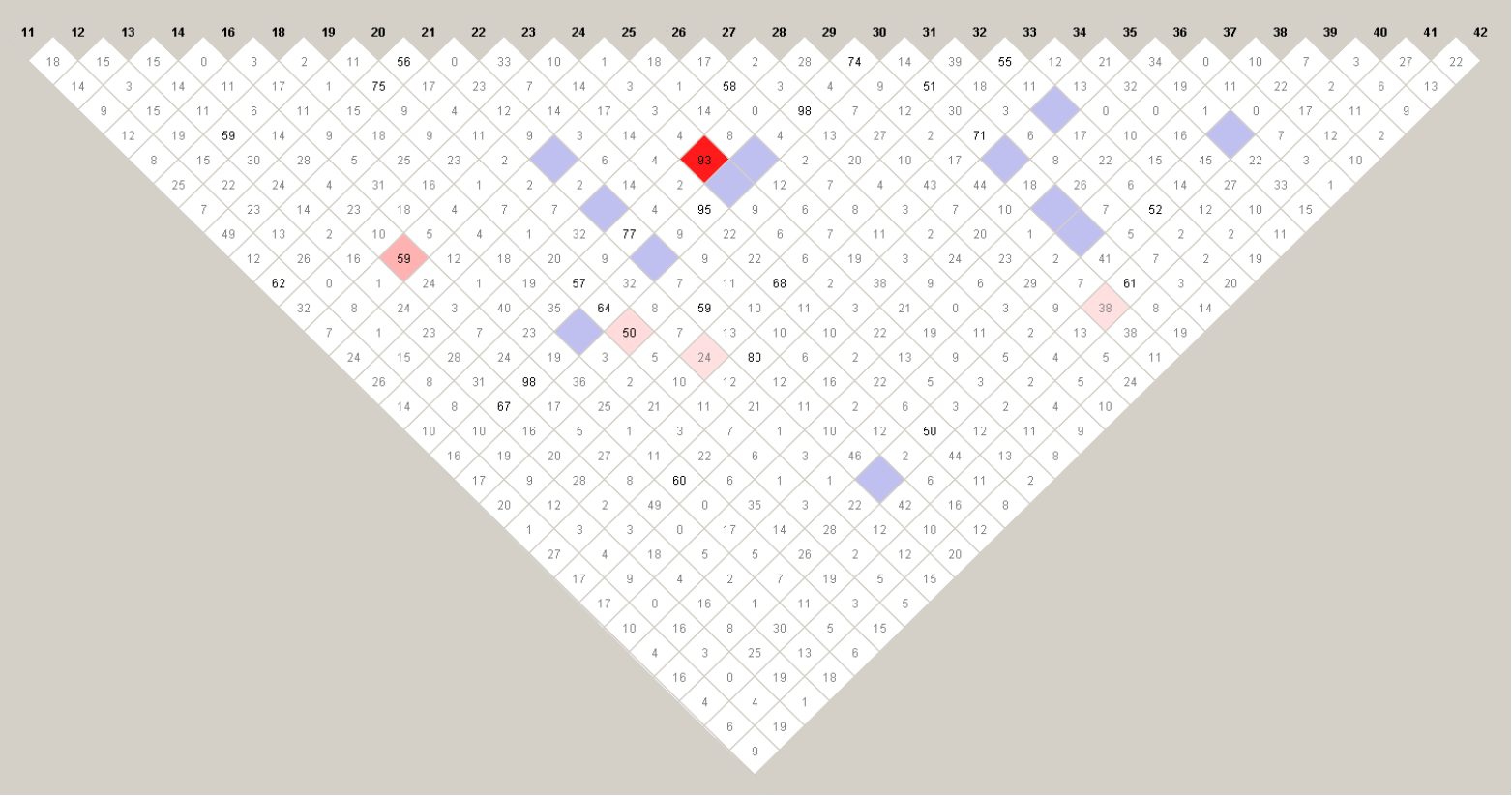
Figure 11 SNP haplotype network for mitochondrial D-loop sequences. Each "step" in the connecting lines marked by a red number represents another mutational SNP between haplotypes. The red circle represents a hypothetical haplotype, indicating where branches split; other circles are coloured according to the locations the mice they contain were captured from. Circles are sized proportionately to the number of mouse sequences they represent, which are marked in brackets (34 sequences total). Circles with multiple colours represent mice from different locations of a single haplotype.

4.3 Immune SNP analysis

4.3.1 Linkage disequilibrium plot

The Haploview plot of linkage disequilibrium (LD) between immune SNPs revealed that just two of the loci displayed strong linkage with each other (Figure 12). These were Il25 and Irf9, numbered 24 and 29 respectively. Both are located on chromosome 14. Several loci pairs showed weak linkage, of which only one pair were on the same chromosome: GATA3_2 and Il2ra_1. The other 3 loci pairs were not on the same chromosomes to each other, and so could not be considered to be linked. SNPs are represented by numbers in Figure 12, which are in the same order as the loci displayed in Table 2, with the first locus, Il17a_1, represented by the number 11, and so on in that fashion. Two loci are not included, these are Gmza_19 and Ifngr1_1, corresponding to 15 and 17, respectively. There was unfortunately no chromosomal information for the Gmza_19 locus available, and as such, it did not appear in the plot. Ifngr1 was likely excluded by the Haploview program as all mice in the sample were homozygous for the same allele at this SNP.

Figure 12 Haploview plot of pairwise LD values between the 32 immune loci. SNPs are in order from left to right starting at 11 (Il17a_1) and ending with 42 (Tyk2_1). Numbers within squares are representative of D' values as percentages. Squares without numbers are indicative of high LD, with a D' value of 1. Red squares indicate high LD (D') with high LOD scores ≥2. Blue squares indicate D'=1 but low LOD scores <2. Shades of pink squares indicate lower D' values less than 1, with LOD scores ≥2. White squares indicate D'<1, with LOD scores <2



4.3.2 STRUCTURE Analysis

Analysis of STRUCTURE results, via calculating the second order rate of change of the likelihood, indicated that the likely number of populations was 4, shown by a peak of ΔK at K=4 (Figure 14). In the corresponding bar plot where STRUCTURE was run for K=4, nearly all individuals assigned their genome roughly equally between all inferred populations; this was also the case when run at K=2 and K=3 (Figure 13).

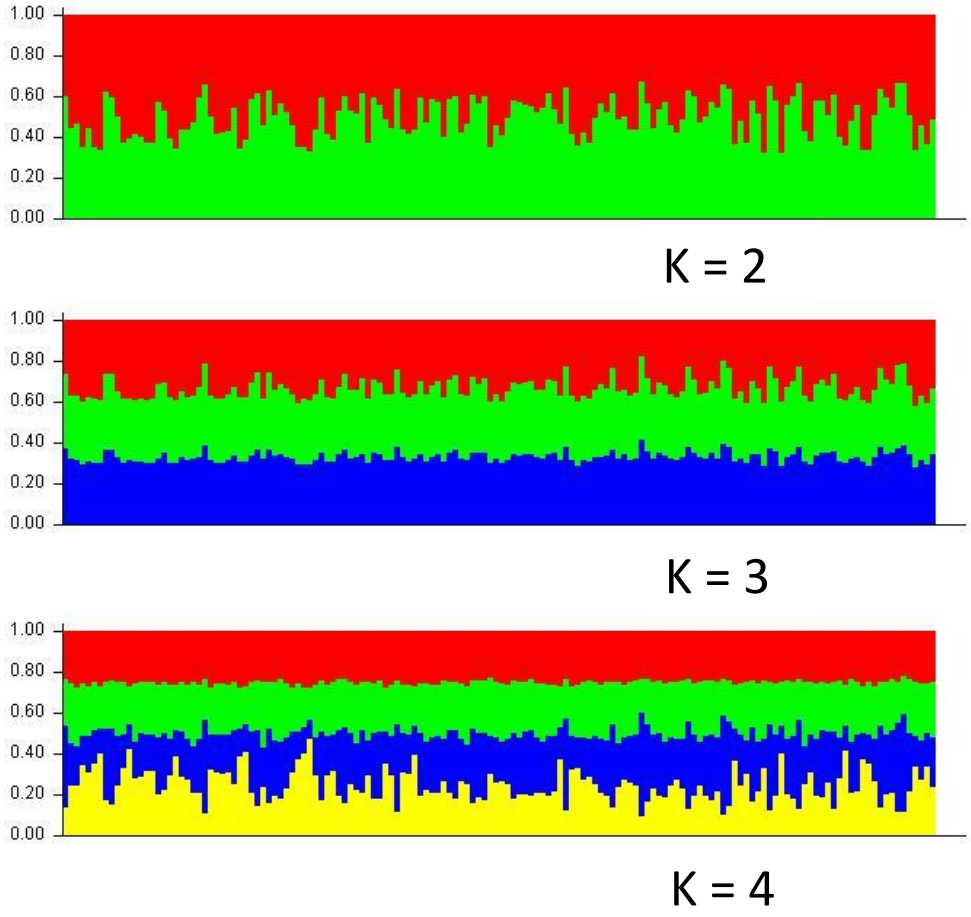


Figure 13 STRUCTURE admixture plots for Isle of May immune SNP data. Results for assumptions of number of population groups, K, are shown for K=2 to K=4. Individuals are represented by vertical lines, which show the proportion of their genotype assigned to each hypothetical population. Each population is represented by its own colour. Individuals are seen to be a mixture of all four populations. Running multiple simulations for each value of K returned similar patterns of genome assignment

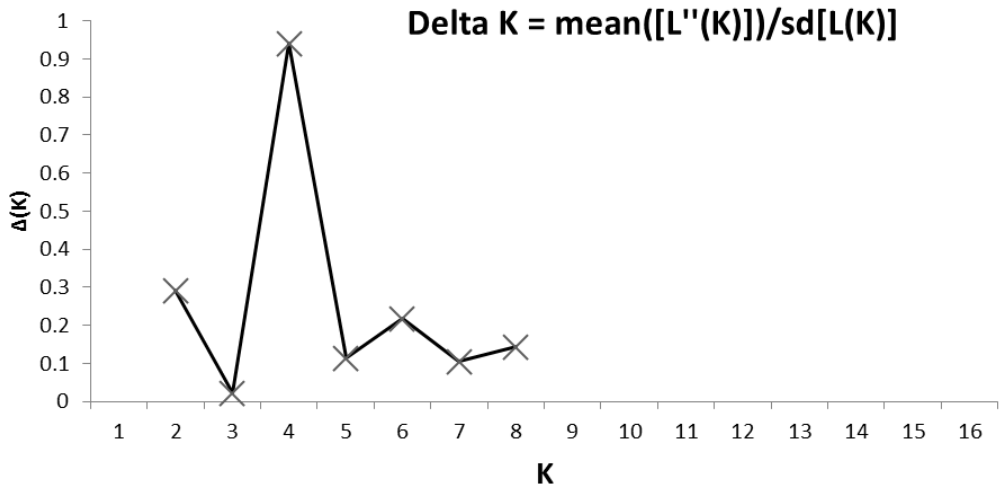


Figure 14 Delta K (ΔK) values plotted for values of K from 2 to 8, based on methods of Evanno et al. (2005), with a clear peak at K=4

4.3.3 Hardy-Weinberg Calculations

Only 3 of the 32 immune loci (9.4%) deviated from Hardy-Weinberg proportions at the population level in the Isle of May population, at the 5% significance level. These were Gmza_19, Il25_2 and Socs7_2. Further inspection of genotypes at these loci in the data, typed for A and T alleles at Gmza_19 and for A and G alleles at Socs7_2, revealed that there were no TT homozygous individuals at Gmza_19, and no AA homozygous individuals at Socs7_2 (Table 4). Gmza_19 deviated with the most significance ($p<0.0001$).

Table 4 Frequencies of genotypes for Gmza_19 and Socs7_2 in the immune SNP data. Question marks indicate unknowns

Locus	Genotype	No. of mice
Gmza_19	AA	32
	AT	83
	TT	0
	??	0
Socs7_2	GG	77
	AG	37
	AA	0
	??	1

4.3.4 Isolation by Distance Analysis

Pairwise estimates of genetic differentiation calculated by ARLEQUIN measured against linear distances of the groups from each other on the island showed a significant negative correlation of isolation by distance, confirmed by a Mantel test (p=0.0097), the opposite way to expected (Figure 13). Multiple Fst values were confirmed as significant at the 5% level by ARLEQUIN, following 16,000 permutations. These values are presented in table 5.

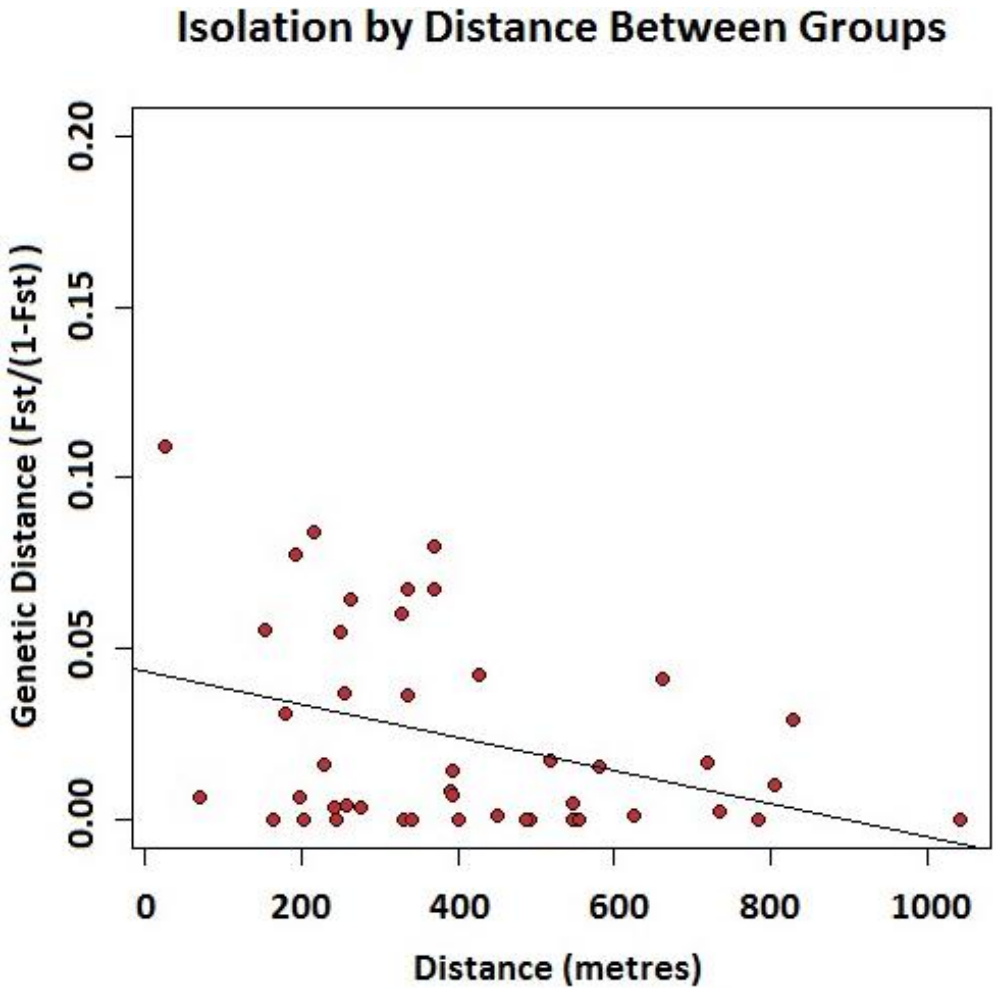


Figure 15 Isolation by distance graph of genetic distance values (indicating differentiation between groups of mice), calculated as (Fst/1-Fst), against linear geographic distance between group locations in metres, with a line of regression.

Table 5 Fst estimates between island groups as calculated by ARLEQUIN. Values significant at the 5% level are highlighted in red.

	SH	BY	HT	FS	ML	G	HR	TTN	LL	BL
SH	0	-	-	-	-	-	-	-	-	-
BY	-0.01106	0	-	-	-	-	-	-	-	-
HT	0.0049	0.00348	0	-	-	-	-	-	-	-
FS	0.00816	0.00632	0.00451	0	-	-	-	-	-	-
ML	0.00274	-0.01053	-0.00022	-0.00575	0	-	-	-	-	-
G	0.03982	0.04049	0.0604	0.07405	0.09828	0	-	-	-	-
HR	-0.00156	0.05665	0.03509	0.03567	0.0069	0.07762	0	-	-	-
TTN	0.02847	0.01556	0.01436	0.01692	0.03026	0.05235	0.06323	0	-	-
LL	0.01003	-0.03752	0.00108	-0.00047	-0.00222	0.07218	0.05189	-0.00033	0	-
BL	-0.00068	-0.01095	0.00145	0.0167	0.00738	0.0632	-0.00403	0.0162	0.00348	0

4.3.5 Phylogeny of Individuals

NeighbourNet mapping of concatenated SNPs, imported as FASTA sequences, between individuals generated a star-like or bushy phylogenetic network, with no clear connection between closeness of individuals on the tree and their group trapping location, indicative of recombination between individuals (Figure 14).

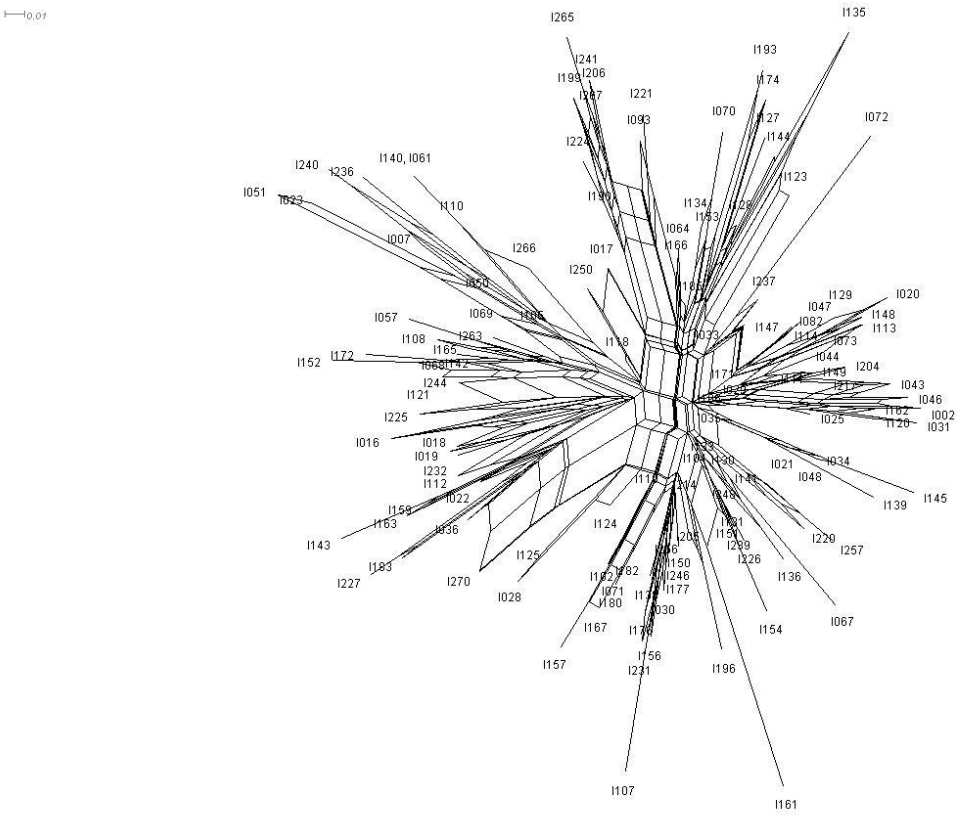


Figure 16 SplitsTree network of concatenated SNPs at the 32 immune loci, with star-like structure. Each individual is considered as a taxa. IDs of individuals are marked at nodes.

5. Discussion and Concluding Remarks

5.1 Phylogenetics

High intraspecific variability has previously been reported in the *M. musculus* species group *Sry* gene fragment downstream of the 5' end of the MHG box, amplified using the primer set (Sry-8653 and Sry-9475) originally described by Albrecht and Eicher (1997), particularly regarding triplet repeats (Terashima et al., 2006). However, regarding variability within *Mus domesticus*, despite much difficulty with sequencing, not enough polymorphism between mice from different locations was discovered such that it could be reliably informative for paternal inheritance of the Isle of May mice. The one polymorphism that was discovered between the mice of different locations distinguished the Eday and May mice from the Inchkeith mice. Though it could not be affirmed, the finding that both the Eday and May Y-chromosomes were identical while the Inchkeith haplotype was distinct may indicate a total replacement of the native May Y-chromosome on the island by the Eday haplotype. This would suggest a continuation of the scenario of differential male genetic success observed by Jones et al. (1995). This may be a function of the Y-chromosome itself, where heritable Y-linked aggression led mice of Eday paternal origin to out-compete less aggressive mice for access to females and possibly shelter (Guillot et al., 1995; Maxson et al., 1996a; Maxson et al., 1996b). On the other hand, it may have been the case soon after the phylogenetic study of Jones et al. (1995), mediated to an extent by female mate choice and inbreeding avoidance (Sherborne et al., 2007).

Whether the lack of differences between Eday and May sequences compared with difference from Inchkeith sequences was indicative of all Isle of May mice being of Eday paternal inheritance was ultimately unclear. Based on this study, this region of the Sry gene can be disregarded for use in phylogenetic and phylogeographic studies of paternal inheritance within *Mus domesticus*, and attention focused on other candidates.

The D-loop analysis shows that the maternal genetics of the original Isle of May mice prior to the Eday introduction persists on the island over 30 years later, and may even be the dominant mitochondrial haplotype, with only 4 Isle of May mice out of 18 sequenced typing for Eday maternal genetics. These results confirm

the earlier findings of Jones et al. (1995), and reinforce the notion that the success of the initial introduction was largely male determined. While not completely absent, the Eday mitochondrial haplotype appears to be conferring no strong selective advantage, or disadvantage, compared to females descended from the original Isle of May population, in stark contrast to the success of the males seen after the initial introduction.

This is in alignment with other studies of island populations of house mice in Western Europe, with male mice generally seen to exhibit greater dispersal and to colonise existing island populations more successfully than females (Pocock et al., 2004; Jones and Searle, 2015). Such sex-specific differences interestingly seem not to be correlated with propensity towards exploratory behaviour, with female mice shown to engage in exploration of unfamiliar environments just as readily as males (Vošlajerová Bímová et al., 2016). In a study measuring competitive behaviour in the Isle of May mice, Grey and Hurst (1998) concluded that males displayed generally reduced aggression with reference to commensal mice, however observed that the Isle of May females displayed scarring and signs of previous aggressive encounters, and preferentially avoided nest sites soiled by other females. Aggression by native May females defending a limited number of suitable nest sites, in addition to avoidance behaviour of female mice may go some way as to explaining the limited introgression of the Eday mitochondrial haplotype on the island. Native females would not be as easy to outcompete as native males, with competition for soiled nest sites risking potentially injurious confrontation by the invading Eday females. However, dispersal to another, unoccupied site suitable for nesting would permit both haplotypes to co-exist on the island. The extent to which this is the case is not known.

Without information on paternal inheritance, the D-loop SNPs were less able to further inform on the whole picture of the phylogenetics of the current population, and the dynamics of interaction between mice of different ancestral origins could not be fully explored.

5.2 Population structure

Conflict within the results of the immune SNP analysis clouded the true structure of the population. However, there was sufficient evidence from that data that the null hypothesis of panmixia of the Isle of May population could not be rejected. Though the ΔK method suggested the true number of populations or subpopulations on the island from the STRUCTURE results was most likely 4, bar plots from the analysis at values of K from 2 to 4 seemed indicate otherwise, with individuals presenting as a mixture of all populations at every value of K, implying a single unstructured population. Erroneous inference of the true number of clusters by the ΔK method is representative of a not uncommon scenario where the method does not work very well (Janes et al., 2017). Based on calculations of the second order rate of change of the likelihood, the method does not allow for K=1 as an answer, for example; this is crucially relevant as that is the cluster number indicated by the bar plots. The authors of the paper introducing the method themselves stressed that it should not be solely relied upon, as it is an ad hoc approach (Evanno et al., 2005).

The genetic markers used must also be critically evaluated when considering confounding results. Selective neutrality and lack of linkage disequilibrium are both assumptions that Bayesian clustering models such as STRUCTURE operate on (Corander et al., 2003). While very little linkage was demonstrated in the genetic markers that I used, the impact of even minor linkage disequilibrium should not be disregarded. Low numbers of markers in linkage disequilibrium may lead to overestimation of clusters by the STRUCTURE program (Kaeuffer et al., 2007). The extent to which the immune loci can be said to be effectively neutral in these analyses is also not defined. Immune genes can be subject to both positive selection for improved resistance to infectious disease and balancing selection for the preservation of polymorphism at the population level (Bernatchez & Landry, 2003; Sommer, 2005; Niskanen et al., 2014), and this may bias Fst estimates (Gharrett et al., 2004).

Though the influence of selection on these estimates can be balanced out in small populations by genetic drift (Elbers et al., 2017), the parameters of population size for which this effect takes place are undefined for mouse populations. However, it has been suggested that selection coefficients less than 0.1% would be sufficient

for natural selection to be less important than neutral processes such as genetic drift in populations comprised of less than 500 individuals (Elbers et al., 2017). Scaling this up to the May population, which is estimated to be able to reach densities of 3000 mice (Triggs, 1991), selection coefficients less than 0.6% may be sufficient for genetic markers at immune loci to be effectively neutral.

That being said, in a comparison of immune SNP genetic markers and putatively neutral microsatellites, differences in estimates of clusters in STRUCTURE, away from the true value, were resolved through principal component analysis (PCA) (Elbers et al., 2017). It was speculated that though population sizes were small and genetic drift was likely the predominant force acting upon variation among the animals, erroneous inference of cluster number using immune SNPs may have been a result of some of the loci displaying non-neutral variation. Use of PCA may be appropriate in this regard, as it has been demonstrated to exceed Bayesian clustering methods such as those used by STRUCTURE in accuracy for detecting subtle structuring, though it may require more loci (Reeves and Richards, 2009).

Multiple studies have demonstrated how unequal sampling size across populations can result in erroneous inference of the true number of clusters (Kalinowski, 2011; Puechmaille, 2016; Wang, 2017). Sample sizes differed between the trapping locations used in the STRUCTURE analysis, with the largest, Fluke Street, containing 27 mice and the smallest containing less than 5. Unequal sampling has been demonstrated to have an effect on the number of clusters inferred using both the ΔK method of Evanno et al. (2005) and the method originally proposed by Pritchard and colleagues (2000) (Wang et al., 2017). Conflict between the results of the ΔK method and the results displayed in the bar plots suggests error, rendering cluster inference by STRUCTURE unreliable in any case.

The number of loci deviating from Hardy-Weinberg proportions was not significant enough to suggest population substructuring, with the probability of 3 out of 32 loci deviating in the population at a 5% significance level being approximately 21.4%, based on the binomial theorem. Furthermore, absence of TT alleles at the Gmza_19 locus coupled with the frequency of heterozygotes typed indicated that there may have been a genotyping error whereby TT type mice were being scored as AT. The frequency of heterozygotes and thus T alleles appear incompatible with the notion that the TT genotype is absent due to producing a deleterious phenotype. The

SplitsTree network, having a star-like structure, implied recombination between the individuals on the island, and as such falls in line with the results of the STRUCTURE bar plots and Hardy-Weinberg calculations.

Taken together, these results may hint at a panmictic population, in which there is no non-random mating. However, the unusual significant negative correlation seen in the isolation by distance analysis seems to indicate slight genetic differentiation across the island, in opposition to these results. Analysis of possible isolation by distance examines whether individuals are more genetically different further away from each other due to the physical distance between them being beyond their dispersal range. It is a simple measure of correlation between linear geographic distances and genetic distances (calculated as Fst/1-Fst), considering sample sites against one another (Jenkins et al., 2010). As such, either no correlation or a positive correlation is expected to be seen, with positive correlation indicating genetic isolation is occurring due to distances of groups of individuals from each other and no correlation indicating genetic isolation is not due to geographically limited dispersal. A negative correlation implies some non-random mating that is occurring as individuals get closer to each other on the island, a scenario which appears counter-intuitive. Significance, by way of the Mantel test of the correlation, cannot be disregarded however.

Aspects of the sampling regime may factor into this unusual result. 2 of the 10 trapping locations used, Main Light and Garden, stand out most prominently in the IBD analysis. They are topographically distinguished from the other sampling sites, being enclosed almost entirely by walls, enclosing denser plant coverage than surrounding areas. These two sites are also walled off from each other, with such barriers potentially restricting gene flow between each of these sampling sites and others on the island. The pairwise Fst estimates hint towards such a scenario, with pairs involving the Main Light and Garden sampling sites accounting for 9 out of the 15 significant Fst values, with the highest Fst value, close to 0.1, arising from the pairwise estimate between the two. Indeed, located quite centrally on the Isle of May relative to the other sampling sites, they account for many of the shorter to medium range geographic distances. Clustering in a small range of geographic distance values, coupled with a large proportion of the significant Fst values, may underlie the peculiar negative association between linear geographic distance and genetic dissimilarity.

The implications of this are important, but caution should be taken in interpretation of these results. It may be first worth considering that unequal sampling size may skew results from isolation by distance analysis. The Main Light and Garden groups considered as populations contained only 6 and 5 mice, respectively. The mice in the Garden sample were also all sampled from the same year, with 4 out of 5 of them being juveniles. As such, they are more likely to have been closely related to one another than mice at the other sampling sites, having yet to have dispersed (Pocock et al., 2005), which may bias Fst estimates at this location. In contrast, sample sizes were much larger at more distant sites such as South Horn and Burnett's Leap, at 19 and 15 respectively, which accounted for only 2 of the significant Fst values each. It could be the case that more variation between individuals at these locations, due to larger sampling size, in congruence with the larger pairwise distances they accounted for, contributed to the pattern seen in the IBD analysis. Imbalance between the sizes of the population samples is potentially much more important to IBD correlations than the total number of individuals such analyses, which is seen to be outweighed in significance by number of variable loci used (Landguth et al., 2012). Whether the degree of genetic differentiation between groups of mice is indicative of non-random mating, rather than an effect of sampling regime, should be considered carefully.

Conclusions derived from use of immune SNPs are further complicated given the discussed complexities of the Isle of May population. There is potential for the house mice on this island to segregate based upon either maternal and/or paternal inheritance, heterogeneity of habitat quality between the sampling sites, or a combination of these factors. Because of this, selection pressures may be operating differentially on different groups of mice, which may lead to differential patterns of immune gene allelic variation in response to fitness demands of different sites (Hellgren, 2015). Despite the small population size on the island, aggregated parasite loads in particular groups of mice by chance may lead to heightened selection on advantageous immune alleles outweighing genetic drift, such that the immune SNPs may no longer serve as neutral genetic markers (Jensen et al., 2008). While allelic variation in immune genes may still be determined primarily by genetic drift in other groups of mice, where selection is involved, results may be unreliable.

Assuming that the loci used in this analysis are neutral and disregarding sampling effects, the isolation by distance result has some important implications for

the structuring of mice within this population. That the topography of sites on the island can influence gene flow in this manner may mean that the feral house mice have the potential to be segregating. Home ranges may shrink based on suitability of the sites that they occupy as protection from both the elements, and to an unknown extent from predators, such that they may not overlap (Triggs, 1991). Should dispersal between sites be limited enough, it would be expected that differential selective pressures have the potential to be operating on groups of mice. Such questions remain unresolved however, due to the unreliability of the results.

As such, these analyses may be benefit from comparison with a panel of "neutral" SNPs in loci with no known gene associations, to see what effects can be attributed to the use of immune genes. It may furthermore be of use to ensure sample sizes are equal between trapping locations considered in both STRUCTURE analysis and isolation by distance models. Principal components analysis (PCA) with the inclusion of more loci may be able to overcome any effects of non-neutrality of the immune loci and give a clearer, less conflicting picture of fine-scale genetic structuring that may be occurring within the Isle of May mice.

5.3 Concluding remarks

The failure of the Sry gene to provide paternal inheritance markers was disappointing but ultimately informative. For studies at this scale attempting to distinguish between populations of the same subspecies, the invariability of the Y chromosome may render sequential investigations of polymorphisms in gene region sequences impractical, as may not be the case for phylogeographic studies determining differences between less recently diverged populations. Long-template PCR has previously been used to screen introns of the Y chromosome to identify sequence polymorphisms in snow voles for use in further phylogeographic studies, and this may be preferable, as it generates substantially more sequence data and thus has increased chance of identifying sufficient numbers of polymorphisms (Wandeler and Camenisch, 2011). Alternatively, development of Y chromosome microsatellite markers, while labour intensive, may be more informative in this regard, as success has been had with phylogeographic studies in *Mus domesticus* (Haasl and Payseur, 2011; Jones and Searle, 2015). Though, as with the SNP approach, the usefulness at the scale of this study is uncertain. As next-generation

sequencing (NGS) technologies improve and become more accessible, and the power of studies utilising SNPs as molecular markers becomes akin to that of microsatellites, studies of this nature will become easier, and more informative.

Given the results of this study, analysis of the genetics of the Isle of May population still has the potential to yield some very interesting results. Further investigation of these mice at the population genetics level is certainly warranted, so as to unravel the ecological web underlying immune variation in this wild population.

References

"The Story of the Isle of May National Nature Reserve". (2014). **2nd ed**. [ebook] 1-28 Available at: http://www.snh.gov.uk/docs/A1516651.pdf [Accessed 27 Jun. 2017].

Adler, G. H. and Levins, R. (1994) "The island syndrome in rodent populations" *The Quarterly Review of Biology* **69(4)**, 473-90

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D. and Wayne, R. K. (2004) "High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal" *Proceedings of the National Academy of Sciences of the United States of America* **101(10)**, 3490-3494

Albrecht, K. H. and Eicher, E. M. (1997) "DNA Sequence Analysis of *Sry* Alleles (Subgenus *Mus*) Implicates Misregulation as the Cause of C57BL/6J-Y^{POS} Sex Reversal and Defines the SRY Functional Unit" *Genetics* **147**, 1267-1277

Anderson, P. K., Dunn, L. C. and Beasley, A. B. (1964) "Introduction of a lethal allele into a feral house mouse population" *The American Naturalist* **98(898)**, 57-64

Andrews, C. A. (2010) "Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations" *Nature Education Knowledge* **3(10)**, 5

Andrews, C. A. (2010) "Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations" *Nature Education Knowledge* **3(10)**, 5

Arnqvist, G., Dowling, D. K., Eady, P., Gay, L., Tregenza, T., Tuda, M. and Hosken, D. J. (2010) "Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect" *Evolution* **64**, 3354–3363

Baird, S.J., Ribas, A., Macholán, M., Albrecht, T., Piálek, J. and Goüy de Bellocq, J. (2012) “Where are the wormy mice? A reexamination of hybrid parasitism in the European house mouse hybrid zone” *Evolution* **66(9)**, 2757-2772

Ballard, J.W. and Kreitman, M. (1995) “Is mitochondrial DNA a strictly neutral marker?” *Trends in Ecology and Evolution* **10**, 485-488

Barreiro, L.B. and Quintana-Murci, L. (2010) “From evolutionary genetics to human immunology: how selection shapes host defence genes” *Nature Reviews Genetics* **11**, 17-30

Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. (2005) “Haploview: analysis and visualization of LD and haplotype maps” *Bioinformatics* **21**, 263–265

Bar-Yaacov, D., Blumberg, A. and Mishmar D. (2012) “Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation” *Biochimica et Biophysica Acta* **1819**, 1107–1111

Beaumont, M. A. and Nichols, R. A. (1996) “Evaluating loci for use in the genetic analysis of population structure” *Proceedings of the Royal Society of London B: Biological Sciences* **263**, 1619-1626

Bedhomme, S., Agnew, P., Sidobre, C., Michalakis, Y. (2004) “Virulence reaction norms across a food gradient” *Proceedings of the Royal Society of London B: Biological Sciences* **271**, 739–744

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

Bernatchez, L. and Landry, C. (2003) “MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years?” *Journal of Evolutionary Biology* **16(3)**, 363-377

Berry, R. J. (1964) “The evolution of an island population of the house mouse” *Evolution* **18**, 468-483

Berry, R. J. (1968) “The Ecology of an Island Population of the House Mouse” *Journal of Animal Ecology* **37(2)**, 445-470

Berry, R. J. (1981) “Town Mouse, Country Mouse: adaptation and adaptability in *Mus domesticus* (*M. musculus domesticus*)” *Mammal Review* **11**, 91-136

Berry, R. J. (1986) “Genetical Processes in Wild Mouse Populations. Past Myth and Present Knowledge” In: Potter, M., Nadeau, J. H. and Cancro, M. P. Eds. *The Wild Mouse in Immunology* New York: Springer-Verlag 86-94

Berry, R. J. (1986) “Genetics of insular populations of mammals, with particular reference to differentiation and founder effects in British small mammals” *Biological Journal of the Linnean Society*, **28**, 205-230

Berry, R. J. and Jakobson, M. E. (1974) “Vagility in an island population of the house mouse” *Journal of Zoology* **173**, 341–354

Berry, R. J. and Peters, J. (1977) “Heterogeneous Heterozygosities in *Mus musculus* Populations” *Proceedings of the Royal Society of London B: Biological Sciences* **197**, 485-503

Berry, R. J., Triggs, G. S., King, P., Nash, H. R. and Noble, L. R. (1991) “Hybridization and gene flow in house mice introduced into an existing population on an island” *Journal of Zoology, London* **225**, 615-632

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W. and Clayton, D. A. (1981) “Sequence and gene organization of mouse mitochondrial DNA” *Cell* **26**, 167–180

Boursot, P., Din, W., Anand, R., Darviche, D., Dod, B., Von Deimling, F., Talwar, G. P. and Bonhomme, F. (1996) “Origin and radiation of the house mouse: mitochondrial DNA phylogeny” *Journal of Evolutionary Biology* **9(4)**, 391-415

Brown, P. R. (2007) “Reducing the impact of feral house mice in agricultural ecosystems” In: Lunney, D., Eby, P., Hutchings, P. and Burgin, S. Eds. *Pest or Guest: the zoology of overabundance* New South Wales: Mosman 8-15

Bryant, D. and Moulton, V. (2004) “Neighbor-Net: An agglomerative method for the construction of phylogenetic networks” *Molecular Biology and Evolution* **21**, 255–265

Carlier, M., Roubertoux, P.L. and Pastoret, C. (1991) “The Y chromosome effect on intermale aggression in mice depends on the maternal environment” *Genetics* **129**, 231–236

Carlier, M., Roubertoux, P.L., Kottler, M.L. and Degrelle, H. (1990) “Y chromosome and aggression in strains of laboratory mice” *Behaviour Genetics* **20**, 137–156

Choquenot, D., Ruscoe, W. A. (2000) “Mouse population eruptions in New Zealand forests: the role of population density and seedfall” *Journal of Animal Ecology* **69**, 1058–1070

Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D. and Lewis, L. C. (2009) “Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis” *Journal of Heredity* **100(5)**, 556-564

Corander, J., Waldmann, P. and Sillanpaa, M. J. (2003) “Bayesian analysis of genetic differentiation between populations” *Genetics* **163**, 367–374

Corby-Harris, V. and Promislow, D. E. L. (2008) “Host ecology shapes geographical variation for resistance to bacterial infection in Drosophila melanogaster” *Journal of Animal Ecology* **77**, 768–776

Cucchi, T., Vigne, J.-D. and Auffray, J.-C. (2005) “First occurrence of the house mouse (*Mus musculus domesticus* Schwarz and Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil house mouse occurrences” *Biological Journal of the Linnean Society* **84**, 429-445

Dallas, J. F., Dod, B., Boursot, P., Prager, E. M. and Bonhomme, F. (1995) “Population subdivision and gene flow in Danish house mice” *Molecular Ecology* **4**, 311–320

Darwin, C. (1871) “The Descent of Man, and Selection in Relation to Sex” London: John Murray

Din, W., Anand, R., Boursot, P., Darviche, D., Dod, B., Jouvin-Marche, E., Orth, A., Talwar, G. P., Cazenave, P. A. and Bonhomme, F. (1996) “Origin and radiation of the house mouse: clues from nuclear genes” *Journal of Evolutionary Biology* **9(5)**, 519-539

Dow, H. C., Kreibich, A. S., Kaercher, K. A., Sankoorikal, G. M. V., Pauley, E. D., Lohoff, F. W., Ferraro, T. N., Li, H. and Brodkin, E. S. (2011) “GENETIC DISSECTION OF INTERMALE AGGRESSIVE BEHAVIOR IN BALB/cJ AND A/J MICE” *Genes, Brain, and Behavior* **10(1)**, 57–68

Dowling, D. K., Abiega, K. C. and Arnqvist, G. (2007) “Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles” *Evolution* **61**, 194–201

Elbers, J. P., Clostio, R. W. and Taylor, S. S. (2017) “Population genetic inferences using immune gene SNPs mirror patterns inferred by microsatellites” *Molecular Ecology Resources* **17(3)**, 481-491

Evanno, G., Regnaut, S. and Goudet, J. (2005) “Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study” *Molecular Ecology* **14(8)**, 2611-2620

Excoffier, L. and Lischer, H. E. L. (2010) “Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows” *Molecular Ecology Resources* **10**, 564-567

Figueroa, F., Tichy, H., Berry, R. J. and Klein, J. (1986) “MHC polymorphism in island populations of mice” *Current Topics in Microbiology and Immunology* **127**, 100-105

Fitzgerald, B. M., Karl, B. J. and Moller, H. (1981) “Spatial Organization and Ecology of a Sparse Population of House Mice (*Mus musculus*) in a New Zealand Forest” *Journal of Animal Ecology* **50(2)**, 489-518

Foster, J. B. (1963) “The evolution of native land mammals of the Queen Charlotte Islands and the problem of insularity” PhD dissertation, University of Columbia, Vancouver.

Foster, J. B. (1964) “The evolution of mammals on islands” *Nature* **202**, 234-235

Frynta, D., Slábová, M., Váchová, H., Volfová, R. and Munclinger, P. (2005) “Aggression and commensalism in house mouse: a comparative study across Europe and near east” *Aggressive Behaviour* **31**, 283-293

Gabriel, S. I., Jóhannesdóttir, F., Jones, E. P. and Searle, J. B. (2010) “Colonization, mouse-style” *BMC Biology* **8,** 131

Gabriel, S. I., Stevens, M. I., Mathias, M. and Searle, J. B. (2011) “Of Mice and ‘Convicts’: Origin of the Australian House Mouse, *Mus musculus*” *PLoS One* **6(12)**, e28622

Gershoni, M., Levin, L., Ovadia, O., Toiw, Y., Shani, N., Dadon, S., Barzilai, N., Bergman, A., Atzmon, G., Wainstein, J., Tsur, A., Nijtmans, L., Glaser, B. and Mishmar, D. (2014) "Disrupting Mitochondrial–Nuclear Coevolution Affects OXPHOS Complex I Integrity and Impacts Human Health" *Genome Biology and Evolution* **6(10)**, 2665–2680

Glaubitz, J. C. (2004) "Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages" *Molecular Ecology Notes* **4(2)**, 309–310

Gonzalez-Quevedo, C., Spurgin, L. G., Illera, J. C. and Richardson, D. S. (2015) "Drift, not selection, shapes toll-like receptor variation among oceanic island populations" *Molecular Ecology* **24**, 5852–5863

Google Maps (2017) Map of the Isle of May, Firth of Forth, Scotland [online]. Available at: https://www.google.co.uk/maps/@56.1856754,-2.5529282,1976m/data=!3m1!1e3 [Accessed 20 July 2017]

Graham, A. L., Hayward, A. D., Watt, K. A., Pilkington, J. G., Pemberton, J. M. and Nussey, D. H. (2010) "Fitness correlates of heritable variation in antibody responsiveness in a wild mammal" *Science* **330(6004)**, 662-665

Gray, S. J. and Hurst, J. L. (1998) "Competitive behaviour in an island population of house mice, *Mus domesticus*" *Animal Behaviour* **56(5)**, 1291-1299

Gray, S. J., Jensen, S. P. and Hurst, J. L. (2000) "Structural complexity of territories: preference, use of space and defence in commensal house mice, *Mus domesticus*" *Animal Behaviour* **60(6)**, 765-772

Greminger, M. P., Krützen, M., Schelling, C., Pienkowska-Schelling, A. and Wandeler, P. (2010) "The quest for Y-chromosomal markers – methodological strategies for mammalian non-model organisms" *Molecular Ecology Resources* **10(3)**, 409-420

Grueber, C. E., Wallis, G. P. and Jamieson, I. G. (2013) "Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species" *Molecular Ecology* **22**, 4470–4482

Guenet, J. L. and Bonhomme, F. (2003) “Wild mice: an ever-increasing contribution to a popular mammalian model” *Trends in Genetics* **19**, 24–31

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lopoittevin, C., Malausa, T., Revardel, E., Salin, F. and Petit, R. J. (2011) “Current trends in microsatellite genotyping” *Molecular Ecology Resources* **11**, 591–611

Guillot, P.V., Carlier, M., Maxson, S.C. and Roubertoux, P.L. (1995) “Intermale aggression tested in two procedures, using four inbred strains of mice and their reciprocal congenics: Y chromosomal implications” *Behaviour Genetics* **25**, 357–360

Haasl, R. J. and Payseur B. A. (2011) “Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites” *Heredity* **106(1)**, 158–171

Haasl, R. J., and Payseur, B. A. (2011) “Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites” *Heredity* **106**, 158–171

Hagström, E., Freyer, C., Battersby, B. J., Stewart, J. B. and Larsson N. G. (2014) “No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline” *Nucleic Acids Research* **42(2)**, 1111-1116

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

Hardouin, E. A., Chapuis, J. L., Stevens, M. I., van Vuuren, J. B., Quillfeldt, P., Scavetta, R. J., Teschke, M. and Tautz, D. (2010) “House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion” *BMC Evolutionary Biology* **10**, 325

Hauser, L., Baird, M., Hilborn, R., Seeb, L. W., and Seeb, J. E. (2011) “An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population” *Molecular Ecology Resources* **11**, 150–161

Hellgren, O. (2015) “Allelic variation at innate immune genes (avian β-defensins), within a natural population of great tits” *Journal of Avian Biology* **46**, 113-118

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A,, Maes, G. E., Diopere, E., Carvalho, G. R. and Nielsen, E. E. (2011) "Application of SNPs for population genetics of non-model organisms: new opportunities and challenges" *Molecular Ecology Resources* **11(s1)**, 123–136

Henig, R. M. (2000) "The Monk in the Garden" Boston: Mariner Books

Hoshino, A. A., Bravo, J. P., Nobile, P. M. and Morelli, K. A. (2012) "Microsatellites as Tools for Genetic Diversity Analysis" In: Caliskan, M. Ed. *Genetic Diversity in Microorganisms* InTech

Hurst, J. L., Hall, S., Roberts, R. and Christian, C. (1996) "Social organisation in the aboriginal house mouse, *Mus spretus* Lataste: behavioural mechanisms underlying the spatial dispersion of competitors" *Animal Behaviour* **51**, 327–344

Huson, D. H. and Bryant, D. (2006) "Application of Phylogenetic Networks in Evolutionary Studies" *Molecular Biology and Evolution* **23(2)**, 254-267

Ingvarsson, P. K. and Whitlock, M. C. (2000) "Heterosis increases the effective migration rate" *Proceedings of the Royal Society of London* **267**, 1321-1326

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I. and Andrew, R. L. (2017) "The K=2 conundrum." *Molecular Ecology* **26(14)**, 3594-3602

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., Knox, J., Long, T., Mannino, M., McGuire, M., Riffle, A., Segelsky, S., Shappell, L., Sterner, A., Strickler, T. and Tursi, R. (2010) "A meta-analysis of isolation by distance: relic or reference standard for landscape genetics?" *Ecography* **33**, 315-320

Jensen, L. F., Hansen, M. M., Mensberg, K. L. and Loeschcke, V. (2008) "Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.)" *Heredity* **100(1)**, 79-91

Jensen, S. P., Gray, S. J. and Hurst, J. L. (2005) "Excluding neighbours from territories: effects of habitat structure and resource distribution" *Animal Behaviour* **69**, 785-795

Jones, A. G., Small, C. M., Paczolt, K. A. and Ratterman, N. L. (2010) “A practical guide to methods of parentage analysis” *Molecular Ecology Resources* **10**, 6–30

Jones, C. S., Noble, L. R., Jones, J. S., Tegelström, H., Triggs, G. S., Berry, R. J. (1995) “Differential male genetic success determines gene flow in an experimentally manipulated mouse population” *Proceedings of the Biological Society of Washington* **260**, 251-256

Jones, E. P. and Searle, J. B. (2015) “Differing Y chromosome versus mitochondrial DNA ancestry, phylogeography, and introgression in the house mouse” *Biological Journal of the Linnean Society* **115(2)**, 348–361

Jones, E. P., Eager, H. M., Gabriel, S. I., Jóhannesdóttir, F. and Searle, J. B. (2013) “Genetic tracking of mice and other bioproxies to infer human history” *Trends in Genetics* **29**, 298-308

Kaeuffer, R., Reale, D., Coltman, D. W. and Pontier, D. (2007) “Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium” *Heredity* **99**, 374–380

Kalinowski, S. T. (2011) “The computer program structure does not reliably identify the main genetic clusters within species: simulations and implications for human population structure” *Heredity* **106**, 625–632

Klein, J. (1975) “Biology of the mouse histocompatibility-2 complex” New York: Springer-Verlag

Krasnov, B. R. and Chochlova, J. S. (1994) “Spatial ethological structure of groups” In: Kotenkova, E. and Bulatova, N. S. Eds. *The House Mouse: Origin, distribution, systematics, behaviour* Moskva: Nauka 188–214

Lambert, M. Vial, F. Pietravalle, S. and Cowan, D. (2017) “Results of a 15-year systematic survey of commensal rodents in English dwellings” *Scientific Reports* **7**

Landguth, E. L., Fedy, B. C., Oyler-McCance, S. J., Garey, A. L., Emel, S. L., Mumma, M., Wagner, H. H., Fortin, M. J. E. and Cushman, S. A. (2012) “Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern” *Molecular Ecology Resources* **12**, 276–284

Langton, S. D., Cowan, D. P. and Meyer, A. N. (2001) “The occurrence of commensal rodents in dwellings as revealed by the 1996 English House Condition Survey” *Journal of Applied Ecology* **38**, 699–709

Lazzaro, B. P., Flores, H. A., Lorigan, J. G. and Yourth, C. P. (2008) “Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in *Drosophila melanogaster*” *PLoS Pathogens* **4**, e1000025

Lindström, K. M., Foufopoulos, J., Pärn, H. and Wikelski., M. (2004) “Immunological investments reflect parasite abundance in island populations of Darwin’s finches” *Proceedings of the Royal Society B: Biological Sciences* **271**, 1513–1519

Livingstone, D. S., Motamayor, J. C., Schnell, R. J., Cariaga, K., Freeman, B., Meerow, A. W., Brown, J. S. and Kuhn, D. N. (2010) “Development of single nucleotide polymorphism markers in *Theobroma cacao* and comparison to simple sequence repeat markers for genotyping of Cameroon clones” *Molecular Breeding* **27**, 93–106

Lomolino, M. V. (1985) “Body Size of Mammals on Islands: The Island Rule Reexamined” *The American Naturalist* **125(2)**, 310-316

Lynch, M., Bobay, L. M., Catania, F., Gout, J. F. and Rho, M. (2011) “The repatterning of eukaryotic genomes by random genetic drift” *Annual Review of Genomics and Human Genetics* **12**, 347–366

Martin, L. B., Weil, Z. M. and Nelson, R. J. (2008) “Seasonal changes in vertebrate immune activity: mediation by physiological trade-offs” *Philosophical Transactions of the Royal Society B: Biological Sciences* **363(1490)**, 321–339

Maxson, S.C. (1996a) “Issues in the search for candidate genes in mice as potential animal models of human aggression” *Ciba Foundation Symposium* **194**, 21–30

Maxson, S.C. (1996b) “Searching for candidate genes with effects on an agonistic behavior, offense, in mice” *Behaviour Genetics* **26**, 471–476

Mayer, A., Mora, T., Rivoire, O. and Walczak, A. M. (2015) “Diversity of immune strategies explained by adaptation to pathogen statistics” *Proceedings of the National Academy of Sciences of the United States of America* **113**, 8630-8635

Morin, P. A., Archer, F. I., Pease, V. L., Hancock-Hanser, B. L., Robertson, K. M., Huebinger, R. M., Martien, K. K., Bickham, J. W., George, J. C., Postma, L. D. and Taylor, B. L. (2012) "Empirical comparison of single nucleotide polymorphisms and microsatellites for population and demographic analyses of bowhead whales" *Endangered Species Research* **19**, 129-147

Morin, P. A., Luikart, G. and Wayne, R. K. (2004) "SNPs in ecology, evolution and conservation" *Trends in Ecology & Evolution* **19**, 208–216

Morse, H. C. (2007) "Building a better mouse: One hundred years of genetics and biology" In: Fox, J. G., Barthold, S. W., Davisson, M. T., Newcomer, C. E., Quimby, F. W. and Smith, S. L. Eds. *The mouse in biomedical research second edition* San Diego: Academic Press 1–11

Mortaud, S., Nicolas, L., Pinoteau, W., Tordjman, S., Carlier, M., Roubertoux, P.L. (2010) "Brain pathways mediating the pro-aggressive effect of the steroid sulfatase (Sts) gene" *Behaviour Genetics* **40**, 211–219

Muir, T. S. (1885). *Ecclesiological notes on some of the Islands of Scotland*. Edinburgh: David Douglas 291-292

Muñoz-Muñoz, F., Sans-Fuentes, M. A., López-Fuster, M. J. and Ventura, J. (2011) "Evolutionary modularity of the mouse mandible: dissecting the effect of chromosomal reorganizations and isolation by distance in a Robertsonian system of *Mus musculus domesticus*" *Journal of Evolutionary Biology* **24**, 1763–1776

Myers, J. H. (1974a) "Genetic and Social Structure of Feral House Mouse Populations on Grizzly Island, California" *Ecology* **55**, 747–759

Myers, J. H. (1974b) "The absence of *t*-alleles in feral populations of house mice" *Evolution*, **27**, 702-704

National Nature Reserve, (2014), Isle of May [online]. Available at: http://www.rspb.org.uk/community/cfs-file.ashx/__key/communityserver-discussions-components-files/1920/5633._5F00_MG_5F00_7854.jpg [Accessed 20 July 2017]

Nielsen, E. E., Hansen, M. M. and Meldrup, D. (2006) “Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in non-model organisms” *Molecular Ecology* **15**, 3219–3229

Niskanen, A. K., Kennedy, L. J., Ruokonen, M., Kojola, I., Lohi, H., Isomursu, M., Jansson, E., Pyhäjärvi, T. and Aspi, J. (2014) “Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population” *Molecular Ecology* **23**, 875-889

Otti, O., Gantenbein-Ritter, I., Jacot, A. and Brinkhof, M. W. G. (2012) “Immune response increases predation risk” *Evolution* **66(3)**, 732-739

Outram, B. (2016, July 14) “Gull Results” Retrieved from https://isleofmaynnr.wordpress.com/2016/07/14/gull-results/

Parkes, J., Murphy, E. (2003) “Management of introduced mammals in New Zealand” *New Zealand Journal of Zoology* **30,** 335–359

Piry, S., Luikart, G. and Cornuet, J. (1999) “BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data” *Journal of Heredity* **90**, 502–503

Pocock, M. J. O., Searle, J. B. and White, P. C. L. (2004) “Adaptations of animals to commensal habitats: population dynamics of house mice *Mus musculus domesticus* on farms” *Journal of Animal Ecology* **73**, 878–888

Pocock, M. J., Hauffe, H. C. and Searle, J. B. (2005) “Dispersal in house mice” *Biological Journal of the Linnean Society* **84(3)**, 565-583

Pörtner, H. O., Peck, L., and Somero, G. (2007) “Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view” *Philosophical Transactions of the Royal Society B: Biological Sciences* **362(1488)**, 2233–2258

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

Puechmaille, S. J. (2016) “The program structure does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem” *Molecular Ecology Resources* **16**, 608–627

Putman, A. I. and Carbone, I. (2014) “Challenges in analysis and interpretation of microsatellite data for population genetic studies” *Ecology and Evolution* **4(22)**, 4399–4428

Råberg, L. (2014) “How to Live with the Enemy: Understanding Tolerance to Parasites” *PLoS Biology* **12(11)**, e1001989

Raymond, M. and Rousset, F. (1995) “GENEPOP (VERSION-1.2)—population-genetics software for exact tests and ecumenicism” *Journal of Heredity* **86**, 248-249

Reeves, P. A., Richards, C. M. (2009) “Accurate Inference of Subtle Population Structure (and Other Genetic Discontinuities) Using Principal Coordinates” *PLOS ONE* **4(1)**, e4269

Rius, M. and Darling, J. A. (2014) “How important is intraspecific genetic admixture to the success of colonising populations?” *Trends in Ecology and Evolution* **29**, 233-242

Robertson, R. (2014) “A study of the parasite biology and immunogenetics of natural populations of the house mouse *Mus musculus domesticus* on the Isle of May, Scotland and two other locations” Unpublished master’s thesis, University of Nottingham

Roubertoux, P.L., Carlier, M., Degrelle, H., Haas-Dupertuis, M.C., Phillips, J. and Moutier, R. (1994) “Co-segregation of intermale aggression with the pseudoautosomal region of the Y chromosome in mice” *Genetics* **136**, 225–230

Rousset, F. (2008) “GENEPOP’007: a complete re-implementation of the GENEPOP software for Windows and Linux” *Molecular Ecology Resources* **8**, 103-106

Ryan, A. W., Duke, E. J. and Fairley, J. S. (1993) “Polymorphism, localization and geographical transfer of mitochondrial DNA in *Mus musculus domesticus* (Irish house mice)” *Heredity* **70**, 75–81

Ryan, M. T. and Hoogenraad, N. J. (2007) “Mitochondrial-nuclear communications” *Annual Review of Biochemistry* **76**, 701–722

Schlötterer, C. (2004) “The evolution of molecular markers — just a matter of fashion?” *Nature Reviews Genetics* **5**, 63-69

Schulenburg, H., Kurtz, J., Moret, Y., and Siva-Jothy, M. T. (2009) “Introduction. Ecological immunology” *Philosophical Transactions of the Royal Society B: Biological Sciences* **364(1513)**, 3-14

Scriven, P. (1992) “Robertsonian translocations introduced into an island population of mice” *Journal of Zoology* **227**, 493-503

Scriven, P. N., Bauchau, V. (1992) “The effect of hybridization on mandible morphology in an island population of the house mouse” *Journal of Zoology* **226**, 573–583

Searle, J. B., Jones, C. S., Gündüz, I., Scascitelli, M., Jones, E. P., Herman, J. S., Rambau, R. V., Noble, L. R., Berry, R. J., Giménez, M. D. and Jóhannesdóttir, F. (2009) “Of mice and (Viking?) men: phylogeography of British and Irish house mice” *Proceedings of the Biological Society of Washington* **276**, 201-207

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S. and Seeb, L.W. (2011) “Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms” *Molecular Ecology Resources* **11**, 1-8

Shaw, D. J., and Dobson, A. P. (1995) “Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review” *Parasitology* **111(Suppl.)**, 111–127

Shaw, D. J., Grenfell, B. T. and Dobson, A. P. (1998) “Patterns of microparasite aggregation in wildlife host populations” *Parasitology* **117**, 597–610

Sheldon, B. C. and Verhulst, S. (1996) “Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology” *Trends in Ecology and Evolution* **11(8)**, 317-321

Sherborne, A. L., Thom, M. D., Paterson, S., Jury, F., Ollier, W. E. R., Stockley, P., Beynon, R. J. and Hurst, J. L. (2007) “The Genetic Basis of Inbreeding Avoidance in House Mice” *Current Biology* **17(23)**, 2061–2066

Silver, L. M. (1995a) “Systematics of the *Mus* species group” In: “Mouse Genetics: Concepts and Applications” Oxford: Oxford University Press (Revised 2008 online edition)

Silver, L. M. (1995b) “Polymorphisms detected by PCR” In: “Mouse Genetics: Concepts and Applications” Oxford: Oxford University Press (Revised 2008 online edition)

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

Southern, H. N. (1938) “A Survey of the Vertebrate Fauna of the Isle of May (Firth of Forth)” *Journal of Animal Ecology*, **7(1)**, 144-154

Spurgin, L. G. and Richardson, D. S. (2010) “How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings” *Proceedings of the Royal Society B: Biological Sciences* **277(1684)**, 979–988

Steel, D (2017, May 18) “Gull-p!” Retrieved from https://isleofmaynnr.wordpress.com/2017/05/18/gull-p/

Steel, D. (2018, January 15) “Short of the Mark” Retrieved from https://isleofmaynnr.wordpress.com/2018/01/15/short-of-the-mark/

Stjernman, M., Råberg, L. and Nilsson, J-Å. (2008) “Maximum Host Survival at Intermediate Parasite Infection Intensities” *PLoS ONE* **3(6)**, e2463

Strand, T. M., Segelbacher, G., Quintela, M., Xiao, L., Axelsson, T. and Höglund, J. (2012) “Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse?” *Ecology and Evolution* **2(2)**, 341-353

Suzuki, H., Nunome, M., Kinoshita, G., Aplin, K. P., Vogel, P., Kryukov, A. P., Jin, M-L, Han, S-H, Maryanto, I., Tsuchiya, K., Ikeda, H., Shiroishi, T., Yonekawa, H. and Moriwaki, K. (2013) “Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA” *Heredity* **111**, 375-390

Suzuki, H., Yakimenko, L. V., Usuda, D. and Frisman, L. V. (2015) “Tracing the eastward dispersal of the house mouse, *Mus musculus*” *Genes and Environment* **37**, 20

Tattersall, F. H., Smith, R. H. and Nowell, F. (1997) “Experimental colonisation of contrasting habitats by house mice” *Zeitschrift für Säugetierkunde* **62**, 350–358

Terashima, M., Furusawa, S., Hanzawa, N., Tsuchiya, K., Suyanto, A., Moriwaki, K., Yonekawa, H. and H Suzuki, H. (2006) “Phylogeographic origin of Hokkaido house mice (*Mus musculus*) as indicated by genetic markers with maternal, paternal and biparental inheritance” *Heredity* **96**, 128–138

Triggs, G. S. (1991) “The population ecology of house mice (*Mus domesticus*) on the Isle of May, Scotland” *Journal of Zoology, London* **225**, 449-468

Väli, Ü., Saag, P., Dombrovski, V., Meyburg, B. U., Maciorowski, G., Mizera, T., Treinys, R. and Fagerberg, S. (2010) “Microsatellites and single nucleotide polymorphisms in avian hybrid identification: a comparative case study” *Journal of Avian Biology* **41**, 34–49

Van Oortmerssen, G.A. and Sluyter, F. (1994) “Studies on wild house mice. V. Aggression in lines selected for attack latency and their Y-chromosomal congenics” *Behaviour Genetics* **24**, 73–78

Van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2015) “Isolation-by-distance in landscapes: considerations for landscape genetics” *Heredity* **114(1)**, 27–37

Viney, M., Lazarou, L. and Abolins, S. (2015) “The laboratory mouse and wild immunology” *Parasite Immunology* **37**, 267-273

Viney, M.E. (2006) “Modelling nematode infections?” *Parasite Immunology* **28**, 239–241

Vošlajerová Bímová, B., Mikula, O., Macholán, M., Janotová, K., Hiadlovská, Z. and Wright, J. (2016) “Female House Mice do not Differ in Their Exploratory Behaviour from Males” *Ethology* **122**, 298-307

Wandeler, P. and Camenisch, G. (2011) “Identifying Y-chromosomal diversity by long-template PCR” *Molecular Ecology Resources* **11(5)**, 835-841

Wang, J. (2017) “The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse” *Molecular Ecology Resources* **17**, 981-990

Weir, B. S. and Cockerham, C. C. (1984) “Estimating F-statistics for the analysis of population structure” *Evolution* **38**, 1358–1370

Weissbrod, L., Marshall, F. B., Valla, F. R., Khalaily, H., Bar-Oz, G., Auffray, J. C., Vigne, J. D. and Cucchi, T. (2017) “Origins of house mice in ecological niches created by settled hunter-gatherers in the Levant 15,000 y ago” *Proceedings of the National Academy of Sciences of the United States of America* **114(16)**, 4099-4104

White, T. A. and Searle, J. B. (2007) “Genetic diversity and population size: island populations of the common shrew, *Sorex araneus*” *Molecular Ecology*, **16**, 2005-2016

White, T. A. and Searle, J. B. (2007) “Genetic diversity and population size: island populations of the common shrew, *Sorex araneus*” *Molecular Ecology* **16**, 2005-2016

White, T. A., Perkins, S. E. and Dunn, A. (2012) “The ecoimmunology of invasive species” *Functional Ecology* **26**, 1313-1323

Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A. and Dowling, D. K. (2014) “Mitonuclear interactions: evolutionary consequences over multiple biological scales” *Philosophical Transactions of the Royal Society B: Biological Sciences* **369(1646)**

Wong, B. B. M. and Candolin, U. (2005) “How is female mate choice affected by male competition?” *Biological Reviews of the Cambridge Philosophical Society* **80**, 559–571

Woolhouse, M. E. J., Dye, C., Etard, J.-F., Smith, T., Charlwood, J. D., Garnett, G. P., Hagan, P., Hii, J. L. K., Ndhlovu, P. D., Quinnell, R. J., Watts, C. H., Chandiwana, S. K. and Anderson, R. M. (1997) “Heterogeneities in the transmission of infectious agents: Implications for the design of control programs” *Proceedings of the National Academy of Sciences of the United States of America* **94(1)**, 338–342

Wright S. (1943) “Isolation by distance” *Genetics* **28**, 114–138

Yakimenko, L. V., Suzuki, H., Usuda, D. and Frisman, L. V. (2015) “Tracing the eastward dispersal of the house mouse, *Mus musculus*” *Genes and Environment* **37**, 20

Yonekawa, H., Tsuda, K., Tsuchiya, K., Yakimenko, L. V., Korobitsyna, K. V., Chelomina, G. N. et al. (2003) “Genetic diversity, geographic distribution and evolutionary relationships of *Mus musculus* subspecies based on polymorphisms of mitochondrial DNA” In: Kryukov, A. P., Yakimenko, L. V. Eds. *Problems of Evolution. Vol V* Dalnauka: Vladivostok 90–108