The Social Structure of the Hazel Dormouse (*Muscardinus avellanarius*)

Deborah Glass

University of Brighton

This dissertation is submitted for the degree of

*Doctor of Philosophy*

PhD 2017
I would like to dedicate this thesis firstly to both by parents for their emotional and financial support and secondly to Dr Andrew Overall, without whom it would not be nearly so interesting!
Declaration

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Deborah Glass
PhD 2017
Acknowledgements

Firstly, I would like to thank both my supervisors Dr Andrew Overall and Dr Dawn Scott, for providing me with the opportunity of embarking on a PhD, sparking my interest in cute fluffy mammals, helping me collect samples and supporting me when deciding to leave my job and make it my full-time occupation. Primarily, I would like to thank Andy, who has guided me throughout the whole process with endless patience and kindness, even when I was being difficult and obstinate. For taking me through the dreaded red book of population genetics and making it an enjoyable and satisfying experience. Also, for being a true mentor, igniting my love for all things population genetics and evolution and steering me towards the subjects of social evolution and kin recognition, which I find endlessly fascinating! For encouraging me and believing I could make a great scientist one day, which gave me the confidence to go on even through the darkest and seemingly endless points of my PhD. I would like to thank Adam Grogan, Richard Thompson, Steven Robinson, Ian Powell, Janet Rose, Nida Al-Fulaij and Ian White for letting me tag along on their box checks to collect samples and imparting their wisdom and hilarious dormouse anecdotes. I would especially like to thank Adam for training me up for my dormouse licence and Nida, Ian White and Ian Powell for collecting samples for me when I couldn’t make it. I would like to give a special thanks to Hazel Ryan for giving up so much of her time to let me take samples and swabs from her captive dormice and her generosity in conveying her extensive knowledge on dormice. I would like to give a big shout out to Laurie, Rachel and Row who all joined me for various weekends of dormouse related fun on the Isle of Wight (especially Laurie, who did lots of other collecting on my behalf too) and for putting up with me when I was tired and grumpy. A substantial amount of my genetics work was undertaken at NERC Biomolecular Analysis Facility, University of Sheffield. It was a privilege to work with so many experts in molecular ecology, which made the experience thoroughly enriching. I would like to show my gratitude to Deborah Dawson, Gavin Horsburgh and Helen Hippperson for all their help and guidance. I would especially like to thank Deborah Dawson for her enthusiasm for the project and her assistance. I would also like to thank Becca and Andy for giving me a place to
stay when I needed to pop back to Sheffield. My family and friends (you know who you are!) were a huge support to me throughout. A PhD can be at times an isolating experience and their unwavering reassurances and encouragement, even when they had no idea what I was talking about, helped me more than they can ever know. This experience has been the most enriching of my life so far. I come away from it filled with fond memories of misbehaving dormice, the fabulous people I’ve met along the way and the amazing subject area I have been fortunate enough to study. I will never be the same again and hopefully I finally have the brain the size of a planet (Dad!).
Abstract

The maintenance of genetic variation is a general conservation concern for endangered species, such as the British hazel dormouse (*Muscardinus avellanarius*). The structuring of genetic variation at multiple hierarchical levels, from individual through to the total population, can provide insights into within-population processes, such as inbreeding and social behaviour, and help understand the causes of population partitioning, both current and historical. Investigating the processes that shape and alter levels of genetic variation within hazel dormouse populations will highlight any factors that may influence their continued persistence and inform on appropriate management strategies.

The current thesis aimed to address whether a) microbial-mediated odour could be a mechanism for kin discrimination, inbreeding avoidance and social cooperation; b) the hazel dormouse displays social structuring and, as a result, evidence for inbreeding avoidance; c) there is within and among population structuring, resulting from identified physical features that restrict gene flow and d) current patterns of genetic variation inform us about historic dispersal.

Captive bred individuals of known pedigree were used to assess the relationship between bacterial community composition and relatedness. Microsatellite markers were used to generate relatedness estimates and analyse levels of genetic variation at the individual, social group and sub-population levels on samples obtained from Sussex and the Isle of Wight. Mitochondrial markers were used to infer patterns of genetic variation at the total population level, including samples obtained from all over the species range. F-statistics were then used to infer any deviations from Hardy-Weinberg equilibrium that could imply for example inbreeding or gene flow.

The main findings of the thesis are that 1) microbial composition correlates with genetic relatedness in the hazel dormouse, indicating that microbial-mediated odour could be a mechanism for kin discrimination; 2) inbreeding levels are not significantly high, even though both male and female close relatives are in close proximity, with neither sex exclusively practising natal philopatry or dispersal; 3) no identified habitat features appeared to influence gene flow and levels of genetic variation did not differ between sites, regardless of the habitat features; 4) on the basis of mtDNA, the British
dormouse population is likely to have been founded by a relatively recent colonisation event, rather than be a remnant of a more ancient indigenous species.

The results of this study contribute to our understanding of how social and dispersal mechanisms determine a population’s composition and hence provide some indicators of how the species has evolved. It is intended that knowledge of the species’ social structure and, in particular, details of their inbreeding avoidance behaviour and tolerance of close relatives, will help optimise population survival in future reintroduction programs.
Glossary

A: Number of alleles

Adult: Age class of individuals that are >12g in weight and have survived their first hibernation

AR: Allelic richness

bp: Base pairs

ESU: Evolutionary Significant Units

$F_{IS}$: Inbreeding coefficient: the probability of identity by decent of alleles within individuals relative to alleles within the subpopulation

$F_{ST}$: Inbreeding coefficient: the probability of identity by decent of alleles within sub-population relative to alleles within the total population

$F_{IT}$: Inbreeding coefficient: the probability of identity by decent of alleles within individuals relative to alleles within the total population

F SS NN: Female Same Sex Nearest Neighbour Pairs

Grey: Age class of individuals that are between 20-40 days old and/or between 6 - 10g in weight

$H'$: Shannon index

$H_E$: Expected heterozygosity

$H_O$: Observed Heterozygosity

ha: Hectare

Hap: Haplotype

HW: Hardy Weinberg

HWE: Hardy Weinberg Equilibrium

IBD: Identity By Decent

IUCN: International Union for Conservation of Nature

Juvenile: Age class of individuals that are over 40 days old 10g+ in weight and before first hibernation.

m: Migration rate

Hd: Haplotype diversity

MHC: Major Histocompatibility Complex

M SS NN: Male Same Sex Nearest Neighbour Pairs

$\mu$: Mutation rate

N: Census number

NN: Nearest Neighbour

n: number of samples

NAP: Number of Private Alleles
NDMS: National Dormouse Monitoring Scheme
\( N_e \): Effective population size
OS NN: Opposite sex Nearest Neighbour Pairs
OTU: Operational Taxonomic Unit
\( \pi \): Nucleotide diversity
PCR: Polymerase Chain Reaction
PTES: Peoples Trust for Endangered Species
\( r \): Relatedness
T: Total number of TRFs
T-RFLP: Terminal Restriction Fragment Length Polymorphism
TRF: Terminal Restriction Fragment
Table of contents

List of figures .......................................................... xix

List of tables ................................................................ xxv

1 The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*) .................. 1
  1.1 The Hazel Dormouse *Muscardinus avellanarius* .......................................................... 1
  1.1.1 The distribution and conservation status of *Muscardinus avellanarius* ......................... 1
  1.1.2 Causes of the population decline and current conservation strategies ......................... 3
  1.1.3 Life history of the hazel dormouse ............................................................................. 6
  1.2 Genetic variation ......................................................... 8
  1.2.1 Processes influencing genetic diversity ................................................................. 9
  1.2.2 Genetic markers ......................................................................................... 10
  1.2.3 Microsatellites .......................................................................................... 12
  1.2.4 Mitochondrial DNA .................................................................................. 12
  1.3 Hierarchical approach to conserving species .............................................................. 13
  1.3.1 Individual level of genetic diversity ...................................................................... 13
  1.3.2 Effects of habitat fragmentation on individual level genetic diversity ....................... 14
  1.3.3 Social structure and genetic diversity ................................................................. 16
  1.3.4 Effects of habitat fragmentation on social structure ............................................ 21
  1.4 Regional scale of genetic diversity ............................................................ 22
  1.4.1 Anthropogenic fragmentation ........................................................................... 24
  1.5 Geographic scale of genetic diversity ...................................................................... 25
  1.6 Thesis rationale and aims ............................................................................. 26
  1.6.1 Thesis plan ..................................................................................................... 27
## Table of contents

### 2 General Methods and Materials 31

#### 2.1 Study sites 31

#### 2.2 Sample Collection 33

#### 2.3 DNA extraction 34

##### 2.3.1 Tissue extraction protocol 34

##### 2.3.2 Hair extraction protocol 36

#### 2.4 PCR Microsatellite amplification 37

#### 2.5 Microsatellite error checking 37

##### 2.5.1 Sample selection 39

#### 2.6 Employing the Hardy-Weinberg Principle 39

### 3 Bacterially Mediated Scent Production as a Possible Mechanism for Kin Discrimination 43

#### 3.1 Introduction 45

##### 3.1.1 Why would kin recognition evolve in the hazel dormouse? 45

#### 3.2 Kin recognition mechanisms 48

##### 3.2.1 Olfactory communication 51

##### 3.2.2 Odour as a kin recognition cue 51

##### 3.2.3 The role of bacteria in scent communication 52

##### 3.2.4 Factors affecting microbial composition of the scent glands and gut 55

##### 3.2.5 Environmental factors 57

##### 3.2.6 Aims and objectives 58

#### 3.3 Methods and materials 59

##### 3.3.1 Sample collection 59

##### 3.3.2 Extractions 62

##### 3.3.3 Digest 63

##### 3.3.4 Analysis 63

##### 3.3.5 Bacterial communities and relatedness 64

##### 3.3.6 Environment vs genetic variation 64

##### 3.3.7 Bacterial 16S rRNA diversity measures 64

#### 3.4 Results 65

##### 3.4.1 Comparing extraction techniques 65

##### 3.4.2 Correlation between host relatedness and TRF profiles of the scent-gland 65

##### 3.4.3 Environment vs genetics 72

##### 3.4.4 Diversity 72

#### 3.5 Discussion 80
3.5.1 Relationship between genetic relatedness and bacterial community structure .................................................. 80
3.5.2 Genetic component versus a shared environment: which one explains the relationship between bacterial community structure and relatedness? .................................................. 83
3.5.3 The influence of age, sex and environment on bacterial community composition .................................................. 83
3.6 Conclusions and Implications .................................................. 86

4 The Social Structure of the British Hazel Dormouse (Muscardinus avellanarius) .................................................. 89
4.1 Introduction ........................................................................ 91
4.1.1 Aims and predictions .................................................. 94
4.2 Methods and Materials .................................................. 95
4.2.1 Sample collection and processing .................................................. 95
4.2.2 Relatedness coefficient .................................................. 95
4.2.3 Genetic vs. geographic distance .................................................. 97
4.2.4 Kin tolerance and avoidance .................................................. 98
4.2.5 Population and individual level inbreeding .................................................. 100
4.3 Results ........................................................................ 100
4.3.1 The relatedness coefficient .................................................. 100
4.3.2 Summary of demographics .................................................. 106
4.3.3 Genetic vs. geographic distance .................................................. 108
4.3.4 Kin tolerance .................................................. 112
4.3.5 Inbreeding .................................................. 115
4.4 Discussion ........................................................................ 118
4.5 Conclusions ........................................................................ 127

5 Genetic Variation and Population Structure of the British Hazel Dormouse (Muscardinus avellanarius) .................................................. 129
5.1 Introduction ........................................................................ 130
5.1.1 Measuring the structuring of genetic variation .................................................. 130
5.1.2 The genetic impacts of habitat fragmentation .................................................. 132
5.1.3 The threat of habitat loss and fragmentation to the hazel dormouse .................................................. 134
5.1.4 Aims and predictions .................................................. 137
5.2 Methods and materials .................................................. 138
5.2.1 Study populations .................................................. 138
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.2</td>
<td>Sampling sites</td>
<td>138</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Sampling procedure</td>
<td>140</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Sample selection</td>
<td>140</td>
</tr>
<tr>
<td>5.2.5</td>
<td>Clustering analysis</td>
<td>140</td>
</tr>
<tr>
<td>5.2.6</td>
<td>Genetic structure</td>
<td>141</td>
</tr>
<tr>
<td>5.2.7</td>
<td>Isolation - by - distance measures</td>
<td>141</td>
</tr>
<tr>
<td>5.2.8</td>
<td>Detection of first generation immigrants</td>
<td>141</td>
</tr>
<tr>
<td>5.2.9</td>
<td>Genetic variation</td>
<td>142</td>
</tr>
<tr>
<td>5.2.10</td>
<td>Relatedness coefficient</td>
<td>143</td>
</tr>
<tr>
<td>5.2.11</td>
<td>Sample pooling</td>
<td>143</td>
</tr>
<tr>
<td>5.3</td>
<td>Results</td>
<td>145</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Microsatellite markers</td>
<td>145</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Sample pooling</td>
<td>146</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Isolation - by - distance</td>
<td>147</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Among population genetic structure</td>
<td>147</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Genetic variation among populations</td>
<td>153</td>
</tr>
<tr>
<td>5.3.6</td>
<td>Within site genetic variation and structure</td>
<td>157</td>
</tr>
<tr>
<td>5.4</td>
<td>Discussion</td>
<td>175</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Among population genetic structuring and variation</td>
<td>175</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Within population structure</td>
<td>179</td>
</tr>
<tr>
<td>5.5</td>
<td>General conclusions</td>
<td>183</td>
</tr>
<tr>
<td>6</td>
<td>The Phylogenetic Scale of British Hazel Dormouse Diversity: The Story So Far</td>
<td>185</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>186</td>
</tr>
<tr>
<td>6.1.1</td>
<td>The current phylogenetic status of the British hazel dormouse</td>
<td>186</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Genetic drift - gene flow equilibrium</td>
<td>190</td>
</tr>
<tr>
<td>6.1.3</td>
<td>Aims and objectives</td>
<td>190</td>
</tr>
<tr>
<td>6.2</td>
<td>Methods and materials</td>
<td>191</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Mitochondrial DNA</td>
<td>191</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Statistical analysis</td>
<td>192</td>
</tr>
<tr>
<td>6.3</td>
<td>Results</td>
<td>193</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Phylogenetic trees</td>
<td>193</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Genetic Diversity</td>
<td>203</td>
</tr>
<tr>
<td>6.4</td>
<td>Discussion</td>
<td>205</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Conclusions and implications</td>
<td>208</td>
</tr>
</tbody>
</table>
7 General Discussion

7.1 Introduction ................................................................. 211
7.2 Inbreeding avoidance and social structuring in the hazel dormouse ... 212
7.3 The influences of potential barriers on levels of genetic variation and
gene flow ................................................................. 214
7.4 Differences between microsatellite and mitochondrial markers ........ 216
7.5 Conclusions and implications ............................................ 218

References ................................................................. 221
## List of figures

1.1 Range map of the hazel dormouse [1] .................................................. 2
1.2 The hazel dormouse distribution a) Victorian dormouse distribution (data from Rope, G.T. (1885) On the range of the dormouse in England and Wales [2]) b) the current distribution of dormice in Britain as of 2013 [3] ...................... 4

2.1 Map of South East England showing all sample site locations. Woodland is depicted in green and urban development in grey. IoW: Isle of Wight (Briddlesford); BW: Binsted Wood; TP: Tilgate Park; W: Worth; WH: Wakehurst Place; BH: Bexhill; MD: Mallydams Wood; FR: Flatropers; HB: Herne Bay .................. 33
2.2 Visual depiction of dormouse hair plucking technique. .......................... 35

3.1 The pedigree depicting the Woods family, which was a four generation family sampled from the Wildwood Trust, with sibships in the same colour. .................. 60
3.2 Photographs taken at the Wildwood Trust a) demonstrating how the anus was exposed and b) demonstrating how the swabs were taken. .................... 61
3.3 Percentage of shared TRFs and degree of relatedness using the basic dataset for both sexes. ................................................................. 67
3.4 Percentage of shared TRFs compared with degree of relatedness using the basic dataset for dams and sires with juveniles. .................. 68
3.5 Percentage of shared TRFs compared with the degree of relatedness using the basic dataset for dams and sires with greys. .................. 69
3.6 Percentage of shared TRFs compared with degree of relatedness using consensus dataset for dams and sires with juvenile offspring. .................. 70
3.7 Percentage of shared TRFs compared with degree of relatedness using consensus dataset for dams and sires with grey offspring. .................. 71
3.8 Percentage of shared TRFs compared with relatedness for relatives and non-relatives housed in the same and separate pens using the consensus dataset. .................. 73
3.9 Percentage of shared TRFs of relatives and non-relatives compared with environment, where 1 are those housed together and 2 are housed separately using the consensus dataset. 74

3.10 A cluster dendrogram of the Jaccard index scores of dissimilarity for T in 2013 (1), 2014 (2) and 2015 (3) using the basic dataset. 75

3.11 A cluster dendrogram of the Jaccard index scores of dissimilarity in adults (1), juveniles (2) and greys (3) using the basic dataset. 76

3.12 A cluster dendrogram of the Jaccard index scores of dissimilarity in 2013 (1), 2014 (2) and 2015 (3) using the consensus dataset. 79

4.1 Distribution of nest boxes with a positive result for dormouse sample collection for a) MW (22ha) and b) IoW (59.26ha). 96

4.2 Illustration of possible scenarios where, considering one individual at a time, its nearest same-sex or opposite neighbour is not necessarily its nearest neighbour. Also, the other individual of the pair may itself have a different same or opposite-sex neighbour (single headed arrow). 99

4.3 Plot depicting five relatedness coefficients obtained from seven captive dormice pairs known to be unrelated. Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times. 102

4.4 Plot depicting five relatedness coefficients obtained from 12 captive dormice pairs known to be related to the degree of \( r = 0.25 \) (grandparent/grandchild). Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times. 103

4.5 Plot depicting five relatedness coefficients obtained from 14 captive dormice pairs with a known parent/offspring relationship and related to the degree of \( r = 0.5 \). Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times. 104

4.6 Plot depicting five relatedness coefficients obtained from 12 captive bred dormice pairs known to be full sibs and related to the degree of \( r = 0.5 \). Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times. 105

4.7 Graphical output of relationship between genetic relatedness (between same-sex pairs) and geographic distance for females a) Wang’s relatedness coefficient; b) recreated pedigree and males c) Wang’s relatedness coefficient; d) recreated pedigree, for Mallydams Wood samples. 109

4.8 Graphical output of relationship between genetic relatedness (between same-sex pairs) and geographic distance for females a) Wang’s relatedness coefficient; b) recreated pedigree and males c) Wang’s relatedness coefficient; d) recreated pedigree, for IoW. 110
4.9 Relationship between genetic relatedness (between opposite sex pairs) and geographic distance for Mallydams Wood a) relatedness coefficient and b) recreated pedigree and IoW c) relatedness coefficient and d) recreated pedigree.

4.10 Plot depicting mean $r$ and mean distance with standard error bars of the total population and nearest pair for same sex individuals at both sites with the sample number of pairs in brackets. Females are represented by circles and males by squares. MW individuals are in black and IoW individuals are in blue, with mean total $r$ for each sex represented by solid colour and nearest pair mean $r$ for each sex by the striped pattern. Mean distances (D) for each category are represented by a red cross.

4.11 Plot shows mean relatedness of females and males opposite sex pairs (OSP) compared to total adult population mean relatedness. MW individuals are in black and IoW individuals are in blue. Mean $r$ for OSP are represented by circles and mean $r$ for total populations are represented by triangle.

4.12 Plot shows mean relatedness for NN pairs compared to total population mean $r$ and all individuals. NN included opposite sex (OS), female same sex (F SS) and male same sex (M SS pairs). MW is represented by black and IoW by blue, with triangles for OS, circles for F SS, squares for M SS and diamonds for the total population. All means have SE bars.

4.13 ConStruct results for MW (a) and IoW (b), with $g$ set at 0.0625. Shows the joint likelihood of the causes of excess homozygosity: cryptic substructure (FST) and/or consanguinity, where $Cg$ is the proportion of the sample inbred to degree $g$. The support limit for the maximum likelihood value is given by G.

4.14 Plot shows individual inbreeding scores for both sexes at both sites. Circles represent females (F) and squares represent males (M). Mallydams (MW) individuals are black for females and grey for males, IoW individuals are light blue for females and dark blue for males. Striped individuals are those found to be significantly inbred using the maximum likelihood method. The dashed red line depicts the limit of confidence (i.e., the possibility that $F=0$ is excluded).

5.1 Map of the positions of all four sites in Sussex (BW: Binsted Wood; TP:Tilgate Park; WH: Wakehurst; MD: Mallydams) and the one on the Isle of Wight (IoW: Briddlesford).

5.2 Results of the POWSIM analyses based on 17 microsatellites. The $y$-axis represents the power of the markers to successfully recover the value of FST indicated on the $x$-axis, expressed as the proportion of 1000 simulations.

5.3 Isolation-by-distance relationship between geographical distance (km) and genetic distance (pairwise FST) for all sites.
5.4 Bar plot showing the estimated membership coefficient (Q) for each individual. Each vertical bar corresponds to one individual and each cluster is represented by a different colour. Groups of individuals are labelled by site (M: Mallydams; W: Wakehurst Place; B: Binsted Wood; T: Tilgate and I: IoW).

5.5 Genetic diversity measures for IoW, East Sussex and pooled West Sussex populations. a) shows mean allelic richness (AR) and b) shows mean expected heterozygosity (HE) with standard error bars.

5.6 Bar plot showing the estimated membership coefficient (Q) for each individual. Each vertical bar corresponds to one individual and each cluster is represented by a different colour, West Sussex is in blue and East Sussex is in green. Groups of individuals are labelled by site (M: Mallydams Wood; W: Wakehurst Place; B: Binsted Wood; T: Tilgate).

5.7 Bar plot showing the estimated membership coefficient (Q) for each individual. Each vertical bar corresponds to one individual and each of the clusters is represented by a different colour, Binsted is in green, Tilgate Park in blue and Wakehurst Place is in light blue. Groups of individuals are labelled by site (B: Binsted Wood; T: Tilgate and W: Wakehurst Place).

5.8 Allele frequencies for 17 loci that differed between males and females in the populations a) M-12 and b) M-15.

5.9 Proportion of alleles across 17 microsatellite loci that occur at different frequencies, calculated per each sample pool. White bars represent samples from the IoW, patterned white and black bars indicate samples pooled into years for Mallydams, East Sussex. Coloured solid filled bars represent sampled populations from Binsted and Tilgate West Sussex and grey and black patterned bars represent samples pooled into years for Wakehurst Place, West Sussex.

5.10 Genetic diversity measures for each site a) shows mean allelic richness (AR) and b) shows mean expected heterozygosity (HE) with standard error bars.

5.11 Contours representing the joint likelihood of the causes of excess homozygosity: cryptic substructure (FST) and consanguinity, where Cg is the proportion of the sample inbred to degree g, which in this case is 0.0625 for the Mallydams Wood sample.

5.12 Bar plot showing the estimated membership coefficient (Q) for each individual from Mallydams wood. Each vertical bar corresponds to one individual and based on a priori of four populations.

5.13 Mean relatedness (r) within and between subpopulations designated by STRUCTURE for Mallydams wood.

5.14 ConStruct results for all individuals within Wakehurst Place, with g set at 0.0625.
5.15 Bar plot showing the estimated membership coefficient ($Q$) for each individual for the Wakehurst Place site. Each vertical bar corresponds to one individual based on three a priori populations. ........................................ 162
5.16 Relatedness within and between assigned STRUCTURE groups including all individuals for Wakehurst Place. ........................................ 163
5.17 Relatedness within and between assigned STRUCTURE clusters only including adults. 164
5.18 ConStruct results for adults only within Wakehurst Place, with $g$ set at 0.0625. .... 165
5.19 ConStruct results for Isle of Wight, with $g$ set at 0.0625. .......................... 166
5.20 Bar plot showing the estimated membership coefficient ($Q$) for each individual based on a priori three populations for the IoW site. Each vertical bar corresponds to one individual and each cluster has a different colour (subpopulation A: light green, B: dark green, M: mixed ancestry) ........................................ 168
5.21 Map displaying the location of each subpopulation as defined by the program STRUCTURE (a priori $K = 2$). Each woodland patch comprises: Big Wood; Sandpit Copse; Dunnage and Moorwood West and Upper Sheepwash. The black dots represent cluster A, the red dots cluster B and those of mixed decent are represented by split black and red. ........................................ 169
5.22 Mean relatedness scores for within (e.g. only including individuals cluster A) and between (e.g. including individuals from cluster A and B) the two subpopulations and the group of mixed ancestry. ........................................ 170
5.23 Mean relatedness scores for within (e.g. only including individuals cluster A) and between (e.g. including individuals from cluster A and B) the two subpopulations and the group of mixed ancestry. ........................................ 172
5.24 a) Mean $F_{IS}$ for males and females, both including and not including juveniles within Cluster A; b) Mean relatedness estimates for males and females, both including and not including juveniles within Cluster A. ........................................ 173
5.25 a) Allele frequencies for 17 microsatellite loci that differed between males and females in Cluster A ........................................ 174
6.1 (a) Bayesian phylogenetic analysis of CytB haplotypes with support values. Geographic origins are followed by haplotype number; (b) detail of minimum spanning network showing the number of mutational steps between haplotypes. .............. 194
6.2 Bayesian phylogenetic analysis of CytB haplotypes with support values, including those from Mouton et al. [4] and Combe et al. [5]. Geographic origins are followed by haplotype number ........................................ 195
6.3 Detail of minimum spanning network for the Central-Northern Europe clade and Turkey [4] showing the number of mutational steps between haplotypes. .............. 196
6.4 Bayesian phylogenetic analysis of D-loop haplotypes with support values. Geographic origins are followed by haplotype number, initials of the wood are used to represent the data collected for the current study: Wakehurst Place (WH), Tilgate Park (TP), Binsted (B) Mallydams (MW), Worth (W), Bexhill (BH) and Flatropers (FR) in Sussex, Isle of Wight (IoW), Herne Bay in Kent (HB) and Carmathan, Wales (CW).

6.5 Detail of minimum spanning network for current study D-loop haplotypes showing the number mutational steps between haplotypes.

6.6 Bayesian phylogenetic analysis of D-loop haplotypes with support values, including haplotypes from Combe et al. [5]. Haplotype number are followed by geographic origins, initials of the wood are used to represent the data collected for the current study: Wakehurst Place (WH), Tilgate Park (TP), Binsted (B) Mallydams (MW), Worth (W), Bexhill (BH) and Flatropers (FR) in Sussex, Isle of Wight (IoW), Herne Bay in Kent (HB) and Carmathan, Wales (CW).

6.7 Detail of minimum spanning network for British and the Lithuanian (Hap5) D-loop haplotypes, showing the number of mutational steps between haplotypes.

6.8 Regional distribution of D-loop haplotypes depicting the number of haplotypes found within each site. Initials represent site names: Briddlesford, Isle of Wight (IoW); Binsted wood, West Sussex (B); Tilgate park, West Sussex; Worth wood, west Sussex (W); Wakehurst Place, West Sussex (WH); Bexhill, East Sussex (BH); Flatropers, Rye, East Sussex (FR); Mallydams wood, East Sussex (MW) and Herne Bay, Kent (HB).

6.9 Isolation-by-distance relationship between geographical distance (km) and pairwise $F_{ST}$ between haplotypes for all UK D-loop samples.
## List of tables

1.1 Table taken from Mitchel – Jones and White [6] showing the progress at reintroduction sites by 2007. * Dispersal beyond area. ** Dispersal beyond site ........................ 7

2.1 List of microsatellite primers used, including the forward and reverse primer sequences, the annealing temperatures (Tm) and the allele size. ................................. 38

2.2 Manually calculated error rates for alleles, loci and genotypes. ................................. 39

2.3 Characteristics of 17 microsatellite loci used in study number of individuals typed per locus ($n$), number of alleles at each locus ($k$), observed heterozygosity ($H_O$), expected heterozygosity ($H_E$) and the estimated null allele frequency, computed across all populations and years. ......................................................... 40

2.4 Summary of the sampling sites, sample numbers and analysis methods used for each chapter ......................................................... 41

3.1 Results of the pairwise Mantel tests of percentage of shared TRFs correlated with pedigree relatedness including the Basic TRF dataset and the Consensus TRF dataset produced by both extraction techniques. * signifies significance (p-value < 0.05) . ........................... 66

3.2 For the basic dataset, mean number of TRFs and mean $H'$ scores for each category and the Kruskal - Wallace outputs for total TRFs and $H'$ scores correlated with the categories age, sex and year. ......................................................... 77

3.3 For only the TRFs produced by both extraction techniques (consensus dataset), mean number of TRFs and mean Shannon Index scores for each category and the Kruskal - Wallace outputs for TRFs and Shannon Index scores correlated with the categories age, sex and year. ......................................................... 78

4.1 Demographic summary of the two study populations with total sampled for each sex, sex ratio, density of each sex and total density ......................................................... 106

4.2 Shows the number and proportion of dormouse pairs for each category of $r$ ........................... 107

4.3 The proportions of nearest neighbour relatives for the Mallydams wood site (MW) . 115

4.4 The proportions of nearest neighbour relatives for the Isle of Wight site (IoW) . 115
4.5 Table showing the minimum allele frequency for each locus (MAF) and the reciprocal (1/MAF) for each site. The * highlights those loci most sensitive to inbreeding. 118

5.1 Pairwise $F_{ST}$ between years and generations for each site: a) IoW b) Wakehurst Place c) Mallydams Wood d) Tilgate Park e) Binsted. * indicate significant values after Bonferroni correction. 146

5.2 Pairwise $F_{ST}$ between years for each site: a) IoW b) Mallydams Wood c) Wakehurst Place d) Binsted e) Tilgate Park. * indicate significant values after Bonferroni correction. 146

5.3 Genetic characteristics of pooled samples hazel dormouse populations including sample number ($n$), mean number of alleles across 17 microsatellite loci ($A$) mean allelic richness across 17 microsatellite loci and standardised for a sample size of 9 loci ($AR$), mean observed heterozygosity ($H_O$), mean expected heterozygosity ($H_E$), multi locus estimates for $F_{IS}$ and the number of private alleles per population ($N_{AP}$). Populations with significant heterozygosity deficiencies are indicated by * and those with significant excess heterozygosity are indicated by † according to an adjusted (5%) nominal level calculated per group. 147

5.4 Pairwise $F_{ST}$ for IoW, East Sussex and pooled West Sussex populations. * indicate significant values after Bonferroni correction. 148

5.5 Pairwise $F_{ST}$ for all sites. * indicate significant values after Bonferroni correction. 148

5.6 Pairwise $F_{ST}$ for all Sussex sites. * indicate significant values after Bonferroni correction. 152

5.7 Pairwise $F_{ST}$ between all designated woodlands present in the Isle of Wight site. BM: Briddlesford copse mid; BN: Briddlesford copse north; BS: Briddlesford copse south; BW: Big Wood; D: Dunnage; MU: Moorwood West/Uppersheepwash and S: Sandpit Copse. * indicates $p \leq 0.05$. 167

5.8 Pairwise $F_{ST}$ between grouped woodlands within the IoW site. a) depicts samples pooled by woodland patch (Brid: Briddlesford Copse; BWS: Big Wood and Sandpit Copse; D: Dunnage, MU: Moorwood West and Uppersheepwash) b) depicts the smallest number of possible pools (DMU: Dunnage with Moorwood West and Uppersheepwash).* indicates significant after Bonferroni correction. 167

5.9 Pairwise $F_{ST}$ between subpopulations assigned by STRUCTURE and those with mixed ancestry.* indicates significant after Bonferroni correction. 168
5.10 Genetic characteristics of subpopulations designated by STRUCTURE: number of samples \( (n) \), mean number of alleles across 17 microsatellite loci \( (A) \) mean allelic richness across 17 microsatellite loci and standardised for a sample size of 9 genes \( (A_R) \), mean observed heterozygosity \( (H_O) \), mean expected heterozygosity \( (H_E) \), multilocus estimates for \( F_{IS} \) and the number of private alleles per population \( (N_{AP}) \). Populations with significant heterozygosity deficiencies are indicated by * according to an adjusted (5%) nominal level calculated per group.

5.11 Genetic characteristics of adults within subpopulations designated by STRUCTURE: number of samples \( (n) \), mean number of alleles across 17 microsatellite loci \( (A) \) mean allelic richness across 17 microsatellite loci and standardised for a sample size of 9 genes \( (A_R) \), mean observed heterozygosity \( (H_O) \), mean expected heterozygosity \( (H_E) \), multilocus estimates for \( F_{IS} \) and the number of private alleles per population \( (N_{AP}) \). Populations with significant heterozygosity deficiencies are indicated by * and those with significant excess heterozygosity are indicated by † according to an adjusted (5%) nominal level calculated per group.

6.1 Geographic location of sample locations, with number of samples obtained, haplotype number for each location and Genbank accession number, where relevant for both Cyt b and D-loop. *Accession number from Mouton et al. [4]. †Accession number from Combe et al. [5].

6.2 Genetic diversity measures for Cytochrome Bb and D-loop for the novel data for each of the main regions. UK 2017: all the novel samples; UK all: all sample from this study with the addition of those reported by Combe et al. [5]. Number of samples \( (n) \), the number of haplotypes \( (H) \), haplotype diversity \( (H_d) \) and nucleotide diversity \( (\pi) \), measures including standard deviations \( (SD) \), and neutrality test statistics \( (Tajima’s D (Taj D) and Fu’s F) \). Significant results \( (p<0.05) \) are depicted with *.

6.3 CytB AMOVA for Sussex and IoW

6.4 D-loop AMOVA for Sussex and IoW
Chapter 1

The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*)

1.1 The Hazel Dormouse *Muscardinus avellanarius*

1.1.1 The distribution and conservation status of *Muscardinus avellanarius*

The hazel dormouse (*Muscardinus avellanarius*) is a nocturnal arboreal mammal that mainly inhabits areas of deciduous forest with a thick layer of understorey, but has also been found in mixed deciduous, coniferous forests and hedgerows [7]. They are distributed across Europe, from the Mediterranean to Southern Sweden, eastwards towards Russia (excluding Iberia) and extends its range to parts of Asia Minor [8] (see figure 1.1). It has also recently been discovered in County Kildare Ireland, where it is considered to be a non-native recent arrival [9]. The hazel dormouse is relatively common and widespread throughout its range and as such, is listed as of least concern on the IUCN red list of threatened species; however, in parts of its Northern range (Britain, Netherlands, Sweden, Germany and Denmark) populations are declining and are protected in Europe under Appendix III of the Bern Convention and Annex IV of the EU Habitats and Species Directive [10].

The northwest limit of its European range is in Great Britain, where it is found over much of southern England, with isolated populations in northern Wales [7] and some scattered populations in the midlands and Northern England (see Figure 1.2). Although dormouse population numbers in Great Britain are unknown [3], they are considered
Fig. 1.1 Range map of the hazel dormouse [1]
1.1 The Hazel Dormouse *Muscardinus avellanarius*

to be under threat [7] and as such, are protected under the Wildlife and Countryside act 1981, the conservation (Natural Habitats) regulations 1994 Countryside and Rights of Way Act 2000, The Natural Environment and Rural Communities Act 2006 and are considered a priority under the UKBAP [7]. They are regarded as a flagship species for woodland conservation and their presence in an area is an indicator of high biodiversity [7].

1.1.2 Causes of the population decline and current conservation strategies

Populations of the hazel dormouse were once widespread in Britain, however the past hundred years has seen the extinction of these charismatic rodents from several English counties (see figure 1.2a and b). Further to these local extinctions, a 64% decline in dormouse numbers since the late 1970s has been noted [8]. Habitat loss due to increasing clearing of woodland for agriculture and urbanisation is considered to be the major contributing factor to the range contraction and population decline of the British dormouse [11][12][13]. Since 1930 more than 32,000ha of ancient semi-natural woodland (preferred dormouse habitat) has been lost from England and Wales, now only occupying 2-3% of the landscape [13]. The remaining woodland habitat are small and fragmented [12] and as a result, populations of the hazel dormouse are often small and isolated, making them vulnerable to local extinction [14]. Dormouse presence is more likely in woodlands greater than 20 ha in size, however woodland fragments smaller than this can support dormice when they are connected by hedgerows [12][13], as hedgerows provide both food resources and safe dispersal routes [13]. Therefore, the loss of hedgerows (around 150,000 km lost since 1945) [13] is a compounding factor to the decline in hazel dormouse numbers within Britain. Commercially important coniferous plantations have replaced much of the natural forests [12]. Coniferous plantations support fewer individuals per ha (1-3) than deciduous woodlands (up to 15.6), again resulting in reduced dormouse numbers [7].

Dormice need a continued supply of food throughout the active season to both gain weight post-hibernation to enable reproduction and to gain weight post-reproduction in order to survive hibernation [15]. As a result, dormouse habitat is highly diverse for plant species to fulfil this requirement [13]. These prerequisite conditions are provided by early successional woodlands (e.g. after storms, natural fires etc.) and as such, traditional methods of woodland management (e.g. coppiced and coppiced-with-
The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*)

Fig. 1.2 The hazel dormouse distribution a) Victorian dormouse distribution (data from Rope, G.T. (1885) On the range of the dormouse in England and Wales [2]) b) the current distribution of dormice in Britain as of 2013 [3]

standards) provided ideal conditions for dormouse populations [12]. These methods are no longer widely used in Britain augmenting the dormouse decline [15] [12].

Further threats to the continued survival of the hazel dormouse in Britain include climate change and competition from other species. Climate change has the potential to alter dormouse survival rates, activity patterns and the onset of breeding, together with timings of food plants [12] and as such, is considered to be a major future threat. Deer stuck browsing reduces habitat quality through the destruction of the understorey [12]. Increases in deer population numbers is widely attributed to impede dormouse regeneration to areas of high deer densities [12]. The noted range contraction and population decline of the dormouse prompted the initiation of a dormouse monitoring scheme in 1988 [16]. The National Dormouse Monitoring Programme (NDMP) (previously the national dormouse monitoring scheme) is a volunteer based, nationwide dormouse monitoring scheme, which utilises nest boxes in order to gather long-term data on dormouse abundance, annual variation in timing, breeding success and population densities in different habitats and geographical areas [12]. The scheme is managed by the Peoples Trust of Endangered Species (PTES) in collaboration with Royal Holloway, University of London and Natural England and covers around 400 sites with known dormouse occupation throughout Britain [16]. Through this monitoring scheme a national
average decline of 19% was detected between 1991 and 2000, however this trend was not universal. Populations in northern England suffered greater losses (about 40% between 1993 and 2000) than populations within southern England, where consensus numbers remained relatively stable [17]. As such, the scheme successfully identified areas that required immediate mitigation. In 2009 trend analysis of NDMP data showed the population decline appeared to be slowing, indicating that the concerted conservation effort was a success ([16]). However, Goodwin et al. [18] analysed the 22-year dataset (1993-2014) provided by the monitoring scheme and showed a population decline of 72%, which is the equivalent to a 5.8% mean annual rate of decline, a much more severe decline than previously thought. Indicating that long-term monitoring of threatened species is essential in order to detect an overall decline in numbers, rather than short term population fluctuations. The nest box surveys enable easy monitoring and has been found to increase the carrying capacity of sub-optimal woodlands (e.g. mixed deciduous and coniferous) by providing much needed nesting holes, thus enabling an increase in dormouse density [7]. The monitoring scheme has also been essential in raising public awareness of the plight of the hazel dormouse.

As discussed above, dormouse abundance is heavily influenced by habitat type and quality, with connectivity considered to be vital for their long-term survival [16]. Current conservation strategies in place to maintain and promote dormouse populations include both the protection of dormouse habitats (e.g. woodland and hedgerows) and sympathetic woodland management, such as coppicing over long periods, woodland thinning and deer fencing. To mitigate the local extinctions of the dormouse, a reintroduction programme was initiated in 1993 in order to restore dormice to areas of previous inhabitation, where the likelihood of natural re-colonisation was low [7]. In order to supplement fragmented wild populations of dormice, individuals are bred in captivity and their offspring released into areas deemed appropriate for reintroduction [7]. All individuals placed into the captive breeding programme are those that would not survive in the wild and have been taken under license for the purpose of the reintroduction scheme. Once taken, individuals are placed with specialist zoos or with trained volunteers [7]. The reintroduction programme is generally considered to be a success [6], as multiple reintroduction sites have reached the stage where reintroduced dormice have dispersed out of the area they were released and at a few sites dormice have dispersed into neighbouring woodland of the site (see table 1.1)
1.1.3 Life history of the hazel dormouse

The hazel dormouse is atypical of rodent species, as they have evolved in stable habitats and employ K-selected life history strategies [7]. In contrast to other small mammals, hazel dormice are found in low densities, between 1.75 - 15.6 individuals per hectare in optimal conditions [7], making them vulnerable to habitat loss. There is little information available on the life span of wild dormice, however four years has been recorded in the wild [19] and 5.3 years in captivity [7][13], making them relatively long-lived (for example woodmice (Apodemus sylvaticus) live on average 1.5 years). The hazel dormouse’s reproductive potential is low, with large interannual variation in reproductive success [7][20]. In most cases only one litter is produced per year with between 1-7 young, with occasional second broods if conditions are optimal [21]. Young become independent between 6-8 weeks [7] and within this time period it is believed that mothers teach their offspring foraging routes and navigation [13], thus displaying high levels of maternal investment. In comparison, wood mice, which prefer similar habitats, can produce between 4-6 litters per year of up to 9 young, become fully independent at 2-3 weeks [22] and therefore, display r-selected strategies, typical of most small mammals.

Hazel dormice are seasonal hibernators, making them very unusual amongst small mammals. The length of hibernation depends on age, sex and geographic location [10]. In Britain the hibernation period is approximately between October and May [23]. In order to hibernate, the dormouse comes to the ground to benefit from the stable soil temperatures and moisture [13]. As a result, they can suffer high winter mortalities due to their vulnerability to predation and disturbance, for example flooding during this period, which can wash them from their hibernation nest [21]. Hibernation is thought to be a major limiting factor of the reproductive output of the dormouse, as the active season is limited and as such, there is a limited time in which to breed [7]. Weight gain is essential to obtain a suitable condition for reproduction and to maximize the chances of surviving hibernation; thus much time during the active season is dedicated to recovery from the previous hibernation and preparation for the next [13], limiting the number of broods possible within this time frame. The hazel dormouse, therefore, is best suited to stable habitats, where low reproductive output and low recruitment are possible. However, this specialisation makes them vulnerable to environmental change, such as habitat fragmentation.
<table>
<thead>
<tr>
<th>County/Site</th>
<th>Year</th>
<th>Feeding</th>
<th>Young Born</th>
<th>Overwinter</th>
<th>2nd Generation</th>
<th>Pop. Growth</th>
<th>DBA*</th>
<th>DBS**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridgeshire 1</td>
<td>1993-4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>−</td>
</tr>
<tr>
<td>Nottinghamshire</td>
<td>1994-5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cheshire</td>
<td>1996-7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Buckinghamshire</td>
<td>1998</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Warwickshire</td>
<td>1998</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>?</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>North Yorkshire 1</td>
<td>1999</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Suffolk 1</td>
<td>2000</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bedfordshire</td>
<td>2001</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cambridgeshire 2</td>
<td>2001</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>Staffordshire</td>
<td>2002</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Lincolnshire</td>
<td>2002</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Derbyshire 1</td>
<td>2003</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>?</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>North Yorkshire 2</td>
<td>2004</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Derbyshire 2</td>
<td>2005</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Suffolk 2</td>
<td>2006</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>North Yorkshire 3</td>
<td>2008</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1.1 Table taken from Mitchel – Jones and White [6] showing the progress at reintroduction sites by 2007. * Dispersal beyond area. ** Dispersal beyond site
1.2 Genetic variation

It is widely accepted that for a population to persist for the long-term, genetic variation within the population is needed in order for the population to respond to natural and anthropogenic change through the process of evolutionary adaptation [14]. Genetic diversity is generally, though not always, greater in large populations and maintained in the long term by mutation and the short term by gene flow [14]. Populations with low numbers, such as the hazel dormouse, often have low genetic diversity and, as a consequence, are more susceptible to extinction. Such populations present fewer opportunities for beneficial mutations to succeed, given that positive natural selection is weak in small populations, they are less likely to adapt to changing environments (biotic and abiotic) [24]. Hence, measuring genetic variation within populations will highlight those populations most under threat from changing conditions, known as the lowest evolutionary potential [25]. The maintenance of genetic diversity requires sufficient individuals for the influence of random genetic drift, which is the process that removes diversity at a rate proportional to the population size, to be negligible. Therefore, the primary concern in conservation, including that of the hazel dormouse, is the increase and maintenance of population size that ideally minimises inbreeding in order to maximise genetic diversity and as a consequence population fitness.

Animals have evolved a number of mechanisms to avoid inbreeding and consequently maintain genetic diversity, such as sex-biased dispersal, multiple mating and kin avoidance [26]. Populations isolated through habitat fragmentation may suffer a reduction in dispersal, due to a reluctance to cross an inhospitable matrix [27], (such as agricultural fields surrounding woodlands) thus increasing the likelihood of encountering relatives and reducing the number of suitable mates. These inbreeding avoidance mechanisms are fundamental to the survival of populations, although they are difficult to measure directly. The logistical constraints of traditional field techniques, for example those requiring direct observation of individuals, often limit data collection and lead to biased estimates of juvenile dispersal [28]. Direct observational methods are also insufficient to detect long-distance migration, rare episodes of mass migration and rare instances of population extinction and recolonization can occur [29]. The use of genetic markers can infer patterns of dispersal and gene flow from population genetic structure and detect the presence of immigrants [30]. Thus, genetic methods can provide more detailed information that is particularly useful for species like the hazel dormouse, which are found in low densities, are difficult to observe and have a relatively high protection status, where gaining permission for direct observational study by handling is difficult.
Traditionally, genetic diversity is measured using heterozygosity (e.g., the presence of two different alleles at a genetic locus) and broadly considers three different hierarchical levels: variation within individuals, variation among individuals in the same subpopulation and variation within the total population [31]. This hierarchical approach to exploring the population dynamics of a species has a long theoretical history, specifically with the use of F-Statistics as a means of describing it [32] and is now the mainstay of molecular ecology [14]. In essence, this hierarchical partitioning of the genetic diversity of a population extracts information on the magnitude of inbreeding within individuals and subpopulations, which offer indirect insights into the breeding and migratory behaviour, which is essential for effective conservation intervention (e.g., creating migratory routes and transplanting individuals). The recent development of genetic markers for the hazel dormouse [33][34] now allows this hierarchical approach to studying this species.

1.2.1 Processes influencing genetic diversity

Genetic variability within a population provides the means for adaptation to changing environments, where high levels of diversity are considered healthy and low levels are seen as a threat to a species ability to respond to environmental changes [35][36]. When intending to measure population genetic variability, it is necessary to understand the natural process that affect allele frequencies. A population’s genetic variability is balanced by opposing forces of gain and loss; gain is through mutation and gene flow from neighbouring populations and loss is through genetic drift and natural selection, with the rare exception of overdominance [35][37]. These forces do not work in isolation and are intertwined with species-dependant life history strategies that are determined by environmental pressures. No two populations are the same and must be studied within the context of the species ecology and habitat. Thus all aspects contributing to the genetic variation of a species must be taken into account. The following will discuss both non-adaptive and adaptive processes, which influence population number and diversity at the level of a single population, all of which are likely to be relevant to our understanding of the hazel dormouse populations.

Genetic drift is the process that explains the stochastic change in allele frequency from one generation to the next [37]. This process occurs for two reasons. The first is because reproductive success within a population is variable: some individuals will produce more offspring than others and some individuals will not reproduce at all, for reasons independent of the natural selection of fit phenotypes. This results in some parental alleles being over represented in the following generation, and others under
The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*) represented. The second is that the offspring generation results from what is effectively the random assortment of the parent’s gametes: heterozygous loci (e.g., *Aa*) are divided into haploid sperm and eggs (*A* and *a* bearing gametes). Which fertilizes which, in order to re-establish the diploid state (*AA*, *Aa* or *aa*), is effectively random. As a result, not all alleles within a population are necessarily sampled for the next generation [37]. Genetic drift is thus responsible for random micro-evolutionary change at the level of the allele frequency and results in non-adaptive evolution. The ultimate result of drift is that an allele at any given locus will either reach fixation or loss, both resulting in a reduction in variation at the locus. The rate of this loss can be measured as a loss of heterozygosity and is proportional to the population size. The reason being that, if the frequency of a particular allele is *1/2N*, where *N* is the population number, then the probability of drawing two copies of the same allele (the probability of homozygosity) is *1/2N* and the probability of drawing two copies not identical (the probability of heterozygosity) is *1 − 1/2N*. Given this, it can be shown that heterozygosity declines over time at a rate:

\[
H_t = H_0 \left(1 - \frac{1}{2N}\right)^t \tag{1.1}
\]

Where \(t\) = the number of generations, \(H_0\) = the frequency of heterozygotes in the starting generation, \(H_t\) = the frequency of heterozygotes in generation \(t\) and \(N\) = the population size [38].

Natural selection, on the other hand, is the process responsible for adaptive evolution. Natural selection explains the genetic diversity of genes responsible for adaptive traits under conditions of large *N*. It therefore does not explain diversity at unlinked neutral loci, such as the microsatellites used in this study. Because the magnitude of change due to drift is a function of *N* (Eq. 1.1), it is generally believed that in large natural populations, drift plays only a minor role in shaping short term allele frequency change and the effects of drift will be most profound in small populations, where it will drive each allele to fixation or extinction and hence reducing diversity in a time period proportional to the population size [37].

### 1.2.2 Genetic markers

In order to measure processes such as genetic drift, genetic markers are required. Genetic markers come in a variety of types and are used to provide information on many aspects of a species’ behaviour of interest to an ecologist, such as quantifying genetic diversity, estimating population size, inbreeding levels, relatedness and past
1.2 Genetic variation

and present migration patterns [39][40][41]. Genetic markers can either be neutral or adaptive (non-neutral), depending on whether the marker is linked to a gene under selection or is unlinked and hence having no (or little) effect on fitness [42]. The choice of marker will depend on the question of interest. Neutral markers are commonly used to investigate genome-wide effects, such as genetic drift, population bottlenecks, gene flow and inbreeding. Only genome-wide effects inform us reliably about population demography and phylogenetic history [43]. Adaptive markers can be used to assess the adaptive potential of populations or species [42], such as local adaptions caused by an ecological gradient [44].

In order for a genetic marker to be used as a tool for detecting demographic patterns, it must be fundamentally selectively neutral and follow Mendelian inheritance [40], thus population studies commonly employ neutral genetic markers to study genetic diversity at multiple levels. In principle, neutral loci across the genome will be similarly affected by demography and evolutionary history [43] and as neutral variation is the result of non-selective forces (e.g. genetic drift and migration), it can be used to understand the effects of such forces on populations [45]. This is especially important when studying endangered species, which generally have small population numbers, and where non-selective forces, such as drift and migration, have a greater effect on allele frequency than selection [46]. For the purposes of this investigation, therefore, neutral genetic markers are considered as the most appropriate.

Selecting the correct molecular marker will depend on the area of interest. As the purpose of this investigation is to successfully establish whether the hierarchical population structure of the hazel dormouse informs us on the species, mating behaviour and social structure, the use of multiple neutral genetic markers is necessary.

The degree of polymorphism (how variable the genetic marker is), is proportional to the underlying rate of mutation at these markers [47]. Thus, when measuring recent changes (10-100 generations), present day demography, connectivity patterns and relatedness, high mutation rates and thus high levels of variation, are preferable [40]. Whereas, when trying to establish historical barriers to gene flow or recolonisation since, for example, the last ice age, lower mutation rates are likely to be more informative and different markers, such as mitochondrial markers are required [40]; in other words, it is necessary that mutations do not have time to mutate back to their previous state, resulting in a loss of information.
1.2.3 Microsatellites

Microsatellites are co-dominant nuclear markers with short tandem repeats of 2-6 nucleotides (e.g., GAC) found throughout the nuclear genome of most taxa [40][47][48]. The key feature of microsatellites is their high mutation rate (estimated $10^{-2} - 10^{-6}$ mutations per locus per generation), which makes them one of the most variable types of DNA sequence [47][49]. The polymorphism found at microsatellites derives from variability in length rather than in the primary sequence [47]. Their hypermutability results in higher allelic diversity and thus, high variation within the same gene or locus [40]. It is this variation that makes them a popular choice in population and conservation genetics, as they have the resolving power to give individuals a unique identification tag (genotype) with the use of only a few loci (typically fewer than 10), estimate relatedness and distinguish between high rates of migration and panmixia, thus allowing insights into fine-scale ecological questions [40]. This is extremely important when studying endangered species, as small or recently bottlenecked populations may be largely invariant at loci with lower mutation rates [40]. This together with their relative ease of analysis [47] and their availability for the hazel dormouse [33][50] makes them an excellent choice for investigating levels of inbreeding, relatedness, population structure and current patterns of movement within and between sub-populations.

1.2.4 Mitochondrial DNA

Mitochondrial DNA (mtDNA) is a small, circular, homologous DNA molecule found within the mitochondria [37]. Mitochondrial DNA is a popular genetic marker in population genetics and phylogeography for a number of reasons, such as the ease of amplification (there are multiple copies of the mitochondrial genome per cell) and conserved arrangement of genes: Mitochondrial DNA is maternally inherited without recombination. This lack of recombination and effective clonal inheritance means all genes share the same history [51]. Thus, even with universal primers and small amounts of DNA, reliable information can be ascertained ([37][51]. Mitochondrial DNA has a relatively high mutation rate of $5.7 \times 10^{-8}$ substitutions per site per generation, which is around 10-fold the average rate of synonymous substitutions per generation in protein coding nuclear genes [37], but lower than that of microsatellites. This slower rate of mutation requires more time for variation to build up between individuals and so is of more use when studying multiple genetic lineages between populations that share a more distant common ancestor and is, therefore, an effective marker for studying broad-scale population structure [52]. The frequency of mtDNA haplotypes are largely
governed by migration and drift, rendering it particularly useful for detecting past bottlenecks, founder events and introductions.

As the mutation rate of MtDNA is slightly lower than that of microsatellites, mtDNA will be used here to study the genetic variation of the hazel dormouse at higher spatial scale and to establish movements that could be used to explain the patterns of genetic variation found today.

## 1.3 Hierarchical approach to conserving species

### 1.3.1 Individual level of genetic diversity

The suggestion of the existence of social interaction between individual hazel dormice [23][13][10][53] and the low dispersal ability of the hazel dormouse [21][13] indicate a high possibility of individuals encountering relatives throughout their daily existence. Mating of close relatives causes the erosion of genetic variation at the level of the individual. It is well documented in captive populations that the loss of genetic variation through inbreeding has detrimental effects on growth, development and survival, for example the wild wolf (*Canis lupus*) [54][55]. Inbreeding increases the expression of deleterious recessive alleles (due to increase in homozygosity) and causes the loss of heterozygotic advantage and, thus, causes a decline in mean fitness [56].

Many studies have found correlations between the heterozygosity at several microsatellite markers and a measurable fitness trait, such as reproductive success and survival, which is known as a heterozygosity- fitness correlation (HFC) [57]. HFCs have been observed for decades [58] and have been detected in numerous taxa, including the house mouse (*Mus musculus*) [59], Egyptian vultures (*Neophron percnopterus*) [60] and Dunlins (*Calidris alpina*) [61]. However, the relevance of HFCs are often questioned, as most investigations have been limited to highly inbred, small populations and may not translate to larger panmitic populations. As found in a natural population of great tits, *Parus major*, where overall microsatellite heterozygosity did not correlate with the measure of brood size or reproductive fitness [62]. Common Buzzards, *Buteo buteo*, also showed no correlation between fitness measures and heterozygosity [63].

A further issue is the normally small number of loci used to measure heterozygosity, which represents a tiny fraction of the genome and thus, it is argued, the heterozygosity found accounts for very little of the total fitness variation [64]. Also, small panels of microsatellites may not reflect general inbreeding depression, as a handful of markers provide very little power in detecting inbreeding [64]. Contrary to these studies,
Forstmeier et al. [65] found that a panel of 11 highly variable microsatellites provided as much power to detect HFCs as 1,359 single nucleotide polymorphisms (SNPs) in captive, moderately inbred Zebra finches, *Taeniopygia guttata*. The authors also stated that the potential for microsatellite markers to detect HFCs and inbreeding is greater than previously reported. Therefore, the outcomes of such studies appear to depend on the chosen loci and the species in question, where the greatest predictors of whether HFCs are likely to be observed are the variance in inbreeding within the population and the magnitude in heterozygosity [57]. Captive populations provide an excellent opportunity to study the process of inbreeding depression, as the most accurate measures of inbreeding derive from known pedigrees [66]. Due to the difficulties in obtaining detailed pedigrees for wild populations [60], molecular markers, including microsatellites, will continue to offer an alternative to estimating inbreeding through the measure of heterozygosity deficits.

1.3.2 Effects of habitat fragmentation on individual level genetic diversity

Habitat loss is likely to result in a decline in population size and, if severe enough, the population may encounter what is termed a population 'bottleneck' [54]. A population bottleneck is an event that drastically reduces the size of a population long enough for drift to decimate the genetic variation. This event can be natural, such as a flood, or disease outbreak or it can be anthropogenic, such as habitat destruction. Human induced fragmentation of a habitat (via road building and modern farming practices) will lead to patches of unconnected habitat. Such loss and fragmentation often result in decimated and isolated populations, where there are restricted mating opportunities and as such, an increased likelihood of randomly mating with a relative and thus, inbreeding [61]. If the population remains isolated, genetic erosion occurs through random inbreeding, which is manifested as random genetic drift. In large populations, say in the many thousands, the random choice of a mate will typically result in the union of non-relatives. As the population size decreases, the probability of a random union between relatives increases: random inbreeding, with the subsequent loss of heterozygosity. Even if the population size remains constant, it is expected to lose heterozygosity each generation at a rate proportional to $1/N$, where $N$ is the population size (see Eq. 1.1). An increase in relatedness between individuals can lead to the disruption of sex biased dispersal and interfere with mate choice, making inbreeding even more likely [67], thus compounding the problem. The resulting loss of
genetic variation will ultimately end in the fixation of alleles and the accumulation of deleterious mutations [61], which is known as inbreeding depression. A population that is neither large nor randomly mating is, therefore, not expected to be in Hardy-Weinberg equilibrium. For example, the expected frequency of homozygous (e.g., AA and aa) and heterozygous (e.g., Aa) genotypes in large randomly mating populations is $p^2$, $q^2$ and $2pq$ respectively, where $p$ and $q$ are the respective allele frequencies (see Chapter 2 for general description of Hardy-Weinberg). This assumes that the alleles have independent ancestries: they are not identical by descent. Once a population becomes small, there is a real chance that relatives encounter each other. Relatives share copies of alleles that are identical by descent, say two copies of a parental allele, A. Their offspring could be homozygous AA either through the inheritance of two A alleles that are identical by descent from a single ancestral allele (hence with frequency $p$) or through the inheritance of two A alleles not identical by descent (IBD) from two ancestral alleles (hence frequency $p^2$)[68]. The likelihood that the offspring is homozygous due to the inheritance of alleles IBD depends on how closely related the parents are, which is often referred to as $f$. For the offspring of sibs, for example, $f = 0.25$: the parents share, on average, half of their genome in common, half of this is lost each generation, hence the inbreeding coefficient of the offspring is half the relatedness of the parents. Deviations from Hardy-Weinberg equilibrium (HWE) due to inbreeding are captured by these two issues: the frequency of homozygosity = that due to inbreeding ($f$) and that due to random mating $(1 - f)$, hence the frequency of a homozygous genotype $\text{prob}(AA) = (1 - f)p^2 + fp$; the frequency of heterozygosity = that solely due to the remaining randomly mating proportion (heterozygotes are alleles not IBD), hence $\text{prob}(Aa) = (1 - f)2pq$. When considering IBD within individuals, $f$ is referred to as $F_{IS}$, which is one of Wright’s F - statistics for measuring loss of genetic diversity within hierarchies [32]. An individual’s homozygosity or heterozygosity at microsatellite markers can therefore be used to estimate $F_{IS}$ as the excess of Individual homozygosity relative to that expected within the Subpopulation [69].

In a population suffering from inbreeding depression, such as the wild wolf, *Canis lupus*, individuals show visible physical signs, for example malformed spines and kinked tails [70]. These malformations greatly reduce the chances of an individual being fit enough to reproduce, or even survive, which may tip the population into "mutational meltdown". Mutational meltdown is the process by which harmful mutations accumulate in small populations, leading to a loss of fitness and a further decline of population size [14]. Once a population goes into mutational meltdown, survival to adulthood decreases and sterility increases, thus the population number begins to quickly drop
and can enter an extinction vortex, due to the fixation of deleterious alleles via genetic drift. This is why habitat loss and fragmentation are considered to be the single biggest threats to biodiversity, as they have a potentially devastating impact on both individual and population fitness [71][24]. Indeed, a number of studies have found evidence for habitat loss and fragmentation having a profound effect on individual fitness and levels of genetic diversity within threatened populations [72][73][74].

The hazel dormouse is naturally found in low population densities [13] and thus the compounding effect of habitat loss and fragmentation makes inbreeding a very real threat to local populations. Naim [75] found that in Wales, populations within fragmented, patchy woodland had relatively lower genetic diversity and higher levels of inbreeding compared to populations in the continuous woodland sampled, emphasising how information on inbreeding and diversity levels can be used to highlight which dormouse populations could be at the greatest risk of local extinction.

1.3.3 Social structure and genetic diversity

We saw with inbreeding that alleles with recent common ancestries (alleles that are IBD) can come together within individuals, increasing homozygosity within the individual and result in individual fitness loss - inbreeding depression [56]. However, if individuals have limited dispersal, alleles with recent common ancestries can also come together in subpopulations, or social groups, relative to the total population. The population is then "structured", which can come about either through physical barriers to random gene flow or due to behavioural constraints upon dispersal. When behavioural constraints impinge upon gene flow, such as kin-avoidance during the mating season, then the population structure is referred to as 'social' structure. Social interactions, such as kin-recognition/avoidance have a profound influence on the dynamics and persistence of a population [31]. Kin recognition/avoidance, in turn, influences dispersal mechanisms and consequently the relatedness of group members, which can bear directly on the survival of individuals [31]. Social structuring thus shapes both local and whole population level structure [76]. As such, the spatial distribution of genetic variation provides insights into the patterns of kinship and social structure [77] and vice versa. It is therefore valuable to consider the sociality of a species when regarding its conservation.

The social behaviour of the hazel dormouse is largely unknown. The arboreal and nocturnal nature of the dormouse makes them difficult to observe in the wild and much information comes from brief snapshots of individual interactions during nest-box surveys. Encounters such as these have lead to the impression that social organisation
varies throughout the active season [10]. Same sex cohabitation of adults within nest-boxes is found outside the breeding season [13], peaking just after hibernation and falling once adults reach breeding condition [10]. The occurrence of groups of dormice in nest boxes suggests a social dimension to their behaviour [23]. The recent development of genetic markers has provided insights into the behavioural ecology of the hazel dormouse. Naim [53] found a number of indicators of social behaviour and cooperation. In the first instance, evidence was found for crèching behaviour in related females; where females related as half-sibs were found to nest with young of different ages, hence possibly displaying some level of social cooperation and raising the question of kin recognition. Territoriality during the breeding season has been observed by both males and females, implying the possibility of dominance behaviour [10]. Naim [53] also found that litter size positively correlated with female weight and the number of sires, which she suggested could be a sign of female dominance and reproductive suppression. However, it was also stated that this could be a simple function of the age of the individual and increased attractiveness to mates. In the same study, possible evidence of alloparenting was found, where male dormice, who do not generally engage in active parenting [13] and were not the father of any of the offspring being parented, were observed. However, other reasons such as mate accession could also explain this observation. Naim’s study [53] raised many questions, but indicated some form of social structure, where relatives (at least during the breeding season) are tolerated and engage in social cooperation.

The benefits gained from sociality on an individual level can be difficult to determine, when immediate gains of social living, such as crèching behaviour, may seem to be outweighed by the costs, such as the increased risk of inbreeding. Kin selection theory states that individuals can increase their fitness indirectly by cooperating with kin, known as inclusive fitness [78], and explains why most social groups are formed from closely related individuals. The costs and benefits that define inclusive fitness through social behaviour are difficult to specifically measure in animals; however most research into the fitness gains for individuals have focussed upon the reproductive success of philopatric females in social groups, largely because reproductive success is easier to measure in females than in males [79]. A number of studies have focussed on female house mice (Mus musculus), where females are known to form crèches. These studies have found that female house mice preferentially nest with related individuals, as they appear to prefer those with more similar MHC (major histocompatibility complex) alleles, which correlates with genetic relatedness [80]. The MHC is a large multigene family found in all vertebrates; it is highly polymorphic and it is essential for the
functioning of the acquired immune response. It has been found that females housed with sisters will nearly always produce offspring, whereas females in groups of non-relatives often fail to reproduce and those offspring that are produced are less likely to survive to weaning [81].

It is possible that a similar mechanism has evolved in the hazel dormouse, as individuals that participate in cooperative acts such as crèching behaviour are expected to be able to recognise kin from non-kin and, possibly, even identify degrees of relatedness [82]. This has been shown to be true in a number of species, such as sweat bees (LasioGLOSSum Zephyrum), Belding’s ground squirrel (Urocitellus Beldingi), golden hamsters (Mesocricetus auratus), rats (Rattus norvegicus) and humans [83][84][85][86][87][88]. Hepper [85] demonstrated that the brown rat has the ability to recognise multiple kin classes, including paternal half-sibs and cousins, thus indicating some ability, believed to be olfactory, that distinguishes genetic relatedness. Kin recognition using a genetic identifier, such as MHC-influenced odour, would also help to avoid inbreeding, as those living in close proximity to relatives face the increased risk of mating with kin.

Odour is an important communicator in mammalian species and can be influenced by diet, the environment or be synthesised by the animal themselves. There is growing evidence to suggest that most odours produced by mammals are actually the byproducts of symbiotic bacteria [89]. It is thought, for example, that the MHC, by influencing which of the odour-producing bacteria are tolerated by the host, is a good candidate for a genetic identifier [90][91]. Scent marking is common in mammals, where urine, faeces and scent-gland secretions are all commonly used as a method of conveying information. This information can include individual identity, genotype (e.g., MHC) and social group membership; in fact, theoretically, there is no limit to the information conveyed [92][93][89][94]. All mammals have scent glands, which, depending on their location on the body, can provide the perfect living conditions for symbiotic bacteria as they are typically warm and moist. The glands within these regions produce ample fatty substances and organic materials for bacteria to feed upon [95][96]. It is the breakdown of these substances that releases volatile odourants that are responsible for the animal’s individual scent. Bacterial surveys of 14 mammalian species’ scent glands found them to harbour diverse bacterial communities, but importantly including fermentative anaerobes [89] and thus offering a potential mechanism for kin discrimination.

The relationship between scent-gland bacteria and odour is referred to as the fermentation hypothesis of chemical based recognition [95][97] which contains two parts:
1. Symbiotic bacteria living in scent glands produce volatile odourants and contribute to mammalian recognition cues.

2. Recognition cues among animals caused by variation in composition structure of bacterial communities inhabiting the glands.

Thus, individual recognition could be based on signature odours produced by bacterial communities unique to the individual, and so bacteria may be the underlying mechanism of kin recognition. Evidence for this has been found in the spotted hyena (*Crocuta crocuta*), where females were found to have group-specific odours and scent-gland bacterial communities were more similar in hyenas from the same clan, than those from different clans, and thus conveying information about group associations [98].

The MHC plays a crucial role in disease resistance, which means that it plays a role in determining which bacteria and how many can survive within the scent gland, which may change due to subtle differences between an individual’s immune system [89]. Evidence of this has been found in the Eurasian badger (*Meles meles*), where the bacterial communities of cub scent glands were found to be more diverse, with different dominant bacterial species, to that of adults [94]. Because a young animal’s immune systems has not fully matured, it was suggested that they acquired a mature colony more attuned to the individual’s immune system, and hence less diverse, as the animal aged. Differences between the bacterial communities of breeding and non-breeding females were also noted, suggesting hormonal changes could affect scent gland bacterial communities [94]. The likely relationship between the MHC and bacterial communities of the scent gland suggests that the bacterial communities of closely related individuals show greater similarity than those of unrelated individuals. This mechanism could ensure that individuals that cooperate do not reduce their inclusive fitness by engaging in such behaviour with non-kin. The recognition of kin could also shape mating systems, whereby the group display non-random mating and select mates through an ability to discern relatedness.

Relatedness between individuals can be measured using information from multiple microsatellite markers. Allele frequency data can be used to calculate the coefficient of relatedness (*r*), which, is the measure of the probability that an allele found in an individual will also be present in a relative using the equation, generally:

\[
r = \frac{\sum (p_y - p)}{\sum (p_x - p)}
\] (1.2)
Where for each allele, $p$ is the frequency within the population, $p_x$ is the frequency within the focal individual and $p_y$ is the frequency within the individual whose relationship to the focal individual is under question.

If the hazel dormouse does employ kin discrimination mechanisms, it may mean the population is more resistant to inbreeding, as they would actively avoid mating with relatives. However, in small populations, inbreeding avoidance through kin discrimination may be coupled with a reduction in mating, due to the lack of suitable candidates [99][100], which may further reduce population size.

Local genetic structure can be generated by social behaviour, where social structure results in non-random mating and close spatial associations between relatives [101][76]. Generally, group living mammalian species display female philopatry and male biased dispersal [102], evidence of which has been found in the hazel dormouse [75]. At the local population level, philopatry and maternal ancestry of group mates can subdivide the genetic structure [76]. This subdivision results in increased within-group relatedness and thus in positive local genetic structuring, whereby relatedness between individuals decreases with increasing genetic distance [101]. If territoriality, which can repulse competitive individuals, or social barriers are sufficient to restrict gene flow between the groups, groups become effectively isolated and the rate of genetic drift becomes augmented, resulting in genetic differentiation between groups [76]. As was done to measure inbreeding within individuals, inbreeding within Subpopulations relative to the Total population, referred to as $F_{ST}$, another one of Wright’s $F$ - statistics [32] is a consequence of homozygosity being more probable within subpopulations due to the higher frequency of alleles that share a more recent common ancestry within the subpopulation, relative to the whole, metapopulation. As before, the frequency of a homozygote, e.g., $prob(AA) = (1 - F_{ST})p^2 + F_{ST}p$, and the frequency of a heterozygote, $prob(Aa) = (1 - F_{ST})2pq$. The difference, however, is that the allele frequencies are the average of the total population and $F_{ST}$, rather than being a measure of increased homozygosity within individuals due to IBD is a measure of increased homozygosity within subpopulations due to IBD. $F_{ST}$ is, then, a measure of the variance in alleles between subpopulations, typically standardised by the maximum allelic variation:

$$F_{ST} = \frac{\sigma^2(q)}{q(1 - q)}$$

where $q$ is the allele frequency in the total population.

Although female philopatry and male biased dispersal have been demonstrated in the hazel dormouse through local genetic structuring amongst females, there has yet been no evidence of significant differentiation between sub-populations [75], suggesting
there is sufficient dispersal between family groups to prevent differentiation and perhaps that the social grouping of female dormice can be flexible.

Patterns of reproduction in a group will influence the genetic patterns of the subpopulation and population [31]. Mating systems are determined by the structuring of resources, which largely determine reproductive strategies in species. By determining who passes on their genes, mating systems influence the rate at which diversity is lost from a population over time: the reduction in the effective population size, $N_e$ [103]. The pattern of reproductive behaviour is typically driven by the investment of each sex in the production and raising of offspring [104]. This polygynous mating system, whereby males invest less energy than females in the resulting offspring, is typical for mammal species. Many aspects of dormouse ecology predict this 'typical' polygynous mating system, such as the lack of paternal care [13], the overlapping of male territories with multiple females [7] and the crèching behaviour of related females. However, female multiple mating, with the potential of sperm storage, has also been demonstrated to varying degrees in dormouse populations in Wales, Devon and Cornwall [53][50].

Genetic analysis has allowed mating systems to be studied in many species, generating increasing evidence that animals are more promiscuous than presumed on the basis of behavioural observations [53]. Multiple mating appears common in species producing more than one offspring per reproductive event and may have substantial influences on sexual selection and genetic structuring of demes [105]. In dormice, larger litter sizes have been associated with a higher proportion of multiple paternity [53]. Such reproductive strategies may serve as a way of increasing genetic diversity, especially in species like the hazel dormouse, known to typically have small population sizes and limited dispersal abilities, possibly explaining the higher than expected genetic diversity found by Naim [75].

1.3.4 Effects of habitat fragmentation on social structure

Habitat fragmentation can disrupt dispersal patterns of within-population processes, such as mating systems, social organisation and kin interactions [26]. These processes are likely to impact upon reproductive success and therefore are fundamental to population declines and extinctions [26]. Reduced dispersal may be caused by a species reluctance to cross an inhospitable environment [27], such as open fields in the case of the hazel dormouse [106]. Reduced dispersal can lead to the retention of offspring, an increase in population density and an increase in inbreeding levels [26].
A major factor leading to group cohesion, or social structure, and hence population structuring, is the retention of offspring in the natal territory [107]. Such retention is thought to be caused by ecological constraints, such as habitat saturation, environmental harshness and unpredictable conditions, which limit the likelihood of the successful establishment and breeding through dispersal to a new area [107]. Such constraints could be artificially induced by habitat loss and fragmentation, as the spatial distribution of resources is altered [71].

Changes in population density could increase competition over limited resources, potentially leading to reproductive suppression [108][31]. This increase in competition may then facilitate tolerance towards kin over non-kin and lead to cooperative behaviours [26]. Should it be the case that an individual is reproductively suppressed, they can, for example, increase fitness indirectly through aiding non-descendent kin (i.e., their inclusive fitness). Reproductive suppression leads to reproductive skew and this has been proposed as an important mechanism in stabilising group cooperation [109]. Therefore, habitat fragmentation could have a profound effect upon relatedness structure and could cause a shift in kin interactions [26]. Few studies have addressed the genetic and behavioural changes associated with habitat fragmentation. Female hairy nosed wombats (Lasiorhinus krefftii) normally disperse out of their natal territory and avoid associations with other individuals. In isolated populations, Walker et al [26] stated that an increase in spatial clustering, increased association of female relatives and an increase in inbreeding avoidance, display how habitat fragmentation can impact upon social organisation, which in turn is expected to alter genetic structuring (e.g., $F_{ST}$) and diversity levels (e.g., heterozygosity).

1.4 Regional scale of genetic diversity

Dispersal of individuals is an important determinant of population dynamics, patterns of gene flow and consequently population genetic structure [110][30]. Individual dispersal is also essential for maintaining the genetic viability of species (assuming high genetic diversity predicts species longevity). Species disperse for a number of reasons, such as to acquire resources, avoid competition and to prevent inbreeding [111][110]. Mammals generally exhibit sex-biased dispersal, the direction of which influences mating strategies [31]. For example, if males defend females and their resources and compete for female access, male-biased dispersal is expected [110][31]. Dispersal can vary in rate, distance and timing, largely depending upon the reason for dispersal (i.e. natal or breeding) [110]. Short-distance dispersal may be enough to prevent inbreeding, but longer distances
may be required to escape overcrowding and to colonise new territory [110]. Thus, the dispersal abilities and behavioural characteristics of species influence the scale at which spatial genetic structuring develops [112]. For instance, female philopatry may increase spatial genetic structure among females relative to males [113].

Populations of dormice, like many species, are typically non-homogenous entities, living at a scale at which movement requires considerable time and effort and is interrupted by geographical topography. The spatial distribution of many species, therefore, is fragmented with interrupted patches of suitable habitat [35]. This can disrupt gene flow, generating subpopulations (or "demes"), within a meta-population, that can vary at the genetic level as a consequence. The effects of this fragmentation on the spatial and genetic structuring of a population will depend upon population size, the species’ dispersal abilities and the penetrability of the inhospitable habitat for the species in question. Thus, the consequences on genetic diversity will depend on the rate of dispersal and gene flow between patches [35]. If gene flow is sufficiently high, then the population will behave as a single unit and the influences of the fragmentation will be negligible; however, if gene flow is restricted enough, genetic differentiation could occur, resulting in subpopulations [35]. This situation can lead to subpopulations comprising individuals that are more closely related than individuals between subpopulations. As discussed, the genetic variation between subpopulations is quantified as $F_{ST}$ [32]. However, because this variation is a consequence of restricted gene flow between populations, such that the effects of drift on the allele frequencies becomes independent within different subpopulations, $F_{ST}$ can be understood in terms of the counteraction between gene flow and genetic drift [69]. For example, if the probability of drawing two IBD alleles from a population is $1/2N$, where $N$ is the total population size, and we define $m$ as the probability that the allele has migrated between subpopulations, then it can be shown that [69]:

$$F_{ST} = \frac{1}{1 + 4Nm} \tag{1.4}$$

Eq(1.4) shows the inverse relationship between gene flow (the number of individuals migrating per generation: $Nm$) and the magnitude of population substructure $F_{ST}$. Once $F_{ST}$ has been estimated from allele frequency data (Eq(1.3)), then Eq(1.4) can be used to arrive at an estimate of gene flow ($Nm$).

Measuring gene flow ($F_{ST}$) offers an insight into dispersal. Gauffre et al. [110] used hierarchical $F_{ST}$ analysis to measure gene flow at multiple scales for the common vole ($Microtus arvalis$). The results were as expected at a fine, local scale, showing female kin clustering (female philopatry) and strongly male biased dispersal. However, at a
landscape scale, dispersal was balanced between genders. Thus, it is important to take scale into account when measuring gene flow.

1.4.1 Anthropogenic fragmentation

Hazel dormice may be particularly vulnerable to habitat fragmentation, as they are habitat specialists, have low population densities, low reproductive potential and limited dispersal \([114][23][115]\). As such, they are an ideal model species for investigating genetic diversity and structuring. The barriers to dormouse dispersal are not yet known. Previous research has found them to be reluctant to cross open ground \([7]\), other research has shown crossing open fields is a rare, but normal event when the distance between woodland patches was less than 500m \([116]\). Mortelliti et al \([117]\) reported that dormice are able to cross both grassland and mowed fields when homing behaviour is initiated, thus it would appear that when there are limited options at least some individuals will risk movement across habitats where they are vulnerable. Man-made structures such as roads are recognised as a threat to the natural movements of species increasing the likelihood of local extinction through reduced gene flow and increased inbreeding \([13]\). Roads are proposed to be a significant barrier to dormouse dispersal, especially when they are too wide to allow overhanging tree branches to aid movement from one side of the road to the other \([13]\); however dormice have been reported crossing from busy central reservations to the side of dual-carriageways in South West England, with one individual performing this movement multiple times \([118]\). It would appear that what constitutes a complete barrier to dormouse movement has yet to be established, although it is likely that the fragmentation of woodland habitats (e.g. with agricultural land, urban development, roads, railways etc.) would lead to a reduced number of individuals dispersing, if not complete restriction of movement.

Lower genetic diversity has been found in patchy hazel dormouse populations compared with those in a continuous habitat \([75]\), although genetic erosion was not significant and no bottlenecks were detected at either site, suggesting there was enough gene flow to maintain diversity. Hence, despite being habitat specialists, hazel dormice may be less susceptible to habitat fragmentation than previously thought. It is important to note, however, that dispersal can also be triggered by fragmentation, as the carrying capacity of the habitat may be reduced along with resources. Thus, it is important to validate these results. The same study \([75]\) found overall levels of gene flow to be limited to a fine geographic scale (1km), however rare interpatch movement was also detected (>1km), indicating higher mobility than previously thought was
possible. At a landscape scale, however, there was substantial genetic substructuring (>15km), indicating that the limit to dispersal is probably less than 15km [75][50]. Bani et al (2017) reported similar findings, with high genetic structuring in a fragmented landscape where woodland remnants were separated by a minimum of 1km. Mills et al. [50] found that dormice at the edge of their range (Cornwall) have reduced genetic diversity and low migration (i.e., potential gene flow) between subpopulations. All population genetic studies on the hazel dormouse have been carried out in regions at the edge of their range, in North Wales and South West England, where conditions (e.g. weather and resource availability) may be less than optimal. It is therefore, possible that local adaptations to naturally difficult conditions may influence the reaction of hazel dormouse populations to fragmentation.

1.5 Geographic scale of genetic diversity

Phylogeography is concerned with the distribution of genetic lineages in a spatiotemporal context [119][120]. Although historical connections of populations may be very different to current patterns of gene flow, both are relevant to the contemporary distribution of genetic variation [37]. European species’ current distributions would have been greatly influenced by ice age events [121] with many studies concentrating on recolonization routes into northern Europe following the last glaciation [122][121][4]. By comparing evolutionary relationships of genetic lineages with their geographic locations, we may be able to understand which factors most influenced the current distribution of genetic variation [37]. Physically neighbouring populations are not necessarily the most related or indicative of past movement, therefore genetic analysis is employed to reveal historic patterns of dispersal. The detection of phylogeographic structuring is important in terms of identifying long-isolated populations that might have distinct gene pools and local adaptations [4]; identifying such evolutionarily significant units is going to be important when embarking upon the conservation of a species [123].

A number of phylogeographic studies of the hazel dormouse have been performed with both mitochondrial and nuclear markers. Naim et al. [75] provided the first, using D-loop and COI mtDNA markers and showed strong divergence between Britain and mainland Europe, together with three clear British lineages separating northern (Wales and Welsh English border populations), central (Worcestershire, Bedfordshire and Suffolk) and southern lineages (Somerset, Dorset, Devon and Cornwall).

Mouton et al. [4], provided the most extensive, where the hazel dormouse was sampled throughout its distributional range, although excluding Great Britain, using a
species specific mtDNA marker (Cytochrome B). Two highly divergent lineages, with an ancient separation estimated at 7.7 million years ago, were revealed. Lineage 1 includes populations of Western Europe (France Belgium and Switzerland) and Italy. Lineage 2 includes populations of Central-Northern Europe (Poland, Germany, Latvia and Lithuania), the Balkan Peninsula and Turkey. The high levels of genetic diversity found, indicated that both Italy and the Balkans were refugia for hazel dormice during the last ice age, but these populations could not have contributed to postglacial recolonisation of western and central northern Europe respectively, leaving certain colonisation routes unconfirmed. The results do show that hazel dormice consist of a minimum of five historically isolated mitochondrial lineages, which could be regarded as independent conservation units.

A further study by Glass et al. [124] (see appendix) found that samples taken from Sussex, Isle of Wight, Devon and Wales all shared the same common Cytochrome B haplotype (Hap15) sequenced by Mouton et al. [4] in Poland, indicating no regional differentiation. However a further study using mtDNA (D-loop and Cytochrome B) and nuclear markers found similar regional clustering as reported by Naim [75] and suggested the level was meaningful enough to warrant specialised conservation measures, though bootstrap support was weak [5]. The magnitude of differentiation between British populations thus, remains unclear and needs to be clarified if effective conservation strategies are to be employed.

1.6 Thesis rationale and aims

Although census numbers are informative in the obvious sense that a decline is a worrying trend and the first indication that a species is under threat [31]; numbers are not considered to be useful indicators of the underlying genetic diversity of a species, which is now known to be a better predictor of species recovery and sustainability [25]. There are many potential causes of a species’ decline and they are not exclusively correlated with environmental degradation. Even when this is the case, the scope of potential causal factors ranges from the macro-scale (e.g., habitat loss and predation) to the micro-scale (e.g., disease vectors), offering a considerable array of possibilities. Although it is important to identify these factors, it is also necessary to uncover whether the decline in numbers is impacting underlying genetic diversity. In short, if it is not, then numbers are not a good indicator of the species’ health and evolutionary potential, which, is often the major concern in conservation. If, on the other hand, genetic diversity is also showing signs of decline, the nature of this decline - inbreeding
depression or lack of gene flow - can offer a useful indication of otherwise elusive causes. For example, barriers to gene flow can inform conservationists of this possible contributory driver of population decline. The focus of the current study is to ask whether analysis of the genetic diversity at the level of 1) the individual; 2) the social group and 3) the regional population provides a more informative insight into the population dynamics of the British hazel dormouse than is currently understood.

Previous research using microsatellite markers has been performed on the British hazel dormouse [75] [53] [125] [50] providing some information on the levels of genetic diversity and gene flow at the sub-population and regional population scales. However, the subject has by no means been exhausted and to date no genetic studies have been performed to ascertain the levels of sociality in the dormouse. Although Naim et al [75],[53] explored mating strategies, providing many questions on potential cooperative behaviours, these questions were not explicitly tested. As sociality can greatly influence individual movement and effective population number, understanding patterns of genetic variation in the context of social behaviour (or lack there of) to comprehend the affects of external influences (e.g. habitat fragmentation) on the persistence of populations. A number of phylogeographic studies have been performed on the hazel dormouse using both mitochondrial and nuclear markers [75] [4] [126] [5]. However, the extent of differentiation between dormice in Britain and in mainland Europe and among British populations remains ambiguous. In order to properly assess the need for putting into practise any extra conservation measures, such as the assignment of ESUs the subject warrants further attention.

The aims of this thesis are to 1) establish whether the fermentation hypothesis is a plausible mechanism for kin discrimination, by exploring whether relatives share more similar anal bacterial profiles than non-relatives. 2) Calculate relatedness (r) to establish social structuring, dispersal bias and any inbreeding consequences (FIS). 3) Quantify the magnitude of within and between population differentiation and estimate the impact of different habitat conditions on gene flow (FST). 4) Reanalyse dormouse phylogeny to determine the level of British and European differentiation and assess whether the British population is in genetic drift-migration equilibrium.

1.6.1 Thesis plan

Chapter 1, Introduction - The hierarchical population genetics of the threatened hazel dormouse (Musc cardinus avellanarius): Does the hierarchical population structure of the Hazel dormouse inform conservation?
Chapter 2, General Methods - Site description, population sampling and molecular ecology tools and analysis.

Chapter 3, Scent-gland microbial diversity as a possible mechanism for inbreeding avoidance - Background. The naturally low densities, evidence of social behaviour and poor dispersal abilities of the hazel dormouse mean that the potential for individuals encountering relatives is high, increasing the likelihood of inbreeding. Inbreeding within an otherwise genetically diverse population can cause the genetic erosion of variation within a population through the combined influences of natural selection against recessive homozygotes and the increased influence of genetic drift. This, coupled with the effects of habitat fragmentation, could greatly increase the chance of local population extinction, hence early detection is important. Previous work on the British Hazel Dormouse has identified lower levels of genetic diversity within fragmented populations of the hazel dormouse although, counter-intuitively, inbreeding levels were not as high as expected. This finding prompts the question that, due to the life history traits of the hazel dormouse, inbreeding avoidance mechanisms may have evolved. Sex-biased dispersal and kin discrimination are two ways in which inbreeding can be avoided. The low dispersal ability and evidence of possible créching behaviour make kin discrimination a viable mechanism for inbreeding avoidance in the hazel dormouse. If such a mechanism is employed, the genetic erosion expected in highly fragmented populations may occur more slowly than expected.

Aims: To test the hypothesis that kin-recognition is a plausible mechanism for inbreeding avoidance, the similarity of bacterial profiles derived from the anal scent glands of hazel dormice will be compared between individuals of known relatedness. Specifically, if scent is influenced by odour produced by host-dwelling bacteria, which in turn are influenced by the host’s immune system, the bacterial genomes of related individuals should be more similar, genetically, than between non-relatives.

Chapter 4, Social structure and inbreeding in the hazel dormouse - Background. Following on from Chapter 3, where the mechanisms of breeding behaviour were explored within a captive population, this chapter broadens out to quantify the magnitude of inbreeding within naturally occurring subpopulations. Little is known about the social structure of the hazel dormouse, however some level of sociality is evident during the active season, begging the question as to what is driving this behaviour and whether it impacts a population’s genetic diversity. It has long been hypothesised by population geneticists that in the absence of dispersal, where social barriers restrict gene flow and increase isolation between groups, the population suffers an increase in its vulnerability
to inbreeding and, consequently, inbreeding depression. Recent research has indicated that this prediction may not in fact hold and sociality may operate to maintain genetic diversity, which would be consistent were the individuals discriminating relatives from non-relatives for mate selection.

Aim: In order to explore this reasoning, the inbreeding coefficient ($F_{IS}$) will be measured from microsatellite marker data. Genetic distances between individuals will be calculated from relatedness measures ($r$) and correlated with geographic distances, in order to determine the social structuring of each population. The hypothesis being that the degree of relatedness between individuals provides an expected level of inbreeding should they mate at random. If the hazel dormouse population is socially structured and not randomly mating, then inbreeding levels should be lower than expected, even in fragmented habitats.

Chapter 5, Population genetic structuring at multiple scales - Background. This chapter approaches the social structure of the Hazel Dormouse by viewing it from a higher hierarchical level, which is that of the dispersal between populations. Dispersal is an important determinant of the population dynamics of any species, where patterns of gene flow influence population genetic structure. Dispersal is also essential for maintaining the genetic viability of species. Dispersal can be measured at various spatial scales Local dispersal; for example, can be inhibited by barriers such as geographic distance (as individuals are more likely to randomly mate with individuals in close proximity), in addition to physical barriers, such as a reservoir or railway line. If these barriers are sufficient to block dispersal, and hence gene flow, local genetic structuring is expected, as the genetic diversity of the isolated breeding groups would be subjected to genetic drift. At the landscape scale, dispersal is considerably influenced by environmental characteristics. Habitat fragmentation can disrupt species’ natural movements. In animals with naturally low dispersal ability, like the hazel dormouse, it is likely that fragmentation further decreases dispersal and, as such, may prevent gene flow. Male dormice have been found to move further than females, thus indicating the possibility of sex-biased dispersal, thus fragmentation may disrupt male-biased dispersal, further increasing the risk of inbreeding. It is yet unknown what constitutes as a barrier to dispersal for the hazel dormouse, but previous research has found them to be reluctant to cross open ground. Therefore, both local and landscape levels of differentiation may be expected. The chosen sample sites include potential barriers to local and landscape level gene flow: local in the form of water reservoirs (Wakehurst place), railway lines and geographic distance (Isle of Wight). Landscape level in the
form of farmland (Isle of Wight and Mallydams), coastline (Mallydams) and the sea (Isle of Wight).

Aim. This chapter will use a measure of genetic variance ($F_{ST}$) to quantify local and sub-population genetic structuring. Assignment tests to measure gene flow and thus detect immigration in addition to autocorrelation analysis will be used to detect genetic sub-structuring at a higher spatial scale, highlighting potential barriers to gene flow. The hypothesis is that the restrictions caused by both local and landscape barriers, will cause genetic differentiation between sites. It is hypothesised that local sub-structuring will be higher between those sites with clear detectable physical barriers (i.e. those found at Wakehurst Place and the Isle of Wight) and landscape level differentiation will be evident for isolated sites (Isle of Wight), as opposed to connected sites (Wakehurst Place).

Chapter 6, The Phylogeographic scale of British hazel dormouse diversity: The story so far - Background. Phylogeography is concerned with the distribution of genetic lineages in a spatiotemporal context. Although historical connections of populations may be very different to current patterns of gene flow, both are relevant to the contemporary distribution of genetic variation. A number of studies to date have investigated the phylogeographic structure of the hazel dormouse, however the magnitude of differentiation between British and European and among British populations remains unclear. Some authors suggest the levels are enough to warrant specialised conservation measures, whereas others have found little evidence of differentiation.

Aim: The Cytochrome b and D-loop regions will be sequenced for novel samples and combined with those generated by previous research in order to construct phylogenetic trees to clarify the levels of differentiation in hazel dormouse populations in Great Britain. The hypothesis is that the levels of differentiation among British populations is minimal and as such, does not warrant extra conservation measures.

Chapter 7, General conclusion - What has been revealed about the hierarchical structure of the hazel dormouse UK populations by application of molecular ecological analysis and how does this inform future conservation?
Chapter 2

General Methods and Materials

2.1 Study sites

Sampling took place at nine sites (see Figure 2.1), however only five had sufficient samples (>25) to be considered for analysis using microsatellite markers (with the other four sites producing <5 samples):

(1) Mallydams (Fairlight, East Sussex; latitude 51.069224 N; longitude 0.63854700; OS Grid reference TQ857122. Mallydams is a 22 hectare, semi-natural, mixed deciduous woodland owned by the RSPCA. Of the 22 hectares, 7 are monitored for dormice with 100 boxes placed in a grid formation [7]. The matrix environment is grassland/agricultural land with small to medium woodland patches, with relatively low connectivity in the form of hedgerows. There are urban developments to the east and west of the site, with a road directly south of the site and the sea 1.75km to the south.

(2) Wakehurst Place (Haywards Heath, West Sussex; latitude 50.879667 N; longitude -0.087440800; OS Grid reference TQ340317). The wood at Wakehurst place is 49.58 hectares of semi-natural ancient woodland owned by the National Trust and leased to Kew Royal Botanical Gardens. Approximately twenty of the 49.58 hectares are monitored for dormice, with 270 boxes currently in place. There is a reservoir running through the site with a number of foot bridges that span the reservoir, as well as a tree-lined road to the South of the site. The footbridges are unlikely roots of passage for the dormouse. The tree-lined road may function as a dispersal aid. The site is mainly surrounded by grassland and remnant patches of woodland, with varying connections by hedgerows. The B2028 road is to the east of the site.
(3) Binsted Wood (Arundel, West Sussex; latitude 50.850652 N; longitude -0.58929360; OS Grid reference SU994066) is a planted ancient woodland (PAWS). Binsted is a 427 hectare continuous woodland, with 170 boxes in one and a half hectares at the time of sample collection. The A27 motorway is situated to the north of the wood and has minor roads to the east and west. The majority of the surrounding area is agricultural/grassland fields, with some urban development and small isolated patches of woodland to the south and large patches of woodland north of the A27.

(4) Tilgate Park (Crawley, West Sussex; latitude 51.092968; longitude -0.18212758; OS Grid reference TQ274342.) is a 67.43 hectare mixed conifer and deciduous woodland owned by Crawley Borough Council. Tilgate Park is situated on the edge of a town, Crawley, West Sussex and contains 67.43 ha of woodland, with the M23 motorway running directly through Tilgate Forest; however, samples were only collected from the north of the M23, where 50 boxes had been placed. A major road (A23) is situated to the west and a double tracked railway to the east of the forest. Urban development surrounds the north side of the park and agricultural land is south of the M23.

(5) Briddlesford Copse, (Ryde, Isle of Wight; latitude 50.710644 N; longitude -1.2195111; OS Grid reference SZ552904) is a 76 hectare site. Briddlesford is situated on the Isle of Wight, Hampshire and as such, is separated from mainland Britain by the 6km Solent Straight. The site is consists of ten designated woodlands and seven woodland patches ranging from 2.55 to 43.66 ha. Here, a 'designated woodland' simply refers to human labelled, whereas a 'woodland patch' refers to a connected area of woodland. For example, Briddlesford Copse is a single woodland patch that has been given the three woodland names: Briddlesford North, Mid and South and thus, is three designated woodlands. Samples were taken from seven of the designated woods (five woodland patches). A railway runs between two woodland patches: Big Wood and Briddlesford Copse. A brook separates Briddlesford copse from a large continuous woodland to the north east of the site. The majority of the surrounding matrix is grassland and agricultural land with small patches of woodland, with varying degrees of connectivity.

For the purpose of investigating whether related individuals had more similar bacterial profiles than unrelated individuals samples were acquired from captive bred dormice of known pedigree. Only with captive bred dormice could the relationships between individuals be garenteed, as dormice display multiple paternity and as such, all fathers are rarely sampled, and creching behaviour, meaning any adult found with offspring may not necessarily be a parent and any offspring found together could be full or half sibs or more distant relatives (e.g. cousins). Having samples from individuals of
known pedigree also allowed the reliability of the microsatellite markers employed to be checked (see section 2.5). Samples taken from captive bred dormice were collected from the Wildwood trust in Herne, (Kent), which is part of the captive breeding program for reintroductions.

Fig. 2.1 Map of South East England showing all sample site locations. Woodland is depicted in green and urban development in grey. IoW: Isle of Wight (Bridlesford); BW: Binsted Wood; TP: Tilgate Park; W: Worth; WH: Wakehurst Place; BH: Bexhill; MD: Mallydams Wood; FR: Flatropers; HB: Herne Bay

2.2 Sample Collection

Samples were collected from populations within Sussex, as at the time of the study’s commencement, only populations at the edge of dormouse range in Great Britain had been sampled. The Isle of Wight site was selected as a comparison population, due to its long separation from the mainland and, as such, could be considered completely
isolated from all other sites. Sites were chosen in Sussex based on accessibility and their varying site structures and surrounding landscapes.

All sites are part of the long term national dormouse monitoring scheme run by the Peoples Trust for Endangered Animals (PTES) and as such, samples were taken during the routine box checks between April and October 2011-2015. Each individual was weighed, sexed and had their breeding status recorded. Sexing is performed by observing the distance between the anus and the urinary opening (shorter distance for females than males), as the testes of males are only apparent when they are ready to breed. Breeding condition is recorded by noting the presence or absence of visible testes for males; whether the abdomen is distended in females, indicating pregnancy or the presence of nipples, signifying recent lactation (NB - non-breeding; TS - testes scrotal; P - pregnant; NH - nipples halo; L - lactating).

More than 10 hairs are required for reliable microsatellite typing [127], thus bundles of approximately 20 -100 hairs were taken using tweezers (see Figure 2.2) and all tissue samples were provided from animals which died of natural causes, collecting a tail tip or an ear for DNA extraction. All samples were collected and stored under a Natural England licence (Licence numbers 20113857, 20122082, 20140438 and 20159472).

2.3 DNA extraction

2.3.1 Tissue extraction protocol

DNA was extracted from tissue (ear or tail) samples using the DNeasy Tissue Kit (QIAGEN) following the manufacturer’s instructions. A tissue sample of approximately 25mg was cut into small pieces, using a sterile scalpel blade and transferred to a sterile 1.5μl microcentrifuge tube. 180μl of Buffer ATL and 20μl of proteinase K was added to the tube, vortexed thoroughly and placed in a water bath at 56°C for between 1-2 hours (or until dissolved), vortexing occasionally. Once the tissue had dissolved, the sample was vortexed for 15 seconds, after which 200μl of Buffer AL was added to the sample and vortexed thoroughly, then 200μl of 100% ethanol was added to the mixture and vortexed thoroughly again.

The whole mixture was then transferred into a DNeasy Mini spin column placed in a 2 ml collection tube. The extraction was then centrifuged for 1 minute at 8000 rmp, after which the flow-through and collection tube were disposed of. The DNeasy spin column was placed in a clean collection tube, then 500μl of Buffer AW1 was added and centrifuged for 1 minute at 8000 rpm. Again the flow-through and collection tube were
Fig. 2.2 Visual depiction of dormouse hair plucking technique.
disposed of and the DNeasy spin column was placed in a clean collection tube. 500µl of Buffer AW2 was added and centrifuged for 3 minutes at 14,000 rpm, after which the flow-through and collection tube was disposed of and the DNeasy spin column was placed in a clean 1.5 ml microcentrifuge tube. 200µl of Buffer AE was added directly onto the DNeasy membrane, incubated at room temperature for 1 minute and then centrifuged for 1 min at 8000 rpm.

### 2.3.2 Hair extraction protocol

Hair sample extraction of DNA required an initial digestion step to maximise DNA yield, following the protocol by Pfieffer et al. [128], before continuing with the DNeasy Kit. A master mix containing 182.6µl of PCR grade water, 34µl of Tris-HCL (100mM), 34µl of sodium chloride (100mM), 1.02µl of calcium chloride (3mM), 68µl of 10% SDS, 13.6µl of 40mM DTT and 6.8µl of 250µg/ml Protease K per sample was added to a 2ml microcentrifuge tube and vortexed for 30 seconds. Each hair sample (>10 hairs with visible root) was placed in a 1.5ml microcentrifuge tube, to which 340µl of the extraction solution was added and vortexed for 5 seconds. The sample was then transferred to a water bath set at 56°C and left to incubate for 2 hours (or until the sample was fully dissolved), vortexing every half hour.

Once the hair was dissolved, the sample was vortexed for 15 seconds, after which 200µl of Buffer AL was added to the sample and vortexed thoroughly, then 200µl of 100% ethanol was added to the mixture and vortexed thoroughly again.

The whole mixture was then transferred into a DNeasy Mini spin column placed in a 2 ml collection tube. The extraction was then centrifuged for 1 minute at 8000 rpm, after which the flow-through and collection tube were disposed of. The DNeasy spin column was placed in a clean collection tube, then 500µl of Buffer AW1 was added and centrifuged for 1 minute at 8000 rpm. Again the flow-through and collection tube were disposed of and the DNeasy spin column was laced in a clean collection tube. 500µl of Buffer AW2 was added and centrifuged for 3 minutes at 14 000 rpm, after which the flow-through and collection tube was disposed of and the DNeasy spin column was placed in a clean 1.5 ml microcentrifuge tube. 200µl of Buffer AE was added directly onto the DNeasy membrane, incubated at room temperature for 1 minute and then centrifuged for 1 min at 8000 rpm.
2.4 PCR Microsatellite amplification

Twenty two microsatellites were available for the hazel dormouse (*Muscardinus avellanarius*) [34], three of which did not amplify in the target populations of this study. Genotyping was performed in 2\(\mu\)l PCR reactions, containing <10ng of lyophilised genomic DNA, 0.2\(\mu\)M of each primer and 1\(\mu\)l QIAGEN multiplex PCR mix (QIAGEN). PCR amplification was performed using a DNA Engine Tetrad PTC-225 thermal cycler (MJ Research, Bio-Rad, Hemel Hempstead, Herts., UK) with the following touch-down program: 95°C for 15 minutes; followed by 13 cycles of 95°C for 30 seconds, primer annealing for 30 seconds (decreasing by 1°C every cycle from 67°C to 55°C) and 72°C for 45 seconds; then 25 cycles of 95°C for 30 seconds, 55°C for 30 seconds and, 72°C for 45 seconds, and a final elongation at 60°C for 10 minutes. Amplified products were loaded on an ABI 3730 48-well capillary DNA Analyser (Applied Biosystems, California, USA) and GENEMAPPER v3.7 (Applied Biosystems, California, USA) was used to assign allele sizes.

2.5 Microsatellite error checking

All of the 19 microsatellites amplified were checked by comparing known parent offspring genotypes, two were found to be unreliable, as known parent and offspring genotypes consistently did not match (A5, Mav021) which left 17 microsatellites (see Table 2.1). In order to calculate error rate for the 17 reliable microsatellites, 10% of samples were randomly chosen to PCR again and then compared to the original genotype. Error rate for locus, genotype and multi genotype calculated manually as the proportion of mismatches (see Table 2.2); only four microsatellites produced any error, Mav30, 32, 53 and SH3.

Microsatellite amplification success averaged across all 17 loci was 99.3%. All 17 loci tested were found to be polymorphic in all populations, accept Mav034, which was found to be monomorphic in the Mallydams population. A total of 155 alleles were identified for 17 loci in 288 individuals, sampled across years in five populations (Table 5.2). Observed and expected heterozygosities, and estimated null allele frequencies were calculated using CERVUS v3.0.2 [129]. Tests for departures from Hardy-Weinberg equilibrium and assessment of linkage disequilibrium were conducted in GENEPOP v4.0.10 [130][131].
<table>
<thead>
<tr>
<th>Locus</th>
<th>EMBL accession</th>
<th>Repeat motif</th>
<th>Primer sequence (5' - 3') and fluor-label</th>
<th>Tm °C</th>
<th>Observed allele size range (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mav003</td>
<td>HE819185</td>
<td>(16)</td>
<td>F:[HEX]TTCTCAATTGCCTTCAGCTC R:TTAGTGAGGCCTTCTGCAAC</td>
<td>58.2</td>
<td>214-240</td>
</tr>
<tr>
<td>Mav011</td>
<td>HE819192</td>
<td>(2) (2) (2)</td>
<td>F:CCCGAGTACTGGGATTACAGG R:AAGGCTGAGGGTTTAGTTCAGAG</td>
<td>60.7</td>
<td>168-214</td>
</tr>
<tr>
<td>Mav015</td>
<td>HE819196</td>
<td>(5) (3) (2)</td>
<td>F:ACACCAGCCTCTGCAACTTAG R:GAGAATGGCCTCTGACGAAG</td>
<td>59.6</td>
<td>240-254</td>
</tr>
<tr>
<td>Mav017</td>
<td>HE819198</td>
<td>(2) (1) (2)</td>
<td>F:ACCATAAGCGAAGGGTGAG R:TCTCCGTCTTGCTTTACAGG</td>
<td>57.8</td>
<td>237-263</td>
</tr>
<tr>
<td>Mav030</td>
<td>HE819211</td>
<td>(16)</td>
<td>F:GGGAACCATCCAACCTATTG R:AAATGTCTCTGGGTTCCATACC</td>
<td>59.1</td>
<td>127-179</td>
</tr>
<tr>
<td>Mav032</td>
<td>HE819213</td>
<td>(2) (1) (1)</td>
<td>F:AGGGTTTCCCACAGAAACAG R:GAATCTATAACCAGTTAGCAAATCTCC</td>
<td>59.0</td>
<td>135-155</td>
</tr>
<tr>
<td>Mav034</td>
<td>HE819215</td>
<td>(15)</td>
<td>F:ATGTACACAACGCGGAAGTG R:GGCAGCTCAGAAGTAGAATGC</td>
<td>59.6</td>
<td>235-253</td>
</tr>
<tr>
<td>Mav038</td>
<td>HE819219</td>
<td>(18) (2)</td>
<td>F:TCTTGTCTCAGCTTCCCAAGT R:AGCACCTCAGAGGGAGTTGT</td>
<td>59.1</td>
<td>252-277</td>
</tr>
<tr>
<td>Mav044</td>
<td>HE819225</td>
<td>(6) (6) (15)</td>
<td>F:CAGCACCCACAGCCTCAT R:GGCAGGCTACTGCTGCAC</td>
<td>60.9</td>
<td>123-162</td>
</tr>
<tr>
<td>Mav048</td>
<td>HE819229</td>
<td>(14) (1) (1)</td>
<td>F:CACAATCCTCCTGTGTCAGC R:AGGACTCTCCATGCCAAGAG</td>
<td>59.3</td>
<td>207-254</td>
</tr>
<tr>
<td>Mav049</td>
<td>HE819230</td>
<td>(11) (1) (1)</td>
<td>F:GGGTTTGCAAAGAACCCAAT R:CCTGGGTTCCGTCCCTAGTA</td>
<td>61.1</td>
<td>202-225</td>
</tr>
<tr>
<td>Mav051</td>
<td>HE819232</td>
<td>(18)</td>
<td>F:AGTGGTGGCATGTACCTGTG R:ACGATTATTCTGCCACTGAGC</td>
<td>59.5</td>
<td>229-249</td>
</tr>
<tr>
<td>Mav053</td>
<td>HE819234</td>
<td>(3) (1) (3)</td>
<td>F:GGCCTTCTATAACTTAGCAAGAACC R:ACGGACGAGTGAGCTGTACC</td>
<td>60.1</td>
<td>217-263</td>
</tr>
<tr>
<td>MavSF10*</td>
<td>GF089509</td>
<td>(43)</td>
<td>F:GCTGAGGGTATAACTTGGAGGTAG R:GTCTGAAATGGCTGGTATTGC</td>
<td>59.6</td>
<td>218-243</td>
</tr>
<tr>
<td>MavSG3*</td>
<td>GF089517</td>
<td>(54)</td>
<td>F:TGTTGACTGATTGAGTGGTGAC R:GGCTGAAGATGTAGCTCATAGG</td>
<td>58.6</td>
<td>123-159</td>
</tr>
<tr>
<td>MavSG6*</td>
<td>GF089511</td>
<td>(38)</td>
<td>F:AGCCCTTCACTTTCCCTGTATC R:TCCTCAGCAACTTAGCAAGACC</td>
<td>60.8</td>
<td>165-205</td>
</tr>
<tr>
<td>MavSH3*</td>
<td>GF089518</td>
<td>(29) (1) (11)</td>
<td>F:CTGGGAGTGTAGCTTGAAGG R:CAAGGATAGGGATACACCTCAG</td>
<td>57.6</td>
<td>201-215</td>
</tr>
</tbody>
</table>

**Table 2.1** List of microsatellite primers used, including the forward and reverse primer sequences, the annealing temperatures (Tm) and the allele size.
<table>
<thead>
<tr>
<th>Allelic</th>
<th>0.014286</th>
<th>0.027778</th>
<th>0.008929</th>
<th>0.046875</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus</td>
<td>0.028571</td>
<td>0.055556</td>
<td>0.017857</td>
<td>0.09375</td>
</tr>
<tr>
<td>Genotype</td>
<td>0.289474</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2 Manually calculated error rates for alleles, loci and genotypes.

2.6 Employing the Hardy-Weinberg Principle

The chapters of this thesis consider the analyses based upon Wright’s concepts of inbreeding [32], such as substructuring \( F_{ST} \) and close-kin inbreeding \( F_{IS} \), which are based upon the basic principles and underlying logic presented here. It is possible to identify the dynamics of a population using information gleaned from genetic diversity. The relationship between allele frequency and genotype frequency is influenced by mutation, population size, mating strategy, natural selection and gene flow. In the absence of these factors, allele frequencies and genotype frequencies conform to the simple relationship known as the Hardy-Weinberg (HW) relationship [69].

The Hardy-Weinberg relationship states that in non-evolving populations the allele frequencies and hence genotype frequencies will remain constant from generation to generation. These populations are then said to be in HW equilibrium (HWE). For example, at a single locus with two alleles, \( A \) and \( a \), at frequencies of \( p \) and \( q \)
Table 2.3 Characteristics of 17 microsatellite loci used in study number of individuals typed per locus ($n$), number of alleles at each locus ($k$), observed heterozygosity ($H_O$), expected heterozygosity ($H_E$) and the estimated null allele frequency, computed across all populations and years.

<table>
<thead>
<tr>
<th>Locus</th>
<th>n</th>
<th>k</th>
<th>$H_O$</th>
<th>$H_E$</th>
<th>$F_{Null}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mav015</td>
<td>288</td>
<td>5</td>
<td>0.33</td>
<td>0.39</td>
<td>0.062</td>
</tr>
<tr>
<td>Mav032</td>
<td>282</td>
<td>6</td>
<td>0.63</td>
<td>0.70</td>
<td>0.050</td>
</tr>
<tr>
<td>Mav49</td>
<td>288</td>
<td>11</td>
<td>0.63</td>
<td>0.73</td>
<td>0.080</td>
</tr>
<tr>
<td>Mav051</td>
<td>288</td>
<td>10</td>
<td>0.60</td>
<td>0.77</td>
<td>0.146</td>
</tr>
<tr>
<td>Mav017</td>
<td>284</td>
<td>7</td>
<td>0.52</td>
<td>0.68</td>
<td>0.151</td>
</tr>
<tr>
<td>Mav053</td>
<td>284</td>
<td>6</td>
<td>0.37</td>
<td>0.57</td>
<td>0.213</td>
</tr>
<tr>
<td>MavSG3</td>
<td>288</td>
<td>10</td>
<td>0.79</td>
<td>0.86</td>
<td>0.042</td>
</tr>
<tr>
<td>Mav011</td>
<td>287</td>
<td>10</td>
<td>0.77</td>
<td>0.85</td>
<td>0.047</td>
</tr>
<tr>
<td>Mav030</td>
<td>285</td>
<td>12</td>
<td>0.75</td>
<td>0.82</td>
<td>0.039</td>
</tr>
<tr>
<td>Mav034</td>
<td>288</td>
<td>6</td>
<td>0.18</td>
<td>0.49</td>
<td>0.477</td>
</tr>
<tr>
<td>Mav003</td>
<td>275</td>
<td>7</td>
<td>0.65</td>
<td>0.70</td>
<td>0.037</td>
</tr>
<tr>
<td>Mav038</td>
<td>289</td>
<td>9</td>
<td>0.51</td>
<td>0.72</td>
<td>0.189</td>
</tr>
<tr>
<td>Mav044</td>
<td>288</td>
<td>12</td>
<td>0.80</td>
<td>0.85</td>
<td>0.030</td>
</tr>
<tr>
<td>Mav48</td>
<td>288</td>
<td>12</td>
<td>0.74</td>
<td>0.80</td>
<td>0.040</td>
</tr>
<tr>
<td>SF10</td>
<td>286</td>
<td>9</td>
<td>0.75</td>
<td>0.83</td>
<td>0.050</td>
</tr>
<tr>
<td>SH3</td>
<td>283</td>
<td>8</td>
<td>0.63</td>
<td>0.79</td>
<td>0.114</td>
</tr>
<tr>
<td>SG6</td>
<td>289</td>
<td>15</td>
<td>0.51</td>
<td>0.78</td>
<td>0.206</td>
</tr>
<tr>
<td>All</td>
<td>288</td>
<td>155</td>
<td>0.60</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td>Site(s)</td>
<td>OS Map Ref.</td>
<td>Origin</td>
<td>N</td>
<td>Sample Type</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>3</td>
<td>Wildwood Trust, Kent</td>
<td>TR172637</td>
<td>Wild/Captive</td>
<td>51</td>
<td>Bacterial DNA</td>
</tr>
<tr>
<td></td>
<td>Mallydams Wood, East Sussex</td>
<td>TQ857122</td>
<td>Wild</td>
<td>74</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Briddlesford Copse, Isle of Wight</td>
<td>SZ552904</td>
<td>Wild</td>
<td>40</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td>4</td>
<td>Binsted Wood, West Sussex</td>
<td>SU994066</td>
<td>Wild</td>
<td>18</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Tilgate Park, West Sussex</td>
<td>TQ274342</td>
<td>Wild</td>
<td>18</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Wakehurst Place, West Sussex</td>
<td>TQ360137</td>
<td>Wild</td>
<td>30</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Mallydams Wood, East Sussex</td>
<td>TQ857122</td>
<td>Wild</td>
<td>127</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Briddlesford Copse, Isle of Wight</td>
<td>SZ552904</td>
<td>Wild</td>
<td>90</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td>5</td>
<td>Binsted Wood, West Sussex</td>
<td>SU994066</td>
<td>Wild</td>
<td>5</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Tilgate Park, West Sussex</td>
<td>TQ274342</td>
<td>Wild</td>
<td>5</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Wakehurst Place, West Sussex</td>
<td>TQ360137</td>
<td>Wild</td>
<td>5</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Mallydams Wood, East Sussex</td>
<td>TQ857122</td>
<td>Wild</td>
<td>11</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Briddlesford Copse, Isle of Wight</td>
<td>SZ552904</td>
<td>Wild</td>
<td>18</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Worth, West Sussex</td>
<td>TQ307356</td>
<td>Wild</td>
<td>2</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Bexhill, East Sussex</td>
<td>TQ756096</td>
<td>Wild</td>
<td>2</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Flatroper, East Sussex</td>
<td>TQ601230</td>
<td>Wild</td>
<td>1</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Wadhurst, East Sussex</td>
<td>TQ630313</td>
<td>Wild</td>
<td>1</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Hillsbrook, Devon</td>
<td>SX273753</td>
<td>Wild</td>
<td>2</td>
<td>Dormouse DNA</td>
</tr>
<tr>
<td></td>
<td>Carmathen, Wales</td>
<td>SU248636</td>
<td>Wild</td>
<td>1</td>
<td>Dormouse DNA</td>
</tr>
</tbody>
</table>

Table 2.4 Summary of the sampling sites, sample numbers and analysis methods used for each chapter
respectively, whatever the initial frequencies of these two alleles, after one generation of random mating the frequencies of the genotypes \(AA\), \(Aa\) and \(aa\) will be in the ratio \(p^2 : 2pq : q^2\). Therefore, at a single locus, any deviations from HWE will be erased after one generation of random mating [14]. Deviations from HWE are therefore indicative that at least one of the assumptions of the HWE are being violated. Unsurprisingly, many natural populations are not in HWE.

When a population deviates from HWE, there is a shift in genotype frequency. For example, when inbreeding occurs in a population, the population is no longer randomly mating and the frequency of homozygote genotypes increase proportionally to the decreasing frequency of heterozygote genotypes. This is because relatives, by definition, are more likely to have inherited the same ancestral alleles from a common ancestor, which are said to be identical by decent (IBD). Were these relatives to have offspring, the offspring could inherit two copies of these IBD alleles, forming a homozygote genotype. This is over and above the chance that this offspring is homozygous just by chance. For an inbred individual, therefore, the probability of homozygosity is greater than that expected simply by chance; i.e., the probability of them having an \(AA\) genotype > Hardy Weinberg expectations and therefore > \(p^2\). A population comprised of inbred individuals would therefore be expected to show excess homozygosity over HW expectations, which can be detrimental when recessive mutations are segregating in the population, resulting in inbreeding depression [56]. Alternatively, gene flow between two previously isolated populations brings together novel genotypes increasing the probability of offspring being heterozygous, a situation that can also be detrimental when isolated populations have evolved locally optimal adaptations, a result known as outbreeding depression [14]. Conservation biologists employ population genetics in this way to predict the scale of potential harm that can befall a population experiencing changes in patterns of gene flow.
Chapter 3

Bacterially Mediated Scent Production as a Possible Mechanism for Kin Discrimination

Abstract

The naturally low densities and poor dispersal abilities of the hazel dormouse means the potential for encountering relatives is high, increasing the likelihood of inbreeding. Previous work on the British hazel dormouse has identified lower levels of genetic diversity within fragmented populations, however counter-intuitively, inbreeding levels were not as high as expected. This finding prompts the assertion that inbreeding avoidance mechanisms may have evolved. Evidence of possible cring behaviour suggests that kin discrimination is a viable mechanism for inbreeding avoidance in the hazel dormouse. Individual recognition could be based on signature odours produced by bacterial communities unique to the individual, and so bacteria may be the underlying mechanism of kin discrimination. If scent is influenced by host-dwelling bacteria, which in turn are influenced by the host immune system and mother to neonatal transmission, bacterial profiles of related individuals should be more similar than between non-relatives. To test this hypothesis, anal swabs of captive dormice of known pedigree were taken. Bacterial profiles were generated using T-RFLP analysis of the 16S rRNA gene and then compared to pedigree relatedness. Related dormice were found to have significantly more similar bacterial profiles than unrelated individuals; and that this relationship was genetically driven, rather than through social contact in a shared environment. Bacterial community structure and diversity differed between age groups, signifying the acquisition of a mature bacterial community and implying an immune element to the attainment of a stable community. The results of the study
indicate that scent driven kin discrimination could be a viable inbreeding avoidance mechanism in the hazel dormouse, although extensive, further research is required to confirm the findings of this chapter.
3.1 Introduction

One of the key elements of kin selection theory is the preferential cooperation of relatives over non-relatives, which requires kin recognition abilities. The mechanisms behind recognition are unclear, one possibility is that relatives use microbial-mediated odour as a cue. Here we ask whether relatives share similar microbial community profiles compared with non-relatives by correlating bacterial 16s RNA similarities (TRF profiles) between individuals with relatedness.

3.1.1 Why would kin recognition evolve in the hazel dormouse?

Inbreeding

The hazel dormouse is an ecological specialist and as a result has evolved low-density populations [12]. Even in optimal habitats there are typically fewer than ten individuals per hectare [12]. This, together with poor dispersal abilities, predicts that the potential for encountering relatives is high, increasing the likelihood of inbreeding.

Inbreeding results from the mating of relatives, which can lead to a reduction in fitness of the inbred offspring, known as inbreeding depression [133]. Reduced fitness through inbreeding depression could occur as a consequence of an increase in homozygosity as well as a decrease in heterozygosity. Increased homozygosity has the potential to result in a greater expression of deleterious recessive traits. The subsequent decline in heterozygosity, on the other hand, could decrease fitness at loci where heterozygotes have the selective advantage over homozygotes (overdominance or heterozygote advantage) [56][133]. Thus, when the likelihood of inbreeding is naturally high, it is expected that some form of inbreeding avoidance mechanism would evolve to mitigate the effects of inbreeding depression [56]. The increased homozygosity resulting from inbreeding should expose deleterious, recessive alleles to natural selection, thus, over time, purging the population of its genetic load [133]. Under these circumstances, low levels of genetic variability would be expected at genes of large effect, as rare detrimental alleles become removed from the population, thus mating between non-relatives over relatives would no longer confer any benefit and further inbreeding would have no effect on individual or population fitness. This is a consequence of the fact that, as the genetic variation tends towards zero, homozygosity tends towards 100%, irrespective of how inbred the individual happens to be. Nevertheless, the hazel dormouse has been found typically to have higher levels of genetic diversity than would
bacterially mediated scent production as a possible mechanism for kin discrimination.

It has also been demonstrated that genetically identified fathers were less related to their mates than other males in the vicinity (those sharing a nest box with the female) [53], indicating inbreeding avoidance or mate selection.

Inbreeding avoidance mechanisms include sex-biased dispersal, polyandry and kin recognition [134]. Sex-biased dispersal, where one sex disperses, while the other stays or returns to its natal site to breed, is common in the animal kingdom [135]. In mammals, male-biased dispersal and female philopatry are generally observed [135] and are expected to reduce the possibility of inbreeding opportunities, as related males are removed from the population before breeding occurs [113]). Removing related males also allows the influx of unrelated males from other populations, increasing the chances of a philopatric female mating at random with unrelated males, thus reducing the risk of inbreeding. If both sexes were to disperse, the likelihood that related females and males arriving in the same area would be high, as their habitat requirements may be similar and suitable habitat may be sparse.

Polyandry has been proposed to reduce inbreeding in two ways, firstly through post-copulatory mechanisms, whereby females bias paternity towards males with higher fitness, or greater genetic compatibility, potentially skewing paternity towards unrelated males or those with higher genetic diversity [136][137][138][139]. Secondly, polyandrous females that have litters from multiple paternities would produce half-sib offspring, rather than full-sibs; thus if individuals from the same litter did mate, the degree of inbreeding in the next generation would be lower [137]. There is some evidence for the former, although many studies have shown that paternity is skewed towards the first mating [138], however, the evidence for the latter is currently inconclusive [137].

Both polyandry and male-biased dispersal have been observed in the hazel dormouse [53][125][34]. The influx of unrelated males through dispersal and the decrease in the degree of relatedness through polyandry may explain the higher than expected levels of genetic diversity observed in the hazel dormouse. However, male-biased dispersal is not exclusively practiced; site fidelity has been reported in male dormice, where the same marked male has been found in the same nest box for several consecutive years [12]. Additional complications arise as a consequence of females who are born early in the season having been observed producing litters that same year [10], making father-daughter mating a possibility. Although multiple paternity was found in one study [75], a different study found the amount of multiple mating to be considerably less [50], thus there appears to be some site dependent factors at play. It is currently unclear
what role multiple mating plays in female dormice, particularly with regard to whether it maintains genetic diversity or reduces inbreeding.

**Cooperation**

When contact between relatives is likely and mate choice is possible, individuals are expected to choose to mate with non-kin over kin [140]. For such a choice to be possible, individuals would need to be able to recognise related individuals. Overlapping generations are required for social behaviour to evolve; however, in populations with overlapping generations, the evolution of kin recognition is also likely to prevent inbreeding [141]. Additional evidence of social behaviour amongst the hazel dormice comes from the observation of same sex cohabitation of adults (both male and female) outside the breeding season [13][10], the occurrence of multiple dormice in nest boxes [12] and evidence of créching behaviour, where large litters consisting of two different ages (juveniles and brown young) were found within the same nest box [75]. Breeding females have also been observed in close proximity to female half-sibs within their nest [75], showing a form of cooperation and raising the possibility that dormice have the ability to recognise kin.

Kin selection theory predicts that animals can increase their inclusive fitness by allocating more cooperation to kin over non-kin [78]. The most familiar example of this is parental care. The evolution of parental care in species that invest heavily in reproduction (such as mammals) is unsurprising when taking into account genetical theory, which predicts that natural selection would favour individuals who maximise their genetic contribution to the next generation.

For diploid organisms, the probability that an offspring inherits any one, specific, parental allele is 0.5, so \( r = 0.5 \), where \( r \) is the coefficient of relatedness: a parent and their offspring share one half of their alleles in common. Caring for ones’ offspring maximises the chances that they will reach maturity, reproduce and pass on their genes. Thus, parental care could be considered genotypically selfish [142].

All relatives, on average, have a higher probability of sharing alleles identical by decent than non-relatives. For instance, for full siblings have an \( r = 0.5 \), for half siblings and grandchildren \( r = 0.25 \), for nephews, nieces and cousins \( r = 0.125 \) and so on. From this we can see that full siblings share, on average, the same proportion of alleles in common as parents and offspring. Therefore, caring for offspring confers, on average, the same fitness benefit as aiding full siblings, as the probability of alleles identical by decent passing to the next generation (i.e. grandchildren, nieces and nephews) is the same, \( r = 0.25 \). Accordingly, aiding a half-sib would confer a smaller
inclusive fitness increase than aiding a full sibling or one’s offspring, as passage of alleles of common decent to the next generation would, in this case, be \( r = 0.125 \). This can be made explicit by Hamilton’s rule. Hamilton’s rule summarizes how relatedness between co-operators biases the trade-off between benefits and costs incurred from the act of cooperation:

\[
Br > C
\]

Here \( B \) represents the benefit conferred from the actor to a recipient of the act, \( C \) represents the cost of the action to the actor and \( r \) is the coefficient of relatedness. From this inequality it can be seen that for the cooperative behaviour to have evolved, the benefit obtained by the actor must outweigh the cost and that the cost would increase with decreasing relatedness [142].

Consequently, kin selection theory predicts the evolution of some form of perceivable recognition component, such as odour, to evolve with sociality/cooperation, in order for the degree of relatedness to be determined and the cost of helping behaviour to be calculated [78][141].

Kin recognition is a means by which individuals are able to assess the genetic relatedness of others and hence the likelihood that a potential mate shares alleles that are IBD to their own. This is typically inferred from kin discrimination, which is the differential response of individuals towards their kin [143][144][145]. Kin recognition is expected to evolve when animals interact repeatedly over time and when discrimination among multiple familiar individuals is beneficial, such as with reciprocal altruism, parental care, nepotism and dominance hierarchies [78][146][145]. Inbreeding avoidance has also been postulated as an additional driver of the evolution of kin recognition and may therefore be expected in species that are philopatric and likely to encounter relatives. If these relationships persist, selection would favour the evolution of discrimination, even among distant kin [78].

### 3.2 Kin recognition mechanisms

There are three main mechanisms used by organisms for the purpose of kin recognition, which are spatial association, prior-association and phenotype matching [147]. Spatial association is an indirect form, which occurs when relatedness reliably correlates with distance and neighbours are treated preferentially, regardless of their true relatedness and as such, could be considered as a failure to recognise kin [147]. Prior association and phenotype matching are both direct forms of recognition and require the production
of individual cues or labels, which are perceivable by other individuals and can be used to compare these cues to a template in order to act accordingly [148][143]. Recognition through prior-association occurs when individuals learn "labels of related individuals" during early development, such as when siblings share a nest, and then later discriminate between familiar and non-familiar individuals on the basis of these differences [141].

Prior association may be sufficient for inbreeding avoidance if the coming into contact with unfamiliar kin is unlikely. Behavioural experiments have shown that reproductively activated naked mole-rats (*Heterocephalus glaber*) discriminate on the basis of familiarity: unfamiliar males over familiar males [149]. The naked mole-rat lives in discreet burrows, where contact with unfamiliar relatives is low [149]. Prior association is also sufficient when relatives interact in the absence of non-kin, such as parent-offspring recognition and nest mates [141][145][150]. However, prior association can lead to recognition mistakes, if non-kin are encountered during the learning phase, as demonstrated in spiny mice (*Acomys cahirinus*) that were cross-fostered at birth: when siblings were reunited, they would not huddle together, but they would huddle with non-sib litter-mates [151].

Phenotype matching occurs when individuals learn their own phenotype label and/or those of familiar kin and later compare and match the phenotype of unknown individuals to this learned template [150]. Phenotype matching allows for the recognition of unknown kin resulting in a more refined discrimination between kin class, which is important when considering kin-directed behaviours, such as nepotism and inbreeding avoidance [150]. For phenotype matching to be reliable, the production cues would have to vary predictably with kinship, thus, a genetic component would be expected as kinship denotes the sharing of alleles between individuals of common descent [143]. Hamilton [78] stated that genes should affect some perceivable feature of an organism, the perception of that feature and the social response resulting from what was perceived.

Phenotype matching is expected to be observed in circumstances when unfamiliar kin are likely to come into contact and when non-kin may be present during the important learning phase, leading to mismatches in preferential treatment. Male dormice have been found nesting with females and litters they have not sired [53] indicating that unrelated individuals may come into contact with juvenile dormice within the nest. Phenotype matching is expected to be favoured in situations when discrimination without prior association is necessary, such as: polygyny (to allow for the recognition of paternal half-sibs), communal nesting, natal or breeding dispersal (dispersers may come into contact with older sibs or parents) and when there are overlapping generations [141]. Self-referent phenotype matching would be favoured
when females are polyandrous and young may need to discriminate between maternal full and half sibs [87]. Being able to discriminate between full and half sibs may be important when deciding whether to engage in cooperative behaviours that impact upon individual fitness (such as, resource sharing or warning signals), when the cost may outweigh the benefits if applied to lesser relatives (i.e., Hamilton’s rule).

The hazel dormouse shows polygyny, polyandry, possible communal nesting and overlapping generations; therefore it is predicted that phenotype matching is used for kin recognition, thus it can be hypothesised that there should be a genetic element to the production of recognition cues and this should in turn vary with relatedness.

It is important to point out however, that multiple recognition mechanisms can be employed by a species and can vary throughout an individual’s lifetime. Spatial recognition, for example, may be sufficient for offspring - parent recognition early on, when contact with other individuals is unlikely, but it may be necessary to employ prior association or phenotype matching once out of the nest.

In some cases it may only be necessary to distinguish between certain classes of individuals, such as close-kin, relative to everything else, when avoiding inbreeding or cooperative behaviour that’s unlikely to be reciprocated or beneficial to inclusive fitness. In these circumstances, labels shared by many individuals would be sufficient [152]. Individual recognition labels would only be required when more complex interactions occur, where the ability to distinguish between specific animals is necessary [152]. Such interactions include long-term partnerships, dominance hierarchies, where it is essential to recognise individuals with a higher rank, so as to avoid social conflict within the group and competitive relationships, where the identification of an individual known to be the best fighter would mean the avoidance of injury [152]. Long-term pair bonding has been reported in the hazel dormouse [13] and, as adult dormice hibernate alone during the winter, they must seek out their long-term mate each breeding season. This implies complex recognition systems, in which temporally stable individual recognition cues are required. For reliable recognition, the individual cue needs to be relatively independent from background influence, be easily disentangled from other information (i.e. breeding condition) and should exhibit a high degree of diversity [152]. However, as hazel dormouse population densities are low, it is equally likely that the same individuals happen upon each other by chance or due to similar preferences of habitat.

Given what is known about the hazel dormouse, it seems reasonable to predict that the hazel dormouse will use phenotype matching in order to distinguish kin classes, at least to some extent. If this were the case, a genetically coded perceivable cue would be expected and odour is a likely candidate for such a recognition label.
3.2 Kin recognition mechanisms

3.2.1 Olfactory communication

Olfactory communication is common in mammals, where urine, faeces and scent gland secretions are all used for the purpose of conveying information through scent marking [94]. Odour is thought to be the most common mechanism for conspecific recognition in animals [153]. Given this, olfactory cues would be advantageous when auditory and visual signals are difficult to discern, for instance at night or near a loud sound source. In addition to this, the constant emission of a visual or auditory cue would use up a lot of energy, which could lead to a fitness cost. Scent marks, however, can remain in the environment for some time and therefore can be easily distributed in space and time, without the immediate risk to safety or the requirement of close proximity to other individuals [154].

Olfactory cues would be a beneficial form of communication in the hazel dormouse, due to their nocturnal nature and because nocturnal predators, such as owls, have evolved a highly developed sense of directional hearing [155]. Predictably, then, the hazel dormouse is more vulnerable to predators with highly developed olfactory capabilities, such as the weasel. Odour cues can be derived from the animal’s diet, the environment or synthesised by the animal themselves [89].

The information conveyed to other individuals through odour can include health, dominance status, genotype (e.g., MHC alleles) and social group membership; in fact there is a multitude of possible information that can be communicated via odour [92][93][89][94]. Therefore, olfactory cues play a prominent role in shaping social and antagonistic behaviours, such as mate choice, parental care, territory maintenance, dominance hierarchy and group cohesion [89]. Kin recognition through scent has been described in a number of mammals, including Belding’s ground squirrels (Urocitellus beldingi)[141], Golden hamsters (Mesocricetus auratus) [86], wild house mice (Mus musculus)[156], bank voles (Myodes glareolus) [157] and beavers (Castor canadensis)[158].

3.2.2 Odour as a kin recognition cue

For odour to function as an effective cue for kin recognition, their would need to be low within-individual variability, high between-individual variability and stability of both over time [159]. Considerable work has been conducted on the role played by armpit odour in human social recognition cues [89]. Axillary odours, such as armpit odours, are individually distinct and appear to remain relatively stable over time [160][161][162] and there is strong evidence that these odours reflect a individual’s genotype and/ or their
kin relationships. For example, Weisfield et al. [88] demonstrated that mothers were able to distinguish the axillary odours of their biological children, but not those of their step-children and children could identify their full siblings, but not half siblings or step-siblings, to whom they cohabited with; indicating the likely role of a genetic element, as continual association and similar environmental factors failed to correlate. Further evidence of a genetic factor responsible for axillary scent is provided by monozygotic twin studies. It was found that untrained dogs are unable to distinguish between the odours of monozygotic twins [161] and human subjects were able to match the axillary odours of monozygotic twins, even when the twins did not cohabit [163].

Further evidence comes from the North American beaver (Castos canadenis), which has the ability to recognise unfamiliar kin [158]. Sun and Müller-Schwarze [159] examined beaver anal gland secretion profiles and found that variation within the same individual was not significant and that there was a positive correlation between degree of relatedness and scent gland secretions, providing evidence that scent gland secretions could be a recognition cue at least in some mammals.

3.2.3 The role of bacteria in scent communication

Bacteria are proposed to play a prominent role in olfactory communication in mammals, particularly those that use scent gland secretions [164]. The fermentation hypothesis provides a model of how bacteria contribute to chemical recognition in mammals [89] and is comprised of two components. Firstly, it is proposed that symbiotic bacteria living in the scent glands produce volatile odourants that contribute to mammalian recognition cues and, secondly, that inter-individual variation in these cues is caused by underlying variation in the composition and structure of bacterial communities inhabiting these glands [95][97]. It is important to note that, although the fermentation hypothesis refers to scent gland microbia, it could also be relevant for urine and faecal scent marks [89].

Most mammals have scent glands, which provide the perfect living conditions for symbiotic bacteria, as they are warm, moist and produce ample fatty substances and organic materials for bacteria to feed upon [95][96]. The breakdown of these substances release volatile odourants that contribute to the animal’s individual scent. Known microbiological breakdown products from carbohydrates, proteins and cholesterol, such as short-chain fatty acids and phenylacetic acid, have been found in the red fox (Vulpes vulpes), the gray wolf (Canis lupus) and the lion (Panthero leo) anal scent secretions, suggesting volatiles of the scent gland are microbial in origin [95]. Bacterial surveys of 14 mammalian species’ scent glands found them to harbour diverse bacterial communities,
3.2 Kin recognition mechanisms

including known odour producers, such as fermentative anaerobes (reviewed in Archie & Theis [89]).

When bacterial species were experimentally removed from the scent glands of the red fox (*Vulpes vulpes*) and the Indian mongoose (*Herpestes edwardsii*) using antibiotics, it was found that short-chain fatty acids implicated in scent production were no longer produced [165][166]. The same studies also showed that when bacteria from the scent glands were cultivated, they produced the same short-chain fatty acids produced in scent marks. Until recently, there has been difficulty in obtaining evidence for the fermentation hypothesis, due to the limitations of culture-based techniques - not all suspected bacterial species culture well - preventing any comprehensive information on bacteria within scent glands being obtained [89][167]. The advancement in molecular techniques using universal primers of the 16s rRNA gene, however, have allowed a return to this topic. Recent studies using such molecular techniques have found extensive bacterial communities in badgers (*Meles meles*)[94], spotted and striped hyenas (*Crocuta crocuta*; *Hyaena hyaena*[98] and meercats (*Suricata suricatta*[168].

Both spotted (*Crocuta crocuta*) and striped (*Hyaena hyaena*) hyenas have been found to harbour obligate or facultative anaerobes in their scent glands, which are known odour producers[98]. Both species-specific and social group specific bacterial communities have been demonstrated in the hyena [98]. It has also been observed that volatile fatty acid profiles (which contribute to odour) covaried with the bacterial profiles of the scent gland, indicating that it is indeed the make up of the bacterial community that contributes to the group smell [98] and thus are important in scent communication. Social group specific bacterial communities have also been demonstrated in the scent glands of meerkats [168]. The studies described clearly demonstrate that bacteria are, at least to some degree, involved in scent gland odour production and group specific odours, thus providing support for the fermentation hypothesis of chemical recognition.

The hazel dormouse has scent glands under the chin and in proximity to the anus [169] and it is possible that the scent produced could enable chemical communication and thus, kin discrimination. Previous studies have shown that scent gland bacterial composition is correlated with group membership [168] [98]. However, it has not been demonstrated that relatives share similar scent gland bacterial profiles, which is the hypothesis for this study.

It is important to note, that it is not only scent gland odour that influences social behaviour, mate choice and kin recognition. The animal gut also contains large and commonly biodiverse bacterial communities [170][171]. Although these microbial communities have a physiological function in digestion and immune defence, they
Bacterially Mediated Scent Production as a Possible Mechanism for Kin Discrimination

can also effect aspects of the host’s behaviour [170][171]. Recent studies show the importance of bacteria in animal recognition [172] and mating preferences [173][174]. Other studies have shown that social behaviour can alter gut bacterial communities [175] and even that the removal of gut bacteria can impede social development [176].

For example, nestmate recognition in termites appears to be driven by gut microbiota. Matsuura [172] demonstrated both that the termite *Reticulitermes speratus* has colony specific intestinal microbial communities and that these communities are used for nestmate recognition. When termites adsorbed unfamiliar odours from bacteria sampled from another colony, they were fiercely attacked by their own nestmates. The removal of gut bacteria via antibiotic inoculation successfully changed the recognition behaviour towards nestmates, demonstrating the essential role bacteria play for termite nestmate recognition.

In *Drosophila*, research also implicates the role of gut bacteria in social attraction [177], mate choice [178], mating investment and kin recognition [171]. Mating preferences of *Drosophila* appear to be based upon similar gut bacterial profiles, due to the same environmental conditions (such as diet), thus through familiarity [171]. However, *D. subobscura*, a species that only mates once in a lifetime, show a strong effect of relatedness in copulatory investment, with investment lessening with increasing relatedness, indicating kin recognition abilities [171]. In other species, which multiple mate such as, *D. pseudoobscura* and *D. melanogaster*, no such relationship was found, but gut bacteria was still indicated as an important identifier of mates that shared a similar diet or environment [171]. Thus, gut bacteria in fruit flies appears to influence mate choice and copulatory investment, either through familiarity or true kin recognition. Therefore, both environmental conditions and host genotype effect the structure of gut microbiota colonies and the evolved system of kin-recognition appears to be due to the ecology and behaviour of the species.

In humans, gut bacteria is highly variable from person to person, but family members tend to harbour more similar microbiota of related individuals and some strains are shared among family members [179]. It still remains unclear, however, whether the similarity between microbial communities of related individuals is due to a shared environment and/or reflect genetic relatedness [179].

Thus, when investigating the possible link between relatedness and bacterial communities it is important to consider the roles that both genetic and environmental cues play in influencing the composition of bacterial communities invoked in olfactory communication.
3.2 Kin recognition mechanisms

3.2.4 Factors affecting microbial composition of the scent glands and gut

A complex interplay of host genetic and environmental factors determines the bacterial communities established within the scent gland and the gut. Related individuals have been shown to harbour more similar bacterial communities than unrelated individuals [180]. However, this could be due to the mechanisms of initial bacterial inoculation, host genotype or due to a shared environment [179]. If true kin recognition occurs via bacterial production of odours, then genetic factors would need to be the dominant influence on commensal microbial composition. However, it is possible that mating preferences and favourable treatment of relatives through odour could be due to the likelihood of similar environment and/or bacterial transmission via social behaviour.

The effect of the host on microbial composition

Initial colonisation by bacteria occurs through transmission from mother to offspring [180]. Bacterial diversity of neonates is initially low and has been found to increase with age, which is accompanied by a shift in the dominant bacterial species. In humans, once this shift has occurred, it is thought that bacterial composition remains relatively stable and resilient to stresses, such as antibiotic treatments [181]. It is likely that a similar pattern is observed in other mammals and could provide a stable recognition mechanism.

Most genes that have been shown to have an impact on the composition of the gut have been components of the immune system [179]. Thus it is reasonable to suggest, that the genes coding for the immune system play a role in the establishment of bacterial communities within scent glands. If odour produced by bacterial action is responsible for true kin recognition cues, then the most likely mechanism linking host-specific bacterial communities to reliable indicators of relatedness is immunological.

New-born mammals are limited to the protection provided by the passage of antibodies from the mother both in utero and through the milk, as well as through the innate immune system. The innate immune system provides non-specific protection in the form of antimicrobial proteins produced by epithelial cells, which directly kill or inhibit the growth of microorganisms [182][183]. This is particularly important in those areas that are directly linked to the external environment, such as the skin and gut and thus could greatly influence the commensal bacterial communities of those areas.
The acquired immune system provides specific immunity and develops with age. Importantly, the major histocompatibility complex (MHC) of the immune system has been found to influence odours in all vertebrate groups [184] and many studies have found that MHC influences odour, mating preferences and cooperation in rodents [185][80][186].

The MHC comprises a large chromosomal region with over 200 coding loci and contains what are referred to as class I and II genes, which control all specific immune responses, such as fighting infectious disease [187]. The MHC also contains genes that influence growth, development, reproduction, odour and olfaction [187].

The MHC class I and II genes are the most polymorphic genes known among vertebrates, with as many as 170 alleles per locus in humans and 100 alleles per locus in house mice (Mus musculus)[187]. In many species class I and II MHC loci are closely linked on the same chromosome within an MHC cluster, thus there is a low chance of recombination separating them and suggests the close link is maintained by natural selection, but the mechanism remains unclear [187].

The high levels of polymorphism means most individuals, even close relatives, have a unique combination of alleles, but, as relatives should share some alleles through common decent the chance of sharing MHC alleles should lessen with decreasing relatedness [188].

Studies have found MHC-dissortative mating preferences in house mice (Mus musculus) [189][190][186][187], humans [191] and Salmon (Salmo salar)[192]. These mating preferences have been attributed to females increasing the genetic compatibility of mates in an attempt to prevent inbreeding or outbreeding, to produce the fittest offspring [187] and as a result of such studies, it has been suggested that MHC genes could be the elusive genetic component needed in true kin recognition.

There are many hypothesis exploring how the MHC contributes to odour profiles. The MHC may be directly responsible for the production of odours: through MHC molecules, by the molecules acting as carriers for volatile odourants or through the metabolites of MHC-bound peptides; or indirectly by shaping an individual’s microbial flora by providing differing environmental conditions through, for example, the production of individually distinct peptides and metabolites, which are then metabolised by the bacteria [187].

A strong relationship between concentrations of volatile compounds (odour) found within scent glands and MHC diversity has been identified in lemurs Lemur catta[188] and mandrils (Mandrillus sphinx) [193]; and the composition of both gastrointestinal and scent mark microbial communities varies with MHC type in otherwise genetically
identical mice [184][194], which supports the idea that the MHC contributes to scent and may influence bacterial community composition. However, it is important to note that the research is conflicting and other genetic loci, including those linked to the MHC have not been ruled out as the primary influence on odour profiles [188]. It has been shown that background genotype, as well as the MHC haplotype, contributes to microbial and volatile signals in murine scent marks [194]. Hurst et al. [93] found that MHC-associated odours were neither necessary nor sufficient for scent owner recognition of territorial urine scent marks. Implying research is still needed to disentangle the effect of the MHC and background genotype on odour production. However, these studies do lend support to the hypothesis that there is a genetic basis to odour production, which could be important in mate selection or social behaviour and, therefore, kin discrimination.

3.2.5 Environmental factors

The host’s genetics clearly play a role in shaping bacterial communities, at least in some species; however, a number of other factors can influence microbial diversity such as, mother-offspring transmission, social contact, shared environments, as well as diet [179][98]. These factors are difficult to disentangle from each other as litter-mates and social group members are usually related, having direct physical contact with each other in a shared environment and consuming a similar diet. Therefore, it is important to take account of all these factors when considering host bacterial composition as a possible signal for true kin recognition.

Maternal transmission of microbiota, in one form or another (e.g., egg smearing in insects), is considered to be a universal phenomenon within the animal kingdom [195]. In mammals, inoculation can be provided through the mother before birth [195], through the birth canal [196] and via the mother’s milk [197]. This is known as the maternal effect and can greatly influence composition and diversity of a host’s microbiota [179].

Research has shown that monozygotic twins share more similar gut bacterial profiles than with their marital partners [198]. The same study found that dizygotic twins and siblings had gut microbial profiles as similar to each other as monozygotic twins, suggesting a strong maternal effect on gut microbial composition and that a shared environment and direct contact is not influential enough to overcome this initial inoculation. In addition, genetically identical mice harbour more similar bacterial profiles with individuals from the same litter, as opposed to genetically identical mice from a different litter [179], again clearly indicating a maternal effect.
Continued social contact does appear to be a driver in shaping microbial community composition, where significant correlations have been found between microbiome composition and social co-residency in a number of species [94][98][168][199]. Breeding pairs of kittiwakes, chimpanzee and baboons, each within the same social group, as well as humans in a shared household exhibit more similar microbial communities than individuals from the wider population.

The sharing of microbial genetic profiles between relatives is clearly indicative of an underlying genetic component, possibly driving behavioural patterns mediated by the immune system. Social contact between relatives is likely to confound this relationship and requires consideration in any experimental design intending to identify a correlation between genetic profiles of scent-gland bacterial communities and their hosts genetic profiles.

### 3.2.6 Aims and objectives

The main aims of this chapter were to, 1) determine whether anal bacterial community composition was correlated with the degree of relatedness of the hosts; 2) establish whether the likely cause of any relationship found was environmentally or genetically driven; 3) determine whether bacterial diversity was influenced by the hosts condition (age and sex) and/or the year in which the sample was taken.

A number of predictions have been made in light of these aims: 1. Related individuals would share significantly more similar bacterial communities (measured by the percentage of shared TRFs) than unrelated individuals; with an expected positive correlation of increasing bacterial similarity with increasing degree of relatedness. 2. If bacterial community composition is more similar in relatives and the relationship is due to a genetic element, rather than through the mother effect, then offspring would be expected to display the same relationship with their father as with they do with their mother. 3. If true kin discrimination mechanisms through bacterially mediated scent production are in operation, then the positive correlation found would be due to relatives sharing more alleles (genetics) in common than non-relatives, rather than due to a shared environment and bacterial transmission through social contact. 4. Under the assumptions that the hosts immune system either directly or indirectly shapes the bacterial community and that a mature immune system develops with age: a) Lower levels of both alpha (species richness and evenness) and beta (species present) diversity would be expected in the age class grey compared with juvenile or adult individuals; and b) greys would contribute the most variation observed for the diversity indices, with adults and juveniles bacterial community composition being most similar. 5.
3.3 Methods and materials

3.3.1 Sample collection

The bacterial DNA was obtained from captivity-bred dormice of known pedigree resident at the Wildwood-Trust, Kent during September 2013, April 2014, September 2014, May 2015 and September 2015 under the licence (numbers 20122082 and 20113857). Animals were housed outside in standard wooden and wire mesh pens used for captive breeding of dormice (see [7]). Those individuals used for breeding were housed in pairs, with one female and one male adult and as such any offspring produced were of known parentage. Both dams and sires were kept with their litter throughout the active season when only one litter was produced. If a second litter was produced by a breeding pair, the first litter was placed in a separate pen with their sire once the second litter had reached 9g (grey furred, with eyes opened) to keep disturbance to a minimum when the second litter was very young. Samples were taken from three families consisting of two generations (parents and offspring), one family consisting of four generations (see Figure 3.1). Anal swabs were taken using sterile cotton buds dampened with distilled water. The tail of the animal was gently pulled back (see Figure 3.2(a)) and the dampened cotton bud was placed directly onto the anus and rotated gently (see Figure 3.2(b)). The swab samples taken in 2013 and 2014 were then placed in a 1.5ml plastic collection tube containing 1ml of S.T.A.R buffer solution (Roche Diagnostics) and frozen within four hours of collection at -20°C. Unfortunately S.T.A.R buffer has been discontinued and, thus, was not available for the storage of 2015 samples. These samples were placed directly into the PowerBead tubes provided by the PowerSoil® DNA Isolation Kit (MO BIO laboratories) and put into the fridge within 4-6 hours of collection. Samples collected in this way were extracted the next morning, after allowing the PowerBead tubes to reach room temperature.
Bacterially Mediated Scent Production as a Possible Mechanism for Kin Discrimination

Fig. 3.1 The pedigree depicting the Woods family, which was a four generation family sampled from the Wildwood Trust, with sibships in the same colour.
3.3 Methods and materials

Fig. 3.2 Photographs taken at the Wildwood Trust a) demonstrating how the anus was exposed and b) demonstrating how the swabs were taken.
3.3.2 Extractions

**Bacterial extraction protocol - Qiagen method**

Swab sample extraction from swabs collected during 2013 and 2014 required an initial digestion step to maximise DNA yield, before continuing with the DNeasy Kit (Qiagen). The swab samples were briefly vortexed, then 200 µl of the S.T.A.R. buffer solution was removed and placed into a clean 1.5 µl collection tube. 200 µl of a master mix containing 20mM Tris HCL, 1.2% Triton X and Lysozyme to 20mg/ml was added to each sample and vortexed thoroughly. 20 µl of proteinase K was added to each sample, again vortexed thoroughly and incubated at 56°C in a water bath for 30 minutes. 200 µl of Buffer AL (Qiagen) was added to the sample and vortexed thoroughly before being placed back into the water bath for a further 10 minutes. 200 µl of absolute ethanol was added to the mixture and vortexed thoroughly again. The Qiagen tissue extraction protocol was followed from step 4.

**Bacterial extraction protocol - PowerSoil method**

The PowerSoil® DNA Isolation Kit (MO BIO laboratories) was used to extract DNA from the anal swabs collected during 2015. The extraction was performed by following the manufacturer’s instructions.

**PCR amplifications**

PCR amplifications were carried out in 25µl reactions containing ≈5ng of DNA template, 10 pmol forward and reverse primers (see below) and 10 µl QIAGEN multiplex PCR mix (Qiagen).

Internal fragments of 16S rRNA genes were amplified from the isolated DNA using universal bacterial primers 341f (5’ - CCTACGGGAGGCAGCAG - 3’; [200]) and 926r (5’ - CCGTCAATTCMTTTRAGTTT - 3’; [201]). Primer 341f was labelled at its 5’ end with the dye 6-carboxyfluorescein (6-FAM) and primer 926r was labelled with hexachloro-6-carboxyfluorescein (HEX). Cycling was performed in Techne TC-4000 Thermal cycler under conditions of initial denature for 3 mins at 95°C and employing 35 cycles (30 s at 94°C, 1min s at 55°C and 1 min at 72°C), with a final extension at 72°C for 10 minutes. PCR amplifications were repeated for each sample, to enable the production of a consensus at the analysis stage.
3.3 Methods and materials

3.3.3 Digest

Terminal, Restriction Fragment Length Polymorphisms (T-RFLPs) provide a relatively simple and cost-effective means of evaluating bacterial community diversity. It achieves this by using a restriction enzyme, in this case *MspI*, to cut the DNA at the position of the closest restriction site (CCGG for *MspI* enzyme) to the fluorescently labeled end of the amplified DNA. The length of this labeled terminal fragment of the *MspI* digest is expected to vary across bacterial taxa as a consequence of substitutions within the restriction site.

Fluorescently labelled PCR products were digested with *MspI* restriction enzymes following the Fast Digestion Protocol (Thermo Fisher Scientific), incubating the PCR product in a water bath for 5 minutes at 37°C.

3.3.4 Analysis

T-RFLP analysis

The T-RFLP products were genotyped as fragments by SourceBioscience (Nottingham, England). T-RFLP profiles were imaged using Peak-Scanner software (ThermoFisher). Only peaks between 50 and 500 base pairs (bps), with heights exceeding 50 fluorescence units were evaluated, following the recommendations of Sin et al [94].

T-RFLP profiles were aligned using T-Align [202], which identified all fragments to within ±0.5bp in all profiles generated and determined the consensus sizes of the fragment length polymorphisms or Terminal Restriction Fragments (TRFs) between two duplicate samples. The resulting consensus TRF profile can be likened to a DNA fingerprint of each individuals’ bacterial community, allowing comparisons of similarities and dissimilarities among individuals. A search for the presence or absence of identical TRFs across all samples was then performed, calculating the percentage fluorescence intensity of the peaks, which is used as a proxy for species abundance.

Comparing extraction techniques

Two different extraction techniques were employed for the 2015 samples, which were compared in order to ascertain whether both had comparable coverage of the bacterial species. Out of the total 120 TRFs (see subsection above for definition of TRF), 98 were found within the samples extracted using the Qiagen method, compared to 64 found within samples extracted using the PK soil extraction kit and a total of 42 TRFs were found to be common. A consensus dataset was created, which only included the
42 TRFs found to be present using both extraction techniques. Analysis was performed on both the consensus dataset and the basic (all 120 TRFs regardless of extraction technique) dataset in order to establish the necessity of a consensus dataset.

### 3.3.5 Bacterial communities and relatedness

All further statistical analysis was performed using the statistical package R [203]. In order to determine whether dormouse relatives have more similar bacterial profiles than non-relatives, the percentage of shared TRFs between all individuals was calculated and correlated with genetic distance, inferred from genetic relatedness between dormice, estimated using their known pedigrees. Assuming the founding individuals of any pedigree are non-inbred, relatedness between any two individuals in a pedigree was calculated using path analysis [204]: if the number of paths (sum of generations back to a common ancestor between the individuals in question) is $n$, relatedness $= r = (1/2)^n$.

The percentage of shared TRFs between all captive individuals was calculated by dividing the number of shared TRFs for each pair, by the total number of TRFs of the two individuals being compared. Both the percentage of shared TRFs and genetic relatedness estimates were converted into pairwise matrices. Mantel tests were performed using the ADE4 package [205], which produced simulated P-values based on 9999 replicates. The data was split by age, sex and parent-offspring groups in order to explore the data further and linear regressions were then performed when clarification was needed.

### 3.3.6 Environment vs genetic variation

To identify whether the relationship between relatedness measures and bacterial community structure was due to a shared environment, pairs of related and unrelated individuals housed in the same pen were compared with pairs of related and unrelated individuals housed separately. Linear regression models were used to establish the relationship.

### 3.3.7 Bacterial 16S rRNA diversity measures

The covariables considered in addition to genetic variation at the 16S rRNA region were age, sex and year sampled. Age was defined as: a) grey - 6 to 40 days old or 2.5 - 10g b) juvenile - 41 days to 8 months or 10 - 12g c) adult: > 8 months or > 12g and after first hibernation.
In order to determine whether these groupings had any influence on bacterial community’s genetic diversity, specific measures were used: Bacterial complexity was described using (i) alpha diversity (species richness and evenness), (ii) beta diversity (species present) and (iii) community structure. Alpha diversity was measured using total number of TRFs (T) per individual (species richness) and, for comparison with other studies, the widely employed Shannon index (H’) (species richness and evenness). Beta diversity was described using presence/absence data and the relative abundance of each species described the structure of the bacterial communities.

The alpha diversity indices for each grouping were compared using a non-parametric Kruskal - Wallace test. The beta diversity and structure were compared using a PERMANOVA (nonparametric multivariate analysis of variance) with 9999 permutations based on Bray - Curtis distances, available through the VEGAN package using the Adonis function in R [206]. non-parametric tests were used as the data did not fit the assumptions of parametric tests.

To determine which categories within the groups (e.g. grey - adult, grey - juvenile, juvenile - adult) contributed most to the variation in the T and presence/absence data, Jaccard distance measures were calculated and visualised using a dendrogram [207].

3.4 Results

3.4.1 Comparing extraction techniques

The composition of anal bacterial communities of hazel dormice were determined by T-RFLP analysis of the 16S rRNA gene. Analysis of 40 individuals collected in September 2013, 2014 and 2015 established a total of 120 TRFs from the anal swabs.

When analysing the basic dataset of TRFs, the mean number found per individual was 19.6 (range 5 - 49, SE ± 1.8). When only considering the consensus dataset of TRFs produced by both extraction techniques, the mean number found per individual was 11.5 (range 2 - 34, SE ± 1.23). None of these TRFs were found to be present in all 42 individuals, but the most common TRF was a 54.35bp fragment and this was present in 30 individuals.

3.4.2 Correlation between host relatedness and TRF profiles of the scent-gland

Pairwise mantel tests were performed in order to establish whether related individuals of known pedigree share more TRFs in common than unrelated individuals. The
percentage of shared TRFs was correlated with the degree of relatedness between individuals. A significant result was found when all individuals sampled during the month of September were included, for both the basic TRF dataset and the consensus 42 TRF dataset (see Table 3.1). The data was then split into biologically relevant categories (age, sex, parent-offspring combinations) to further investigate whether hormonal, age related and maternal effects were at play. Females, offspring, parents and offspring, dams with greys, dams with juveniles, sires with greys and sires with juveniles all produced significant results when considering the basic TRF dataset. The same result was found when considering the consensus TRF dataset with the exclusion of the female category, which was no longer significant (see Table 3.1).

When including the basic dataset, a significant result was obtained for females (p=0.0279) but not for males (0.0522), thus indicating a predictive relationship between degree of relatedness and percentage of shared TRFs in females, but not in males. The linear relationship is presented graphically in Figure 3.3, which highlights the similarities between the two relationships. When considering the consensus dataset, females did not show a significant relationship (p = 0.199), whereas males did (0.0498).

A significant relationship was obtained for both dams and sires with juvenile offspring and with dams and sires with grey offspring (p < 0.05). The visualisations of the relationships can be seen in Figures 3.4 and 3.5. Both graphs depict slight positive correlations, with sires and dams displaying the same relationship with juveniles. The same relationships were found when considering the consensus dataset, depicted in
Figures 3.6 and 3.7. However, a slightly steeper slope (slope = 0.278) was observed for dams and greys, when compared to sires and greys (slope = 1.68), (see Figure 3.5).

![Graph showing percentage of shared TRFs and degree of relatedness for both sexes.](image_url)

**Fig. 3.3** Percentage of shared TRFs and degree of relatedness using the basic dataset for both sexes.
Fig. 3.4 Percentage of shared TRFs compared with degree of relatedness using the basic dataset for dams and sires with juveniles.
3.4 Results

Fig. 3.5 Percentage of shared TRFs compared with the degree of relatedness using the basic dataset for dams and sires with greys.
Fig. 3.6 Percentage of shared TRFs compared with degree of relatedness using consensus dataset for dams and sires with juvenile offspring.
3.4 Results

Fig. 3.7 Percentage of shared TRFs compared with degree of relatedness using consensus dataset for dams and sires with grey offspring.
3.4.3 Environment vs genetics

To investigate the relative contributions of genetic similarity and a shared environment to bacterial profile similarity, the percentage of shared TRFs amongst related individuals occupying the same and different pen and unrelated individuals in the same and different pens were compared. Neither relatedness nor environment was found to correspond significantly to the percentage of shared TRFs when the basic dataset was included in the analysis (degree of relatedness, \( p = 0.56 \); environment, \( p = 0.86 \)). However, when analysing the consensus dataset, relatedness with the percentage of shared TRFs was found to be significant (relatedness, \( p = 0.028 \); environment, \( p = 0.9 \)). Figure 3.8 shows the relationship between the percentage of TRFs shared between relatives and non-relatives housed in separate and the same pens compared with degree of relatedness. Figure 3.9 depicts the relationship between the percentage of TRFs shared between relatives and non-relatives housed in separate and the same pens compared with environment. The steepness of the slope in Figure 3.8 clearly demonstrates a significant and predictive relationship between degree of relatedness and the percentage of TRFs in common between relatives, regardless of the housing situation. Whereas the slope in Figure 3.9 is almost flat, showing no relationship between a shared environment and the number of TRFs in common.

3.4.4 Diversity

Two different diversity indices were used to describe alpha diversity for the different age classes, sexes and years. The total number of TRFs (\( T \)) was used for richness and Shannon index (\( H' \)) was used to take into account both richness and evenness. When the basic dataset was included in the analysis, diversity was highest for \( T \) and \( H' \) in adults, females and 2013, however the differences within each group were not significant (see Table 3.2; Figure 3.10). When analysing the consensus dataset, a different result was observed. Alpha diversity varied significantly with year with both \( T \) and \( H' \). Age was a significant factor for bacterial community complexity when tested using \( T \) (see Table 3.3). Jaccard’s index scores showed that adults and grey offspring had more similar bacterial profiles to each other than to juvenile offspring (see Figure 3.11).
3.4 Results

Fig. 3.8 Percentage of shared TRFs compared with relatedness for relatives and non-relatives housed in the same and separate pens using the consensus dataset.
Fig. 3.9 Percentage of shared TRFs of relatives and non-relatives compared with environment, where 1 are those housed together and 2 are housed separately using the consensus dataset.
3.4 Results

Fig. 3.10 A cluster dendogram of the Jaccard index scores of dissimilarity for T in 2013 (1), 2014 (2) and 2015 (3) using the basic dataset
Fig. 3.11 A cluster dendrogram of the Jaccard index scores of dissimilarity in adults (1), juveniles (2) and greys (3) using the basic dataset
### Results

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean No. TRFs</th>
<th>TRFs</th>
<th>H'</th>
<th>TRFs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>KW</td>
<td></td>
<td>KW</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\chi^2$</td>
<td>p-value</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>11</td>
<td>22.55 ± 3.98</td>
<td>2.049</td>
<td>0.359</td>
<td>2.47</td>
</tr>
<tr>
<td>Juveniles</td>
<td>17</td>
<td>20.47 ± 2.71</td>
<td>2.96</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Greys</td>
<td>12</td>
<td>15.67 ± 2.48</td>
<td>2.85</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>2.63</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>21</td>
<td>22.05 ± 2.67</td>
<td>2.142</td>
<td>0.1433</td>
<td>2.14</td>
</tr>
<tr>
<td>Males</td>
<td>19</td>
<td>16.89 ± 2.26</td>
<td>2.94</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td></td>
<td>2.68</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>11</td>
<td>22.91 ± 3.94</td>
<td>1.131</td>
<td>0.5681</td>
<td>1.13</td>
</tr>
<tr>
<td>2014</td>
<td>16</td>
<td>16.63 ± 1.79</td>
<td>2.97</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>13</td>
<td>20.46 ± 3.80</td>
<td>2.73</td>
<td>0.41</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 For the basic dataset, mean number of TRFs and mean $H'$ scores for each category and the Kruskal-Wallace outputs for total TRFs and $H'$ scores correlated with the categories age, sex and year.
Bacterially Mediated Scent Production as a Possible Mechanism for Kin Discrimination

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean No. TRFs</th>
<th>KW $\chi^2$</th>
<th>p-value</th>
<th>Shannon’s Index $\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>11</td>
<td>13.55 ± 2.96</td>
<td>12.95</td>
<td>0.0015</td>
<td>2.37 ± 0.22</td>
<td></td>
</tr>
<tr>
<td>Juveniles</td>
<td>17</td>
<td>12.06 ± 1.72</td>
<td>2.33 ± 0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greys</td>
<td>12</td>
<td>8.83 ± 1.85</td>
<td>2.02 ± 0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>2.98</td>
<td>0.085</td>
<td>3.15</td>
<td>0.076</td>
</tr>
<tr>
<td>Females</td>
<td>21</td>
<td>22.05 ± 2.67</td>
<td>2.42 ± 0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>19</td>
<td>16.89 ± 2.26</td>
<td>2.04 ± 0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td></td>
<td>8.2</td>
<td>0.017</td>
<td>8.74</td>
<td>0.013</td>
</tr>
<tr>
<td>2013</td>
<td>11</td>
<td>22.91 ± 3.94</td>
<td>2.39 ± 0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>16</td>
<td>16.63 ± 1.79</td>
<td>1.87 ± 0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>13</td>
<td>20.46 ± 3.80</td>
<td>2.56 ± 0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3 For only the TRFs produced by both extraction techniques (consensus dataset), mean number of TRFs and mean Shannon Index scores for each category and the Kruskal - Wallace outputs for TRFs and Shannon Index scores correlated with the categories age, sex and year.
3.4 Results

Beta diversity ($\beta$) of anal swabs was measured with presence/absence data and community structure was measured with relative abundance (RA) data, both of which varied significantly with age when the basic dataset was analysed ($\beta$: $F_{37:2} = 7.99$, $R^2 = 0.30$, $p = 0.001$; RA: $F_{37:2} = 5.52$, $R^2 = 0.23$, $p = 0.001$). The Jaccard index scores of $\beta$ diversity were more similar between adults and greys than each was with juveniles, the same relationship found with $\alpha$ diversity (see Figure 3.11).

When only including the consensus dataset, $\beta$ and RA varied significantly with year ($\beta$: $F_{38:1} = 0.140$; $p = 0.001$; RA: $F_{38:1} = 0.140$, $p = 0.0002$). $\beta$ diversity was more similar in the years 2013 and 2015 than either were to 2014, see Figure 3.12.

Fig. 3.12 A cluster dendogram of the Jaccard index scores of dissimilarity in 2013 (1), 2014 (2) and 2015 (3) using the consensus dataset.
3.5 Discussion

3.5.1 Relationship between genetic relatedness and bacterial community structure

If bacterial diversity provides a mechanism for kin discrimination, then bacterial community profiles would be expected to be more similar between related individuals than unrelated individuals. Females and males were analysed separately, as hormonal differences between the sexes have been found to influence scent gland bacterial profiles [208][94]. Therefore, there is a possibility that one sex, but not the other, has gender specific bacterial species. When considering the basic dataset, females displayed a significant, positive correlation, with the percentage of shared TRFs, increasing with degree of relatedness (p=0.03), but males showed no significant relationship (p=0.05). However, when considering the consensus dataset, the opposite trend was found with males displaying a significant, positive relationship (p<0.05) and females showing no significant relationship (p=0.2). The consensus dataset was constructed from the common TRFs of two extraction procedures to reduce noise and, thus, was expected to give a more trustworthy result. The basic dataset was included to establish whether or not a consensus dataset was necessary, as if both datasets gave the same result, then the basic dataset could be deemed dependable. The differing results found between the datasets for males and females indicates the need for a consensus and thus only the consensus dataset is considered for discussion.

Pairwise Mantel tests revealed a significant correlation between the relatedness of dormice and the number of shared TRFs (p = 0.0013), indicating that the number of TRFs shared increases with increasing genetic similarity. A number of reasons could explain this relationship. Firstly, it could be that relatives generally share more alleles in common than do non-relatives, and thus have more similar immune systems due to sharing of MHC alleles, which in turn influences the type of bacterial communities that are tolerated by the host. Secondly, the relationship could be due to a mother effect, where mothers have more similar bacterial profiles to their offspring due to the initial bacterial inoculation via exposure to the mother’s birth canal and the mother’s milk. Thirdly, it could be due to a shared common environment, as close relatives are often housed together, as is the case for dormice at the Wildwood Centre, where they are in constant physical contact with each other, allowing bacteria to be passed via their social behaviour [89].

To examine the possibility that the relationship found between relatedness and shared TRFs was due to the presence of individuals with a mature immune system,
adults and offspring were considered separately. Interestingly, only the offspring were found to have a significant relationship (adults p = 0.77, offspring p = 0.0008). Further exploration, whereby pairwise Mantel tests were performed including parents and offspring (n=37) (i.e. excluding those adults that did not reproduce) revealed a significant correlation (p=0.0003). Thus, the lack of significance found for adults alone was ascribed to the low sample number (n=11) and the lack of adult relatives available for sampling.

Close relatives are typically housed together at Wildwood, however, when a breeding pair has two litters, the second litter is kept with the mother in the original pen and the first litter (post-weaning) is housed separately with the father in a new pen. Out of the nine litters sampled, six of them were separated in this way. This meant that it was possible to further explore the influences of the mother effect, maturation of the immune system, a shared environment and genetic similarity on the observed significant correlation. The data was further divided into dams with grey offspring, dams with juvenile offspring, sires with grey offspring and sires with juvenile offspring (see Table 3.1).

There can be no doubt that initial colonisation of bacteria from the mother during birth structures the bacterial communities found in the offspring, as this has been established a number of times this influence persists to some degree into adulthood [209]. However, it cannot be assumed that the bacterial profiles would necessarily correlate between them more so than between individuals. In addition, different scenarios predict quite different patterns of relationship between offspring bacterial profiles and that of their fathers. For instance, if the mother effect was the only determinant of bacterial community structure, then no significant relationship would be found between relatedness and bacterial communities for sires with their offspring. If bacterial communities are determined initially by the mother effect, but later correspond more to the genetics of the offspring due to the maturation of their immune system, then sires may be expected to correlate with their juvenile offspring, as it is likely that grey offspring lack an acquired immune system (however, this has not been investigated to date). If more similar bacterial profiles occur due to a shared environment and social contact, such as bacterial transmission between individuals in a shared pen, then the same scenario would be expected as above, with sires only sharing a significant percentage of TRFs in common with juveniles. Whereas, if bacterial communities are determined by a genetic component, then a comparable relationship would be expected with both categories of age group and sires.
Significant positive correlations were found between percentage of TRFs in common and degree of relatedness, both with dams and sires with both categories of offspring \((p \leq 0.05)\) with all parent/offspring combinations showing a significant, positive correlation between degree of relatedness and the percentage of shared TRFs. As a significant relationship was found for both sires and dams, this indicates that the similarities seen between bacterial communities of relatives cannot merely be due to the mother effect or the maturation of the immune system, but a genetic component in shaping bacterial communities. The slope of the line of best fit was similar for both dams and sires with juveniles (see Figure 3.6), however the slope is steeper for dams with greys, than sires with greys (see Figure 3.7), indicating a stronger positive relationship. This could be due to a combination of the mother effect and an immature immune system or the mother effect and a shared environment, but disentangling these factors requires further study. However, the relationship is complicated by the fact that sires and first litters were only separated from dams and second litters, when the second litters had reached 9 g in weight. This means that the fathers shared an environment with all their offspring for the first 20 - 40 days of the second litters' life [13]. This contact could be enough for bacterial transmission from the father to the offspring, meaning that the explanation of a shared environment cannot be ruled out.

From these results, it is clear that the percentage of TRFs shared between individuals and thus, bacterial community similarity, increases with degree of relatedness. Consequently, the fermentation hypothesis could be a viable mechanism for kin discrimination. What cannot be confirmed is whether the relationship is truly due to genetic similarity influencing bacterial composition, either directly or indirectly, or it is simply due to the sharing of bacteria through social behaviour. The latter could still result in the ability to use odours produced by bacteria to discriminate between known kin and unknown kin and unrelated individuals, a mechanism known as familiarity [145]. But only the former could be employed if it is necessary to discriminate between unrelated individuals and unknown kin; a mechanism known as phenotype matching [145]. The hazel dormouse displays a number of behaviours that would indicate the necessity of the evolution of phenotype matching and hence was investigated further.
3.5.2 Genetic component versus a shared environment: which one explains the relationship between bacterial community structure and relatedness?

In order to investigate further the possibility that related individuals have more similar bacterial profiles due to a higher number of shared alleles, or due to cross contamination within a shared environment, the percentage of TRFs shared between unrelated pairs housed together; unrelated pairs housed separately; related pairs housed together and related pairs housed separately, were compared to degree of relatedness and environment (same or separate pens) using a one-way ANOVA. If the similarity of bacterial profiles was influenced by a shared environment, then those individuals sharing an enclosure should have more similar profiles that those in separate enclosures, regardless of genetic relatedness. The relationship between TRF profile similarities and relatedness was found to be significant (p = 0.03). The regression model was found to be reasonable ($R^2 = 0.41$) and the relationship is depicted clearly in Figure 3.8, with slope = 0.412, between the percentage of shared TRFs. No relationship was found between TRF profile similarities and a shared environment (see Figure 3.9), indicating that the enclosure had no effect on the percentage of shared TRFs between individuals. The fact that the relationship was uncovered using a small sample size (only 10 pairs, to prevent pseudo-replication) and with fewer TRFs, suggests a large effect. Therefore, it appears that the relationship of increasing bacterial similarity with increasing relatedness is due to genetic similarities between relatives, rather than a shared common environment, which is consistent with true kin recognition mediated by the fermentation hypothesis.

3.5.3 The influence of age, sex and environment on bacterial community composition

Many environmental and physiological conditions can influence bacterial community diversity and composition. Both alpha (α) (i.e. TRF richness and evenness) and beta (β) diversity (i.e. species presence) and bacterial community structure (i.e. relative abundance of each species) were examined in relation to sex, age and year. The hypotheses being that bacterial communities: vary with age, due to the maturation of the immune system; differ between the sexes due to hormonal differences and that differences between years occur due to variation in extraction techniques and/or environmental fluctuations in climate.
The composition and complexity of bacterial communities within each sex did not differ significantly, which could be due to low sample numbers (female $n = 21$; male $n = 19$). Alpha diversity levels differed between age categories for species richness ($T$), with the highest found in adults and lowest in grey offspring (see Table 3.3), however the relationship was only significant with richness alone. Both beta diversity levels and bacterial community structure also varied significantly with age. Age-related differences of bacterial communities of the scent gland have been reported in badgers ((Meles meles), meerkats (Suricata suricatta) and striped hyenas (Hyaena hyaena) [94]; [168] [98]. Alpha diversity was reported to be significantly higher in cubs compared to adults in badgers, with the acquisition of a mature colony with age. Cubs had more diverse, pioneering communities within their scent pouch compared to the more specific and less diverse communities established within the adult scent pouch [94]. Beta diversity and bacterial community structure differed significantly between adults and juveniles striped hyenas [98]. The variation between the two age classes was driven primarily by a core adult scent pouch community present in all adults, which was made up of 19 operational taxonomic units (OTUs), whereas only three of these core OTUs were found in juveniles [98]. Both of these studies found that diversity decreased with age, however for dormouse anal swabs, diversity increased with age, this relationship was also found for richness in female meerkats [168].

As diversity appeared to increase with age, juveniles and adults would be expected to have more similar bacterial communities. However, Jaccard distance measures for richness and composition indicated that adults and grey offspring were more similar to each other than to juvenile offspring. The similarity between adult and grey offspring could be driven by initial inoculation of bacteria from the mother (mother effect), which could govern both the diversity levels and bacterial community structure.

Only $\beta$ diversity differed significantly with year, with Jaccard distance measures revealing that the 2014 sample differed from 2013 and 2015, which clustered together (see Figure 3.10). Extremes of environment (e.g. humidity and temperature) have been found to effect gram negative bacteria on the skin [210]. The same could be true for the dormouse if environmental conditions at the time samples were taken in 2014, differed considerably from when samples were taken in 2013 and 2015. However, conditions in the summer and autumn of 2013 were much warmer and dryer than those in 2014 and 2015, which had much cooler and wetter conditions [211], therefore this explanation is unlikely to be feasible. More likely, the difference was due to sample number (2013 = 11; 2014 = 16; 2015 = 13) as more individuals were sampled in 2014.
The current study found related dormice had more similar anal swab bacterial communities than unrelated individuals; and that this relationship was genetically driven, rather than due to repeated contact via a shared environment. Thus, bacterial mediated scent production could be a viable mechanism for true kin recognition through phenotype matching, however whether the bacterial species present represented known odour producers would need to be determined. To date, no behavioural studies have been performed on whether dormice have the ability to recognise kin over non-kin. Kin recognition would be expected to evolve when the chances of encountering opposite sex relatives is high (inbreeding avoidance) and when social cooperation (such as créching behaviour) is employed. Sex-biased dispersal is not practiced exclusively in dormice, with opposite sex relatives found in close proximity, neither actively avoiding nor clustering with one another (see Chapter 4). In such circumstances as these, high levels of inbreeding would be expected within a random mating scenario, however low levels of inbreeding have been described (see Chapters 4 and 5), indicating the employment of other inbreeding avoidance mechanisms. As inbreeding levels were zero in Mallydams Wood (see Chapter 4), the employment of polyandry alone would be insufficient to prevent all inbreeding, thus kin recognition as an inbreeding avoidance mechanism is a plausible mechanism. Mate selection through odour has been observed a number of times in various species [180][181]. In order to resolve the use of mate selection within hazel dormouse populations further research using direct (behavioural) or indirect (genetic) methods would be required. The use of genetic methods may be a more feasible option, due to the elusive nature and conservation status of the hazel dormouse. A genetic study would still require extensive sampling of one or more dormouse populations over multiple generations to identify any bias in non-kin mating.

There is evidence of complex social interactions in the hazel dormouse [7][15][53][11] and female relatives have been found to spatially cluster, with individuals related to the degree of full sibs ($r = 0.5$) driving the observed relationship (see Chapter 4); therefore, kin recognition could be employed for social behaviour and inbreeding avoidance. However, the level of kin tolerance and sociality remains undetermined. Research into the degree of same sex home range overlap could clarify this and whether or not 'clan' odours are employed could help to illuminate whether social groups are formed. Ancillotto et al. [212] detected the use of ultrasonic communication in the hazel dormouse, with differing vocalisations between the sexes and thus such communication could also play an important role in reproductive and social behaviours.

Bacterial community diversity and structure significantly differed with age group, signifying the acquisition of a mature bacterial colony. This could imply an immune
element to the attainment of a stable bacterial community, as the acquired immunity develops with age. Further research using molecular markers known for their role in the acquired immune system response (e.g. MHC) would be necessary in order to reliably deduce the viability of this theory.

### 3.6 Conclusions and Implications

The research presented here establishes a link between the degree of genetic relatedness and anal swab bacterial community composition, eluding to the possibility of kin discrimination via bacterially produced scent. The application of such a mechanism would be expected in species where the likelihood of encountering relatives is high and social cooperation occurs within groups of related individuals, both scenarios would shape genetic variation within populations. Hazel dormouse populations have been found to contain large percentages of relatives ($r \geq 0.06$; see Chapter 4). Thus, active inbreeding avoidance would be expected to evolve. If kin discrimination through odours is used for mate selection purposes under these conditions, reduced levels of homozygosity within a population would be predicted, with excess heterozygosity found within offspring. Cooperative behaviours between relatives would result in the clumping of relatives (as found in Chapter 4) resulting in genetic structuring, together with reduced within population genetic variation and increased among population genetic variation. Further research would be required to: 1) determine whether known odour producing bacteria are present, using next generation sequencing; 2) evaluate the level of kin tolerance and mate selection in the hazel dormouse; 3) infer the ability of the hazel dormouse to discriminate kin from non-kin to determine whether true kin discrimination through bacterial mediated odours occurs.

The use of bacterial mediated odours for kin discrimination in dormice could have practical conservation implications. Reintroduction schemes release captive bred individual into areas where dormice have become locally extinct. Captive breeding programmes tend to house one female/male pair together for breeding purposes. If mate selection through odours is in operation, then allowing access to multiple potential mates may increase offspring fitness and litter success. If odours mediate social cooperation, such as behaviour, it may be useful to house female relatives together to allow communal raising of young. Also, the release of female relatives together into reintroduction sites may improve the likelihood of population persistence. The use of medications such as antibiotics and antiparasitic chemicals, could disrupt the natural conditions found within scent glands and the gut, resulting in a bacterial community change and altered
behaviours. Thus, it may be important to avoid such medicating before the release of captive dormice. If such changes in bacterial composition were found, there would be conservation implications regarding pharmaceuticals and hormone mimicking chemicals entering the water systems, which could disrupt natural social behaviours and mate selection in species that employ discrimination through odour.
Chapter 4

The Social Structure of the British Hazel Dormouse (Muscardinus avellanarius)

Abstract

Little is known about the social structure of the hazel dormouse, however some level of sociality is evident during the active season, begging the question as to what is driving this behaviour and whether it impacts a population’s genetic diversity. It has long been hypothesised by population geneticists that in the absence of dispersal, where social barriers restrict gene flow and increase isolation between groups, the population suffers an increase in its vulnerability to inbreeding and, consequently, inbreeding depression. Recent research has indicated that this prediction may not in fact hold and sociality may operate to maintain genetic diversity, which would be consistent were the individuals discriminating relatives from non-relatives for mate selection. The hypothesis being that the degree of relatedness between individuals provides an expected level of inbreeding should they mate at random. If the hazel dormouse population is socially structured and not randomly mating, then inbreeding levels should be lower than expected, even in fragmented habitats. A general pattern of female philoptrsy and male biased dispersal was found, however this pattern was not absolute. Both sexes were found to cluster with same sex close relatives (to a lesser extent in males) indicating a level of tolerance to related individuals, however the degree of kin tolerance and sociality could not be confirmed. Males were significantly more philopatric within the smaller and more isolated woodland, implying possible disruptions to natural movement, however further research with habitat replicates would be needed to verify this. Opposite sex relatives were present within close proximity to
each other at both sites, thus providing the opportunity for inbreeding. The lack of inbreeding found within the smaller, more isolated site that displayed higher levels of male philopatry implies that inbreeding avoidance mechanisms other than sex-biased dispersal are in operation.
4.1 Introduction

Little is known about the social structure and behaviour of the hazel dormouse (*Muscardinus avellanarius*) and what is known is often conflicting, stemming generally from snap shots of dormice that have been observed in nest boxes together. As the family structure, or more generally referred to here as social structure, separates populations into breeding demes, there is clear scope for such non-random mating to restrict gene flow and reduce the effective population size, resulting in unrepresentative census data and consequently a misinterpretation of a population’s health. When considering endangered species, such as the dormouse, it is therefore important to fully understand the degree to which sociality operates and the genetic consequences it imposes.

The spatial distribution of genetic variation can reveal patterns of kinship and dispersal behaviour in wild populations [213][113][112][214]. The spatial distribution of closely related individuals, i.e., those whose genotypes share a recent common ancestry, are affected by dispersal patterns [77]. Many social groups are formed through natal philopatry, with the result that related individuals are spatially clustered together in a habitat [215]. In terms of genetic structure, the persistence of closely related clusters of individuals is expected to decrease genetic variation within these groups, but increase variation among groups [213]. Most mammal species display male-biased dispersal and female philopatry [216]. This is due to the different selective forces that each sex is subjected to [214]. The reproductive success of males is considered to be more limited by mate competition [213]. For males, the cost of dispersal away from their natal site in the form of increased energy expenditure and predation risk, is proposed to be lower than the costs to individual fitness incurred through the increased risk of inbreeding and mate competition were they to stay within their natal site [217]. For female mammals, resource acquisition is considered to be an important limiting factor for reproductive success, as females invest more energy and time in their offspring to ensure greater survival, thus the risks of dispersal may, in most cases, outweigh the benefit to potentially acquiring better or more resources [213]. In light of this, the attainment of the relevant resources to increase reproductive output may be enhanced by the familiarity with natal area and could be an important evolutionary reason for the classic female philopatry situation [217]. The sedentary nature of most females is also proposed to promote social tolerance and cooperation between kin through kin selection [214]. As such, kin clustering is believed to be the initial step in developing sociality [218][219]. Many studies have shown that females of highly social mammal species live nearer their kin [219]. Females in particular, may benefit from an increased
inclusive fitness through cooperating with kin, for example through the shared care of young. For one thing, communal nesting can increase the likelihood that offspring will reach reproductive age by reducing the risk of predation, as the young are not left alone, thus directly increasing fitness; whereas caring for the progeny of close kin can indirectly increase individual fitness through kin selection [78]. For mammals, the social complexity of female kin groups varies [79]. The structuring of kin can alter in response to shifts in the costs and benefits of life history trade-offs, such as reproduction and dispersal [220]. Recent research has also established that many solitary species share more space with their kin or live close to each other [213][219][216]. Some solitary mammals have matrilineal, genetic structure similar to social species [221][220]. Thus, kin selection could be an important driver in evolutionary adaption in non-cooperative as well as cooperative species [222].

Levels of dispersal and social interactions are governed by ecological conditions, such as resource availability, which influences factors such as population density. In certain environmental conditions, where food and/or space is limited, philopatry and restricted dispersal may increase competition even among relatives, negating the benefits of kin selection [219]. Many studies have revealed a flexibility in sociality in response to environmental conditions, with species exhibiting different behaviour within different parts of their range [223]. For example, badgers in the British Isles live in much larger social groups and display higher levels of sociality compared to those of mainland Europe, due, it is believed, to the limitations of space [223]. It is also important to note that, although female philopatry and male-biased dispersal is the most common situation in mammal species, it is not exclusively the case. Some mammals display the opposite trend, with females being the dispersing sex and males remaining faithful to their natal habitat [224][225], whilst in other species, neither or both sexes disperse [226].

The spatial organisation of individuals within a population can give some indication of the mating strategies and social composition within breeding groups. Adult dormice (>12g and after first hibernation) maintain permanent home-ranges, are relatively sedentary and, as a result, form the foundation of any social organisation [11]. Male home-ranges are typically larger than those of females of which they usually overlap [11]. This arrangement generally occurs where polygynous mating strategies are adopted, which has been reported in dormice [54][50]. Female dormice appear to be philopatric, with males being the main dispersers, which is consistent with the majority of mammal populations [227]. However, females have also been found to disperse [11] and genetic studies have revealed that litters often have multiple paternity [53][50], suggesting that
the dormouse may employ a promiscuous mating strategy. Same-sex home-range overlap has been reported to varying degrees and is reported to be more common in females in Great Britain [106]. However, male overlap has been found to be more common in Russia and Lithuania [10][11]. Regardless of which sex overlaps to a greater degree, the overlap of home-ranges indicates some tolerance to neighbours.

The fitness benefits of defending territories must outweigh the fitness costs, otherwise overlapping use of space is expected [228]. It is not clear to what extent dormice are territorial, as there is evidence of aggressive displays between males on the boundary of their home-range [229]. Home-ranges are predictably smaller in high-density populations, the average size decreasing by one half when dormouse density reaches 3-4 adults per hectare, compared with one adult per hectare [230]. This flexibility could explain why in some areas dormice overlap to a greater or lesser extent. When population density increases, the costs to territory defence may be too high [231] and thus overlapping same sex home-ranges are expected. If individuals’ home-ranges overlap such that they are competing for shared resources, kin selection may reduce fitness costs to individuals. Preferential sharing of resources with close kin over non-kin is predicted to result in an inclusive fitness advantage, even if limited resources curtail direct fitness. In such circumstances, related individuals are expected to be in closer proximity to one another, as kin will be more tolerant than to non-kin.

Social organisation does, however, change throughout the active season, with cohabitation of same sex occupants occurring outside the breeding season (active season occurs between April and October, with breeding starting in May) [229][10][11]. Although, during the breeding season, cohabitation is mostly found between opposite sex pairs [10]. In captivity, reproductively active males housed together were never found sharing nest boxes [232] [233]. Thus, the onset of breeding condition in males may cause more aggressive and territorial behaviour. In meadow voles (*Microtus pennsylvanicus*), gonadal steroid secretions have been associated with increased aggression between unfamiliar males and when these steroids decrease, males aggregate into groups [234]. It could be possible that a similar effect occurs within dormice, though no research has currently been performed in this area.

The degree of sociality in dormice is unknown, although the occurrence of groups of dormice in nest boxes lends support to the idea that there is a social dimension to their behaviour [23]. It has been argued that such occurrences should be considered rare in the wild and that dormice are in fact solitary rodents [11]. However, the limited data suggests that such a position may be too strong. For example, females in captivity have been found to nest and raise young communally, in addition to there being evidence for
créching behaviour in the wild, where large litters of varying ages have been observed nesting with a single female [7][10][229]. Naim [53] also found that communally nesting wild female dormice were related and may use creches to support their young. Another Gliridae family member, the edible dormouse (*Glis glis*), does in fact communally raise young [229], however it is believed that wild hazel dormice are less communal [11].

Sociality is known to be a flexible adaptation, dependent on resource levels, competition and population density. Thus, it is possible that wild hazel dormice behave differently in different parts of their range. Understanding the social structuring of dormice populations is of particular interest due to their continuing decline in numbers in Great Britain. The decline is mainly attributed to habitat loss and fragmentation, together with increasingly warm winters through climate change. Due to habitat fragmentation disrupting dispersal, populations of the hazel dormice are becoming increasingly isolated from each other. As dispersal and kin clustering are closely linked, it is possible that populations found in differing conditions, such as differing levels of isolation, could display different spatial structuring of relatives. For example, if male-biased dispersal is disrupted by habitat fragmentation, it is expected that males in isolated populations will be spatially closer to their relatives than in populations that allow for dispersal. It may then be expected that higher levels of inbreeding ensue. Because of this threat, it has become paramount to identify the prevalence of social structuring in dormice and determine whether habitat fragmentation is indeed likely to impact upon their social structure.

### 4.1.1 Aims and predictions

The main aims of this chapter are to 1) identify any spatial genetic patterns that could clarify the social structuring within populations of the hazel dormouse; 2) establish whether levels of inbreeding found would be expected in relation to the presence of opposite sex relatives and 3) establish whether these patterns are consistent across the two sites.

A number of predictions were made in light of these aims and previous findings: 1) female philopatry and male-biased dispersal is expected. This will be observed through the spatial clustering of female relatives and the absence of male relatives (both same and opposite sex) within sites. 2) Opposite sex relatives, where present (e.g. cross-generational) will actively avoid one another. Opposite sex relatives would be expected to be significantly further away from each other than two members drawn at random from the population. 3) Levels of inbreeding are predicted to be low or absent, due to the expectation of male-biased dispersal. 4) Disruption to male-biased
4.2 Methods and Materials

4.2.1 Sample collection and processing

The sample sites are part of the long term national dormouse monitoring scheme run by the Peoples Trust for Endangered Animals (PTES) and, as such, samples were taken during the routine nest-box checks between April and October 2011-2014. All samples were collected and stored under a Natural England licence (Licence numbers 20113857, 20122082, 20140438 and 20159472). Hair samples were collected from Mallydams Wood (MW) in East Sussex (TQ857122) and Briddlesford on the Isle of Wight (IoW), Hampshire (SZ552904), UK (Figure 4.1). Mallydams wood and the Isle of Wight sites were selected due to the number of adults sampled exceeding 30 individuals, which was deemed enough for the purpose of the analysis. All other sites sampled had fewer than 30 adult individuals sampled and so were excluded from the analysis. See general methods (Chapter 2) for extraction technique, PCR amplification and genotyping methods.

4.2.2 Relatedness coefficient

Relatedness can be considered at two levels, depending on the means by which it is calculated. If pedigree data are available, then 'relatives' are identified, such as 'parents', 'offspring', 'cousins' and so on. Our knowledge of Mendelian inheritance then allows us to predict the degree to which relatives share portions of their genome through identity by descent, using pathway analysis [235]. This procedure gives us an expected relatedness, such as 0.5 between parents and offspring, 0.25 between half-sibs, and so on. However, there can be considerable variation around these expected values, for example identical twins share 100% of their genetic material, not 50%, which can be important when addressing questions that relate specifically to the genetic relatedness of the individuals. In such situations it can be preferable to measure directly the degree to which individuals share genomic variants using genetic markers. This is the realised relatedness and, clearly, has an accuracy that depends upon the number and coverage of markers throughout the genome. Due to the absence of pedigree data for our study animals, we adopted the latter approach of calculating the realised relatedness using microsatellite data. However, we adopt the approach that, on average, individuals of the
Fig. 4.1 Distribution of nest boxes with a positive result for dormouse sample collection for a) MW (22ha) and b) IoW (59.26ha)
same generation calculated to have a realised relatedness of, for example, approximately 0.5, will be referred to as "sibs"; if approximately 0.06, "cousins" and so on, with the caveat that these may sometimes be inaccurate due to the small probability that the realised relatedness value is the accumulated product of consecutive generations of inbreeding.

The genetic relatedness amongst pairs of individuals was estimated using the programme COANCESTRY V1.0.1.5 [76] in order to evaluate the spatial distribution of relatives and kin tolerance. COANCESTRY allows the computation of the relatedness coefficient using seven different methods, the triadic likelihood estimator [236], several moment estimators including: Wang [237]; LynchLi [238] [239]; LynchRd [240]; Ritland [241]; QuellerGt [242] and a dyadic likelihood estimator [243]. In order to decide on the most appropriate method for the data, individuals of known pedigree from a family group of three generations ("Woods family"), sampled from the Wildwood Trust (Herne Bay, Kent) (for details see Chapter 3) were used. The genotypes from the Woods family group were combined with genotypes from the two wild populations sampled from Mallydams (MW) and the Isle of Wight (IoW). COANCESTRY was then used to generate allele frequencies, from which relatedness coefficients were produced, along with 95% confidence intervals (bootstrapping set at 1000 iterations) for each pair, using five of the seven available estimators. The triadic likelihood estimator and the dyad likelihood estimator were not employed, as neither method provided negative values of $r$, which was considered unhelpful when populations were likely to be substructured and, hence, comprising individuals less related than expected in a panmictic population. The estimator that provided the most accurate ($r = 0.5; r = 0.25; r = 0$) and precise (the smallest confidence intervals) estimate for individuals of known pedigree was selected. Relatedness coefficients range between -1 and 1. Positive values indicate that two individuals are more closely related than expected by chance, i.e. they share alleles that are identical by descent more frequently than alleles drawn from two individuals randomly drawn from the same population. Negative values represent those which are less related than expected by chance [244].

4.2.3 Genetic vs. geographic distance

In order to explore the social composition of dormice, the spatial arrangement of relatives was assessed for adults only, as adults are considered to have settled in the location where they will remain for the remainder of their lives, whereas juveniles have a tendency to migrate from their natal location. Pairwise geographical distances between all nest boxes, from which an adult dormouse was sampled, were produced using GPS point data,
which was then analysed using the distance matrix function employing straight line distances in QGIS 2.12.0-Lyon [245]. Estimated relatedness coefficients between adults found within the same year were correlated with their geographical distance using linear models (see Table 4.2 for all comparisons made), employing the statistical software program R (3.3.2) [203]. Linear models were employed over other commonly used tests, such as the mantel test, to assess whether any relationships found were predictive, rather than simply correlated. Both straight-line ($Geographic = a(Genetic) + b$) and polynomial ($Geographic = a(Genetic)^2 + b(Genetic) + c$) relationships were explored for best fit (employing the maximum likelihood default in R). As the realised relatedness is estimated using a limited number of microsatellite markers ($n = 17$), the estimates were likely to include some degree of error. However, given that one of our aims was to identify the family structure of the local groups, these estimates were translated into conventional, pedigree, forms of relatedness such as 'sib', 'half sib' and 'cousin' etc (see Table 4.2). Both realised relatedness and recreated pedigree relatedness were compared to geographic distance between pairs.

4.2.4 Kin tolerance and avoidance

In order to establish the likelihood of kin tolerance between females and males and the likelihood of kin avoidance between the sexes, individual’s nearest neighbours were identified (Figure 4.2). The relatedness estimates between nearest neighbour for same sex pairs and opposite sex pairs were compared to that of the mean relatedness of the total population for same sex adults of both sexes, in order to determine whether neighbours were more or less related on average than random pairs from the total population. Nearest neighbours were determined for adult same sex and opposite sex pairs. Nearest neighbours were defined as the adult sampled in the closest nest box to the individual in question (including those found in the same box). Nearest neighbours were determined for each year separately, as 60% of adults perish during hibernation and thus individual home-ranges would change from year to year [31]. Samples from different months of the same year were included together, as during one breeding season, home-ranges remain constant [11]. Mean relatedness was generated along with standard error bars for each pedigree category. Two-tailed t-tests were performed in Excel to determine significance between relatedness estimates categories (e.g. nearest neighbour relatedness vs. population mean relatedness). F-tests were performed in order to determine whether the variance between the two data sets were equal or unequal.
Fig. 4.2 Illustration of possible scenarios where, considering one individual at a time, its nearest same-sex or opposite neighbour is not necessarily its nearest neighbour. Also, the other individual of the pair may itself have a different same or opposite-sex neighbour (single headed arrow).
4.2.5 Population and individual level inbreeding

In order to investigate the relative contributions of inbreeding and population substructure within each site, the R script ConStruct [246] was employed for each site separately, where the algorithm estimates the joint likelihood of the percentage of the population with inbred parents and $F_{ST}$ within an existing data-set, using the method of Overall Nichols [247]. The value of the inbreeding coefficient ($g$) was set to 0.0625, to represent parents related as first cousins. Individual inbreeding was estimated estimating the probability of observing each individual’s genotype using a maximum likelihood method scripted in R.

The sensitivity of each microsatellite marker to inbreeding, within each population, was assessed using a minimum allele frequency method. If we consider that inbreeding is most likely to be detrimental due to the increased degree of homozygosity for rare, possibly deleterious, alleles that it generates [56], and that the allele frequencies at each locus reflect to some extent the situation across the genome, then the magnitude of influence can be given by the ratio:

$$\frac{Pr(homozygosity|F > 0)}{Pr(homozygosity|F = 0)} = \frac{(p^2(1 - F) + pF)/p^2}{(p(1 - F) + F)/p}$$

where $p$ is the frequency of the rarest allele. We then see that as $F$ tends to 1, this ratio tends towards $1/p$. Hence, populations with rare alleles are going to be more susceptible to the effects of inbreeding than populations lacking rare alleles.

4.3 Results

4.3.1 The relatedness coefficient

The accuracy of five relatedness coefficients was tested using individuals from across all three generations, comprising captive dormouse families of known pedigree, including: unrelated individuals ($r = 0$); known grandparent / grandchild relationships ($r = 0.25$); known parent/offspring and known full sib relationships ($r = 0.5$). All estimators showed a high level of uncertainty with large confidence intervals (see Figures 4.3 - 4.6). The estimator ‘Ritland’ was the least accurate across all degrees of relatedness, closely followed by Lynch and Ritland (LynchRd), as both consistently underestimated known values of $r$. Queller and Goodnight (QuellerGt) consistently underestimated $r$ for parent/offspring relatedness (see Figure 4.5). Both Wang and LynchLi showed relative
high accuracy in obtaining $r = 0$ (unrelated) and for all degrees of related individuals. However, Wang’s moment-based estimator appeared superior when approximating the parent/offspring relatedness (see Figure 4.5).
Fig. 4.3. Plot depicting five relatedness coefficients obtained from seven captive dormice pairs known to be unrelated. Error bars represent 95\% confidence intervals determined by bootstrapping over all loci 1000 times.
4.3 Results

Fig. 4.4 Plot depicting five relatedness coefficients obtained from 12 captive dormice pairs known to be related to the degree of $r = 0.25$ (grandparent/grandchild). Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times.
The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*)

Fig. 4.5 Plot depicting five relatedness coefficients obtained from 14 captive dormice pairs with a known parent/offspring relationship and related to the degree of $r = 0.5$. Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times.
Fig. 4.6 Plot depicting five relatedness coefficients obtained from 12 captive bred dormouse pairs known to be full sibs and related to the degree of $r = 0.5$. Error bars represent 95% confidence intervals determined by bootstrapping over all loci 1000 times.
The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*)

Table 4.1 Demographic summary of the two study populations with total sampled for each sex, sex ratio, density of each sex and total density.

<table>
<thead>
<tr>
<th>Site</th>
<th>Year/Wood</th>
<th>No. ha. monitored</th>
<th>Sex</th>
<th>N</th>
<th>Sex ratio</th>
<th>Density</th>
<th>Adults/ha.</th>
<th>Average density</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>MW</td>
<td>9</td>
<td>3:5</td>
<td>1.3</td>
<td>3.3</td>
<td>3.5</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>MW</td>
<td>10</td>
<td>3:5</td>
<td>1.4</td>
<td>3.7</td>
<td>3.7</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>MW</td>
<td>11</td>
<td>4:5</td>
<td>1.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>Big Wood</td>
<td>4.99</td>
<td>F 1</td>
<td>1:3</td>
<td>0.2</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Briddlesford Copse</td>
<td>43.7</td>
<td>F 14</td>
<td>3:5</td>
<td>0.32</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 9</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IoW Dunnage</td>
<td>3.66</td>
<td>F 4</td>
<td>3:4</td>
<td>1.09</td>
<td>1.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 3</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IoW Sandpit</td>
<td>1.74</td>
<td>F 3</td>
<td>1:3</td>
<td>1.72</td>
<td>2.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 1</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IoW Uppersheep</td>
<td>5.17</td>
<td>F 1</td>
<td>1:1</td>
<td>0.19</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 1</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3.2 Summary of demographics

A total of 74 individuals were sampled over three years at MW (2012 - 2015) and 53 individuals were sampled on the IoW over two years (2014 and 2015), however the demographics were only considered for 2014, as the sampling effort was higher (n = 40). Mallydams Wood demographics were estimated for each year, with averages taken. The IoW site was much larger (MW=22ha, IoW = 59.26ha) and made up of multiple woods, thus demographics were estimated for each wood with averages taken. The summary of demographic information can be found in Table 4.1. Population density remained relatively stable over the three years for MW, fluctuating from 3.3 ad/ha to 3.7 ad/ha and averaging across years to 3.5 ad/ha. IoW density ranged from 0.4 to 2.3 ad/ha in different woodlands and averaging at 1.2 ad/ha. Sex ratios differed between sites, with males outnumbering females with an average of 3:3:5 for MW and females outnumbering males 3:5:5 for the IoW; although the difference was not significant.

Table 4.2 shows the number and proportions of dormouse pairs within categories of \(r\) for each site and year. The proportion of pairs in each category of \(r\) remained relatively stable between the years 2012 - 2013 at MW, with 45% and 42% of the pairs being related above 0.06 (equivalent to first cousin) and 30% and 32% of the population being more unrelated on average than expected by chance. The percentage of related pairs in the sampled population dropped to 27% in 2014, with the percentage of those related less than would be expected by chance rising to 44%. The proportions for the IoW were very similar as those found at MW, with 32% of the dormouse pairs being related above the level of first cousin and with 44% of the population being related less than you would expect by chance.
<table>
<thead>
<tr>
<th>Site</th>
<th>Year</th>
<th>Categories of $r$</th>
<th>No. of pairs</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>2012</td>
<td>$&gt;$0.4</td>
<td>11</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2 - 0.4</td>
<td>35</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.11 - 0.2</td>
<td>33</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06 - 0.11</td>
<td>26</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.06 - 0.06</td>
<td>56</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt;$-0.06</td>
<td>70</td>
<td>0.3</td>
</tr>
<tr>
<td>MW</td>
<td>2013</td>
<td>$&gt;$0.4</td>
<td>12</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2 - 0.4</td>
<td>33</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.11 - 0.2</td>
<td>27</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06 - 0.11</td>
<td>34</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.06 - 0.06</td>
<td>66</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt;$-0.06</td>
<td>81</td>
<td>0.32</td>
</tr>
<tr>
<td>MW</td>
<td>2014</td>
<td>$&gt;$0.4</td>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2 - 0.4</td>
<td>22</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.11 - 0.2</td>
<td>27</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06 - 0.11</td>
<td>21</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.06 - 0.06</td>
<td>78</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt;$-0.06</td>
<td>119</td>
<td>0.44</td>
</tr>
<tr>
<td>IoW</td>
<td>2014</td>
<td>$&gt;$0.4</td>
<td>29</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2 - 0.4</td>
<td>79</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.11 - 0.2</td>
<td>98</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06 - 0.11</td>
<td>44</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.06 - 0.06</td>
<td>190</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt;$-0.06</td>
<td>340</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 4.2 Shows the number and proportion of dormouse pairs for each category of $r$. 
4.3.3 Genetic vs. geographic distance

Both male and female same sex comparisons showed significant, negative correlations, where relatedness increased with decreasing geographical distance for Mallydams Wood, when using both Wang’s [237] relatedness coefficients and reconstructed pedigree relationships (see Figures 4.7 - 4.9). However, little of the variance in geographic distance ($R^2$) was explained by the linear models.

Relatedness increased significantly with decreasing geographical distance for IoW male and female same sex comparisons, when using both relatedness coefficients and the reconstructed pedigree (Figure 4.8). Again, little of the variance in geographic distance ($R^2$) was explained by the linear models. Relatedness also increased significantly with decreasing geographical distance for IoW opposite sex comparisons, when using both relatedness coefficients and the reconstructed pedigree (see Figure 4.9). No significant relationship was found for MW opposite sex pairs.
Fig. 4.7 Graphical output of relationship between genetic relatedness (between same-sex pairs) and geographic distance for females a) Wang’s relatedness coefficient; b) recreated pedigree and males c) Wang’s relatedness coefficient; d) recreated pedigree, for Mallydams Wood samples.
Fig. 4.8 Graphical output of relationship between genetic relatedness (between same-sex pairs) and geographic distance for females a) Wang’s relatedness coefficient; b) recreated pedigree and males c) Wang’s relatedness coefficient; d) recreated pedigree, for IoW.
4.3 Results

Fig. 4.9 Relationship between genetic relatedness (between opposite sex pairs) and geographic distance for Mallydams Wood a) relatedness coefficient and b) recreated pedigree and IoW c) relatedness coefficient and d) recreated pedigree.
4.3.4 Kin tolerance

When comparing mean relatedness of same sex, the mean total $r$ for females at both sites was slightly higher than mean total $r$ of males, however this difference was only significant for the IoW (IoW: male mean $r = -0.08 \pm 0.017, p > 0.05$; female mean $r = 0.007 \pm 0.014, p = 0.0001$). Mean $r$ for females nearest same sex pair (FP) at both sites are significantly more related compared to mean $r$ of the total population of females (MW: FP mean $r = 0.17 \pm 0.096, p = 0.041$; IoW: FP mean $r = 0.17 \pm 0.096, p = 0.0076$). Male nearest same sex pairs (MP) at both sites are more related compared to total population of males, but only significantly so for Mallydams (MW: male MP mean $r = 0.128 \pm 0.053, p = 0.0013$). Both female and male nearest pairs are on average significantly closer to each other than two same sex pairs from the total population for both sites (IoW FP: $p = 2.3E-24$; IoW MP: $p = 4.5E-07$)(see Figure 4.10).

![Fig. 4.10 Plot depicting mean r and mean distance with standard error bars of the total population and nearest pair for same sex individuals at both sites with the sample number of pairs in brackets. Females are represented by circles and males by squares. MW individuals are in black and IoW individuals are in blue, with mean total r for each sex represented by solid colour and nearest pair mean r for each sex by the striped pattern. Mean distances (D) for each category are represented by a red cross.](image)

Figure 4.11 shows the opposite sex nearest neighbour mean relatedness and total population mean relatedness for both sites. Mean $r$ for nearest opposite sex pairs (MW OSP mean $r = 0.021$; IoW OSP mean $r = 0.003$) was higher than total population mean $r$ for both sites (MW mean $r = 0.014$; IoW mean $r = -0.023$), but not significantly. Nearest opposite sex pairs were significantly nearer to each other than compared to
4.3 Results

The rest of the population for both sites (MW OSP mean $D = 0.06$; MW mean $D = 0.26$, $p = 0$; IoW OSP mean $D = 0.075$; IoW mean $D = 0.58$, $p = 0$).

![Image](image-url)

**Fig. 4.11** Plot shows mean relatedness of females and males opposite sex pairs (OSP) compared to total adult population mean relatedness. MW individuals are in black and IoW individuals are in blue. Mean $r$ for OSP are represented by circles and mean $r$ for total populations are represented by triangle.

All nearest pair comparisons (same sex and opposite sex) included both related and unrelated pairs, with the proportion of female nearest pair close relatives being higher at both sites, than the proportion of close relative pairs of males; whereas male nearest pairs contained a higher proportion of unrelated ($r < 0$) pairs for both sites. Fewer opposite sex nearest pairs were closely related than same sex pairs drawn from the Mallydams population and no nearest pairs were closely related for samples drawn from the IoW.

Both sites showed the same pattern for mean $r$ for nearest neighbours (NN), with same sex females, being the most related on average, followed by male same sex NN and then opposite sex NN, though all nearest neighbours were on average closer to each other at MW than IoW. For both sites, Female NN were related to each other on average to the equivalent of half sibs (MW mean $r = 0.24 \pm 0.16$; IoW mean $r = 0.23 \pm 0.11$) and were significantly more related on average from the total population. Males were related on average to the equivalent of third degree relatives, such as uncle/nephew relationship (MW $r = 0.125 \pm 0.071$; IoW $r = 0.147 \pm 1.7$) and were significantly more related on average when compared to the mean $r$ of the total population. Opposite sex NN were not significantly more related than two members of the total population drawn at random (see Figure 4.12).
Fig. 4.12 Plot shows mean relatedness for NN pairs compared to total population mean $r$ and all individuals. NN included opposite sex (OS), female same sex (F SS) and male same sex (M SS) pairs. MW is represented by black and IoW by blue, with triangles for OS, circles for F SS, squares for M SS and diamonds for the total population. All means have SE bars.

The proportions of nearest neighbour relatives are shown in table 4.3 for the MW site and table 4.4 for the IoW site. Of the total number of nearest neighbour pairs, 55% were opposite sex, 14% were female pairs, 31% were male pairs and 47% were considered to be relatives ($r > 0.06$). Whereas, the IoW had 54% opposite sex pairs, 27% female pairs and 19% male pairs, with relatives making up 49% of the total NN pairs. Of the percentage that made up the opposite sex NN pairs, 50% and 35% were related above the level of first cousin ($r > 0.06$) for MW and IoW respectively. Close relatives ($r > 0.2$) made up 19% of the NN opposite sex pairs for MW and 15% for the IoW, however none were related as much as full sibs for the IoW. Of the percentage that made up the female NN pairs for MW, 50% were related to the degree of full sibs and 50% were related less than expected. For the IoW, 70% of female pairs were related as $r > 0.06$, 30% being closely related. For MW males 39% of the NN pairs were comprised of relatives, with 22% being closely related as $r > 0.2$. For IoW site, 57% of male NN pairs were related as $r > 0.06$ and 22% were close relatives.
4.3 Results

<table>
<thead>
<tr>
<th>Total no. of pairs</th>
<th>% of total population</th>
<th>Category of $r$</th>
<th>Number of pairs</th>
<th>% of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS NN</td>
<td>33</td>
<td>0.55</td>
<td>&gt;0.2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06 - 0.2</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;0.06</td>
<td>16</td>
</tr>
<tr>
<td>F SS NN</td>
<td>8</td>
<td>0.14</td>
<td>&gt;0.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06 - 0.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;0.06</td>
<td>4</td>
</tr>
<tr>
<td>M SS NN</td>
<td>18</td>
<td>0.31</td>
<td>&gt;0.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06 - 0.2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;0.06</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 4.3 The proportions of nearest neighbour relatives for the Mallydams wood site (MW)

<table>
<thead>
<tr>
<th>Total no. of pairs</th>
<th>% of total population</th>
<th>Category of $r$</th>
<th>Number of pairs</th>
<th>% of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS NN</td>
<td>20</td>
<td>0.54</td>
<td>&gt;0.2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06 - 0.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;0.06</td>
<td>13</td>
</tr>
<tr>
<td>F SS NN</td>
<td>10</td>
<td>0.27</td>
<td>&gt;0.2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06 - 0.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;0.06</td>
<td>3</td>
</tr>
<tr>
<td>M SS NN</td>
<td>7</td>
<td>0.19</td>
<td>&gt;0.2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06 - 0.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;0.06</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.4 The proportions of nearest neighbour relatives for the Isle of Wight site (IoW)

4.3.5 Inbreeding

ConStruct analysis revealed that for the MW population the maximum likelihood values of $F_{ST} = 0$ and $Cg = 0$, indicating no substructure or inbreeding. Whereas for the IoW population $F_{ST} = 0.876$ and $Cg = 0.47$, indicating both substructure and inbreeding, with 47% of the population inbred to the level of parents being related by 0.0625 (the support limit for the maximum likelihood value is $G = 0.00013$, which almost corresponds to the outermost contour line of Figure 4.13 (b)).

Individual inbreeding scores for each sex at each site are shown in Figure 4.14. From this it can be seen that the IoW has more inbred individuals than MW, with 19% (11 of 58) of significantly inbred individuals compared to 1% (1 of 69) for MW. The majority of significantly inbred individuals on the IoW are male (8 of 11). The majority of individuals at both sites are not significantly inbred, with 43% and 58% scoring $F = 0$ for IoW and MW respectively.

The sensitivity of each microsatellite marker to inbreeding, within the MW and IoW populations, was assessed using the minimum allele frequency method, the results of which can be seen in table 4.6. Nine markers were very sensitive to inbreeding for the IoW and seven markers were sensitive for MW, with Mav 015, 030, 038, 044 and
Fig. 4.13 ConStruct results for MW (a) and IoW (b), with $g$ set at 0.0625. Shows the joint likelihood of the causes of excess homozygosity: cryptic substructure ($F_{ST}$) and/or consanguinity, where $Cg$ is the proportion of the sample inbred to degree $g$. The support limit for the maximum likelihood value is given by $G$. 
4.3 Results

Fig. 4.14 Plot shows individual inbreeding scores for both sexes at both sites. Circles represent females (F) and squares represent males (M). Mallydams (MW) individuals are black for females and grey for males, IoW individuals are light blue for females and dark blue for males. Striped individuals are those found to be significantly inbred using the maximum likelihood method. The dashed red line depicts the limit of confidence (i.e., the possibility that $F=0$ is excluded).
The Social Structure of the British Hazel Dormouse (*Muscardinus avellanarius*)

<table>
<thead>
<tr>
<th>Loci</th>
<th>MAF</th>
<th>1/MAF</th>
<th>MAF</th>
<th>1/MAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mav015</td>
<td>0.0086</td>
<td>116*</td>
<td>0.0072</td>
<td>138*</td>
</tr>
<tr>
<td>Mav030</td>
<td>0.0089</td>
<td>112*</td>
<td>0.0072</td>
<td>138*</td>
</tr>
<tr>
<td>Mav038</td>
<td>0.0086</td>
<td>116*</td>
<td>0.0072</td>
<td>138*</td>
</tr>
<tr>
<td>Mav044</td>
<td>0.0088</td>
<td>114*</td>
<td>0.0072</td>
<td>138*</td>
</tr>
<tr>
<td>MavSH3</td>
<td>0.0089</td>
<td>112*</td>
<td>0.0075</td>
<td>134*</td>
</tr>
<tr>
<td>Mav049</td>
<td>0.0086</td>
<td>116*</td>
<td>0.1087</td>
<td>9.2</td>
</tr>
<tr>
<td>Mav034</td>
<td>0.0086</td>
<td>116*</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mav003</td>
<td>0.0098</td>
<td>102*</td>
<td>0.1269</td>
<td>7.9</td>
</tr>
<tr>
<td>MavSG6</td>
<td>0.0086</td>
<td>116*</td>
<td>0.0145</td>
<td>69</td>
</tr>
<tr>
<td>Mav032</td>
<td>0.0345</td>
<td>29</td>
<td>0.0075</td>
<td>134*</td>
</tr>
<tr>
<td>Mav053</td>
<td>0.0862</td>
<td>11.6</td>
<td>0.0078</td>
<td>128*</td>
</tr>
<tr>
<td>Mav051</td>
<td>0.1638</td>
<td>6.1</td>
<td>0.0362</td>
<td>27.6</td>
</tr>
<tr>
<td>Mav017</td>
<td>0.0357</td>
<td>28</td>
<td>0.0896</td>
<td>11.2</td>
</tr>
<tr>
<td>MavSG3</td>
<td>0.0351</td>
<td>28.5</td>
<td>0.0290</td>
<td>34.5</td>
</tr>
<tr>
<td>Mav011</td>
<td>0.0431</td>
<td>23.2</td>
<td>0.0368</td>
<td>27.2</td>
</tr>
<tr>
<td>Mav048</td>
<td>0.0603</td>
<td>16.6</td>
<td>0.0147</td>
<td>68</td>
</tr>
<tr>
<td>MavSF10</td>
<td>0.0172</td>
<td>58</td>
<td>0.0362</td>
<td>27.6</td>
</tr>
</tbody>
</table>

Table 4.5 Table showing the minimum allele frequency for each locus (MAF) and the reciprocal (1/MAF) for each site. The * highlights those loci most sensitive to inbreeding. SH3 showing sensitivity for both sites. Both populations are sensitive to inbreeding with IoW slightly more so, due to having more loci with rare alleles (see Chapter 5).

### 4.4 Discussion

The purpose of this study was to identify any spatial genetic patterns that could clarify the social structuring of the hazel dormouse native to two sites within the British Isles. Due to the endangered status of the dormouse in Great Britain and the genetic consequences of social behaviour, understanding the social structuring of dormice could give a clearer indication of population health and dynamics. The degree to which the hazel dormouse is social is largely unknown, as the evidence available is often conflicting. Same sex territory overlap has been reported to varying degrees [11] and is dependent on population density and the place of study [11]. This overlap does suggest a level of tolerance for their neighbours. Social organisation varies through the active season, with same sex cohabitation reported outside the breeding season, but never during [228][22][222], implying that breeding condition may influence aggression levels...
in dormice. The communal nesting of females in captivity [248] and the occurrence of possible créching in wild populations implies cooperative breeding between relatives [53] and a high level of sociality. However, such occurrences are considered rare in the wild and as such, the dormouse is generally deemed to be solitary [248]. Therefore, there is currently little consensus of the role sociality plays in wild dormouse populations and insufficient data on the role relatedness plays in population structure. The current study employed 17 microsatellite markers to address the question of social structure in wild populations of the hazel dormouse.

Firstly, because the decision for dormice to disperse was hypothesised to be governed by their degree of relatedness, geographic distance was considered a function of genetic distance and was quantified by employing simple linear models between same sex pairs of dormice for each site separately. This was considered in order to establish whether relatedness could reliably predict distance between same sex pairs. Females at both sites showed significant negative correlations between genetic and geographic distances (Figures 4.7 and 4.8), leading to the conclusion that geographic distance between pairs decreased, as levels of relatedness increased. Matrilineal structure is typical of female philopatry [216] and supports the prediction, based upon available evidence for dormouse female philopatry [248]. This result also concurs with the findings of Naim [53], where kinship was found to be higher between females within close proximity. The same significant negative relationship was found between males at both sites (Figures 4.7 and 4.8), which was an unexpected result given that male - biased dispersal is currently believed to operate. This result is also in disagreement with previous findings suggesting that adult males displayed lower levels of relatedness at shorter distances [53]. However, it must be noted that all models explained very little of the variation in geographic distance with $R^2$ values being $< 0.1$. Hence, not all neighbouring same-sex individuals appear to be related, indicating that some are in fact not philopatric and disperse from their natal areas.

Opposite sex models for Mallydams Wood (MW) showed no significant relationship between genetic and geographic distance, whereas a significant negative trend was found for the Isle of Wight (IoW)(Figure 4.9). Both results are surprising, as it was predicted that opposite sex relatives would avoid each other (either directly through kin recognition or indirectly through natural dispersal mechanisms) driven by inbreeding avoidance, predicting that opposite sex relatives would be further away from one another. Though the relationship was not significant for MW, there was a slight negative trend using both Wang’s relatedness coefficient and the recreated pedigree.
comparison, indicating that some opposite sex close relatives are in close proximity to one another.

The results were the same regardless of whether Wang’s relatedness estimates (realised) or recreated pedigree relatedness was used, however the recreated pedigree relatedness allowed for a clearer visualisation of the relationship. Figures 4.7 and 4.8(c & d) illustrate that for the significant models, the relationship between geographic and genetic distance is largely driven by first degree relatives \( r = 0.5 \), as this category invariably showed the smallest mean geographic distance compared to other relatedness categories (Figures 4.7 - 4.9). The degree of variability found for the other relatedness categories \( r < 0.5 \) clearly impacted on the models’ \( R^2 \) values.

Both same/opposite sex pairs and nearest neighbour analysis were employed to further investigate the possibility of kin tolerance between both same sex pairs and opposite sex pairs. It is important to emphasise that the sampling sites used were part of the PTES national dormouse monitoring scheme and as such, the position of the nest boxes conformed to being placed 1.2 - 1.5 meters above ground level, each being 10 - 20m apart in a grid formation, and thus were not designed specifically for this study. However, as adults are expected to maintain permanent home ranges, nearest neighbour analysis using the nest boxes was still considered to provide useful insights into the level of tolerance displayed by relatives.

For nearest same-sex pair comparisons for the MW site, the mean relatedness for the total population of females was higher than males, but the difference was not significant. Both sexes were significantly more related to same-sex nearest pairs on average than same-sex pairs from the total population (female - female \( p = 0.04 \); male - male \( p = 0.001 \), see Figure 4.10), indicating a level of tolerance to related individuals of the same sex in close proximity.

The IoW analysis produced different results for nearest same sex pair comparisons, as the total population of females mean \( r \) was significantly higher than that of the mean \( r \) of the total population of males. This conformed to the expectation of a species displaying female philopatry and male biased dispersal, where the presence of migrant males in a population would decrease the overall relatedness between males. Female and male same sex nearest pairs had higher mean \( r \) compared to the total population, but the result was only significant for females. All average relatedness measures for male - male pairs were negative, indicating that relatedness between males is less than would be expected by chance [244], which would be anticipated for the dispersing sex. The average relatedness for opposite sex nearest pairs did not differ significantly from average relatedness of opposite sex pairs for the total population for either site. The
4.4 Discussion

lack of difference implies that nearest opposite sex pairs were no more or less related to one another than two individuals drawn randomly from the population.

When considering the nearest neighbours (NN) to each individual, the majority of pairs were made up of the opposite sex, with both sites having roughly the same proportions: 55% of the total number of pairs for MW ($n = 58$) and 54% of the total number of pairs at IoW ($n = 37$). Female - female NN pairs made up 14% of the total pairs for MW and 27% of the total pairs for IoW, whereas male - male NN made up 31% and 19% for MW and IoW respectively (Tables 4.4 & 4.5). The differences seen in the percentages of same-sex NN at each site most likely reflects the sites sex ratios (see Table 4.2). Because males outnumber females at MW, a higher percentage of male same-sex NN was predicted and found. Female NN pairs were on average related to the degree of half sibs ($r = 0.24$) whereas the males were related on average to the equivalent of third degree relatives ($r = 0.125$) and both sexes were significantly more related to their same sex NN than two individuals drawn at random from the population for both sites (Figure 4.12). Opposite sex NN did not differ significantly from the total population.

The relatedness value for nearest opposite-sex pairs and opposite-sex nearest neighbours supports the idea that neither males nor females cluster with or disperse from opposite sex relatives more so than expected by chance. This corresponds with the lack of relationship found between relatedness and geographical distances for MW opposite-sex pairs (Figure 4.9(a& b)), however not with the significant, negative relationship found for opposite sex relatives and distance (Figure 4.9(c& d)), where close relatives were more likely to be in close proximity. The IoW sample is made up of two separate subpopulations (see Chapter 5), thus it is possible that the significant relationship between relatedness and distance is an artefact of this, as members from one population will be in closer proximity to each other and appear to be related to one another on the basis of using total population mean allele frequencies for the relatedness estimates.

All categories of nearest neighbours contained both related and unrelated neighbouring pairs. Female neighbouring pairs at both sites had equal percentages of close relative pairs and unrelated pairs, 50% and 30% for MW and IoW respectively. The pattern was broadly the same for male nearest neighbours from the IoW, with 43% being close relatives and 43% non-relatives; whereas MW had 22% close relatives and 61% non-relatives. The slight differences in percentage was most likely due to sample size and sex ratio differences for MW.

For the MW site, opposite sex nearest neighbours were made up of 51.5% relatives ($r > 0.2 = 18.2\%$) whereas the IoW had 35% relatives ($r > 0.2 = 15\%$). Thus, there
is the potential for inbreeding at both sites. When considering the minimum allele frequency at each locus, it is clear that both populations are broadly sensitive to inbreeding, therefore inbreeding avoidance behaviour would be expected. The MW sample showed no evidence of inbreeding (Figure 4.12), however the ConStruct analysis revealed that there was a significant amount of inbreeding within the IoW site, with as much as 47% of the population being inbred to the degree of parents being first cousins. Individual inbreeding scores showed that some of the population was highly inbred, whereas many individuals had an $F$ of zero (Figure 4.13).

The matrilineal relationships found at both sites (Figures 4.6 and 4.7 (c) and (d)) indicate that females show a level of philopatry, with close relatives appearing to cluster together. Nearest female pair and neighbour analysis confirms this, as females had, on average, a higher relatedness estimate to other females in close proximity and their nearest neighbour than the background, average, level of relatedness (approx. zero). Interestingly, the same relationship was found for male dormice at both sites, with close relatives clustering together in closer proximity (6 and 7 (a) and (b)). The small $R^2$ values produced for the linear models and the presence of both related and unrelated nearest neighbour pairs at both sites for both sexes, indicates that neither sex exclusively practises philopatry or dispersal. Both nearest male pairs and nearest neighbour pairs supported the evidence of clustering for MW, whereas only nearest neighbour pairs suggested the clustering of male relatives for IoW. Finding the same pattern amongst males was unexpected, due to the existing evidence suggesting male-biased dispersal and female philopatry in the dormouse [53] [11] [106] [232] in addition to fewer studies reporting males associating with kin [219]. However, male philopatry and spatial proximity have been described in a number of mammal species, [249] [244] [250] [26] [219] and thus, our result would not be the first example of male philopatry. There are two main explanations for the spatial clustering of relatives, one is natal philopatry and the other is settlement in nearby territories [251]. Firstly, natal philopatry has fitness benefits for individuals settling near relatives, either directly through cooperative breeding or indirectly through reduced aggression in territorial defence [251]. Levels of philopatry can also be influenced by population turnover, which creates opportunities for local settlement [252]. In this case, proximate relatives result from the fact that the first available territory happens to be within their natal range [253] [251] [216]. Winter mortality in the hazel dormouse has been found to be as severe as 60% [31]. Given that young born late in the season (e.g. September/October) have been found to wait until after their first hibernation before dispersing [11] [232], the high mortality during the winter months of adults that previously held territory,
may explain the apparent settlement of both male and female dormice in their natal range.

Female dormice at both sites exhibited the same spatial patterns across all analyses. The consistency at both sites indicates that relatives naturally cluster together, rather than it being the random inheritance of newly opened territories. The higher levels of relatedness found between nearest female neighbours could imply the preferential treatment of close kin over more distant kin for the purposes of co-operative breeding and reduced territorial defence. However, the scenario in which close relatives are proximate due to the probability of settling near nest mates, through lack of movement, is equally plausible. The presence of related clusters of females intermixed with unrelated females is notable. If all neighbours were related, the opportunity for cooperative behaviour to evolve would be lower, as all competitors are kin [254]. Local differentiation is important for the operation of kin selection, as it allows for competition and selection to act at the kin group level [254]. The same could be true for the male dormice, as clusters of relatives were found with unrelated individuals. However, nearest male pairs were significantly further away from each other, on average, compared to female nearest pairs (p = 0.049) in MW and the total male population were on average further away from each other than the female total population for the IoW (p = 0.005). Coupling this with female nearest neighbours being more related on average than male nearest neighbours, shows that females are the more philopatric sex and males are more likely to disperse. Sex dependent selective pressures to either disperse from or remain in a natal range is, however, to be expected. Familiarity with the natal area may be more important for the successful attainment of sufficient resource to maximise reproductive output [213] for female dormice.

The presence of male relatives could be explained by a number of scenarios. Limitations of suitable habitat within a reachable distance could account for the presence of closely related males. Related males from one population may disperse to the same woodland purely due to the lack of any other suitable habitat near their natal range. Mallydams Wood has few, large surrounding woodlands and what is available has limited connectivity (see Chapter 5). It could also be that male relatives disperse together and establish territories near each other and thus are not, in fact, within their natal territory. Although this was not explicitly tested here, this scenario is considered unlikely due to the fact that there were opposite sex relatives found in close proximity in both populations, making the likelihood that some males remain within their natal site high. There would be the priviso, however, that the opposite sex relatives could be cross-generational.
Although there were many similarities between the two sites, with the relatedness structure of each population being very similar (Tables 4.4 and 4.5), there were some marked differences. Firstly, the MW population was significantly more related on average than the IoW population (t-test, p = 0.005). The IoW sample had a significantly higher mean $r$ for the total population of females, compared to the total population of males, whereas the averages for the MW sample were positive, but much the same across sexes (Figure 4.11), which could explain the higher levels of relatedness seen in the total population of MW. Mean $r$ between female pairs was not significantly different between sites, but the MW male pairs were significantly more related than IoW male pairs ($p = 0.0025$), with IoW males being on average unrelated (Figure 4.9). This result indicates that males within the MW site are significantly more philopatric than males in the IoW site. Nearest male pairs were significantly more related to each other compared to the total population of males for MW (Figure 4.9), however, this was not true of nearest male pairs on the IoW, corroborating the theory that MW males are more philopatric than those found within the IoW population. The fact that males are more philopatric in MW than the IoW, suggests that there are differing conditions. For example, dispersal and philopatry are influenced by ecological constraints such as habitat saturation, environmental harshness, unpredictable conditions and the likelihood that dispersal will lead to successful reproduction in a new territory [255][107].

Species inhabiting fragmented habitats may be reluctant to migrate across inhospitable matrix environments, which alters successful movement between isolated patches [27]. Thus, landscape alterations separating once continuous populations can result in restricted gene flow through reduced dispersal [24]. Those species at the greatest risk from habitat fragmentation are habitat specialists with low dispersal ability [256], like the hazel dormouse. Mallydams Wood is a small (22 ha) wood with poor connectivity (Chapter 5) to other woodlands and, as such, it is possible that the natural movements of the hazel dormouse have been disrupted, resulting in males remaining in their natal habitat. By comparison, the Isle of Wight site is a medium sized woodland with good levels of connectivity to other medium and large sized habitats, which may explain the differences seen. Individuals are more likely to remain within their natal area when ecological constraints impose higher costs to dispersal, [257]; [248], such as those imposed by habitat fragmentation, and when food and space become limited at higher population densities [107]. Thus, habitat fragmentation could have the double effect of limiting dispersal and increasing population densities, which further limits dispersal. In Britain, average population densities of hazel dormice tend
to be between 1.75 and 2.5 ad/ha based on 83 National Dormouse Monitoring sites of varying habitats [11]. Mallydams Wood had a density (averaged across years) of 3.5 adults/ha, which is higher than average, and unexpected considering that Mallydams Wood is not an optimal habitat for dormice. The Isle of Wight site is managed specifically for dormice, but has a slightly lower than average density of 1.2 adults/ha (averaged across woodlands). The difference in population densities is possibly due to the connectivity and surrounding matrix of each site and offer a possible explanation for differences seen in male retention. Walker et al [26] found that higher densities of southern hairy-nosed wombats (Lasiorhinus latifrons) caused by habitat fragmentation, led to decreased female dispersal and an increased association with kin. In prairie voles (Microtus ochrogaster), both males and females became more philopatric under conditions of habitat saturation and higher population densities [258]. Thus, as population densities increase, habitats may become saturated, leading to territories becoming limited with the effect that dispersal decreases and group size increases [107]. A similar scenario could be playing out for the dormice in Mallydams Wood and may explain the significantly higher relatedness of males dormice within MW compared to the IoW. The presence of opposite sex relatives would be expected if neither sex exclusively practiced sex-biased dispersal, as is the case for both sites. However, such a situation increases the likelihood of inbreeding and as both populations appear to be susceptible to inbreeding (Tables 4.4 and 4.5), further inbreeding avoidance mechanisms would be expected to evolve. The Mallydams sample had a higher proportion of opposite sex nearest neighbour relatives, compared to the Isle of Wight and, thus, the potential for inbreeding could be higher, which is of concern for an endangered species. The disruption of gene flow through a decrease in dispersal can lead to increased relatedness within a population, hence the increased risk of inbreeding and heightened rates of genetic erosion, potentially lowering genetic variation [71]. However, significant levels of inbreeding were found for the IoW and not for the MW site. The absence of any inbreeding within the MW site, even with close relatives making up 18.2% of the opposite sex nearest neighbours, implies that inbreeding avoidance mechanisms are in operation. Polyandry has been proposed to reduce inbreeding levels through the overall reduction in relatedness between litter mates and the absence of a paternal skew [113][136][137][138]. Thus, if litter mates breed, the offspring will be less inbred, with lower levels of homozygosity, due to the lower levels of relatedness. Polyandry has been observed in the hazel dormouse [50] [53]. However, the majority of the population at MW had an inbreeding coefficient of $F \approx 0$ and if polyandry was the only mechanism, individuals would still be expected to show inbreeding. Another possible mechanism
could be inbreeding avoidance through bacterial driven scent recognition, used to
differentiate between kin and non-kin potential mates. Anal bacterial communities of
the hazel dormouse have been found to be more similar in kin compared to non-kin
(Chapter 3), thus it is possible that odour production through bacterial action may
allow mate selection based upon relatedness. The significant inbreeding levels within
the IoW population in comparison to no inbreeding within MW was surprising, as the
overall relatedness for the IoW population was significantly lower than for MW (Figure
4.11) and levels of male philopatry were higher for MW. The difference between the
two sites may suggest that an external force, which has been in place for some time, is
disrupting gene flow and inbreeding avoidance mechanisms. The two subpopulations
within the sampled IoW site were found to be separated by an unidentified, permeable
barrier to gene flow (Chapter 5), which could explain the presence of both non-inbred
and highly inbred individuals. It should be noted, however, that although the levels
of inbreeding were low and not significant for MW, the population is susceptible to
inbreeding and, thus, there is the potential for future inbreeding if males continue to
display increased philopatry.

A possible major drawback to this study was the use of the nest box scheme to
acquire samples. Nest boxes are fixed at 1 meter above the ground [7] and as such, they
do not provide a random sample of the entire habitat that dormice use. Collecting data
in this way is a compromise between obtaining scientifically valuable information and
the ease of collection for volunteers. Adopting this standard approach likely limits the
proportion of the population sampled, especially sites such as the IoW, which provide
optimum conditions for dormice, which likely means much of the population chooses
to nest in natural tree hollows. The low sample numbers may reflect this, rather than
low densities. Low sample numbers sampled during the beginning of the active season,
(April/May) prevented exploration of how sociality changes during the active season
and the possible mechanism behind it. During the start of the active season, dormice
recently out from hibernation often return to their hibernation nests during the day,
rather than build a new nest, which could explain the low sample numbers. If a large
proportion of the population is not being sampled, it is possible that the individuals
in the closest proximity to one another were not sampled, which could mean that
the spatial clustering of relatives would need revisiting. More traditional methods of
small nocturnal mammal trapping may provide a larger sample size and a more robust
method of sampling, in order to fully understand neighbour tolerances and whether
social structure does change throughout the active season. However, such methods are
not widely used for dormice, as they often elude traditional mammal traps and due to welfare concerns [23].

4.5 Conclusions

This study provides the first investigation into the spatial genetic structuring of the hazel dormouse. A general pattern of female philopatry and male biased dispersal was uncovered, though this pattern was not absolute. Both sexes were found to spatially cluster with relatives, but to a higher degree between females, indicating a level of tolerance to related individuals. The presence of both related and unrelated same sex individuals in close proximity provides the opportunity for kin directed behaviours to evolve, such as cooperative breeding. However, the level of kin tolerance and sociality could not be confirmed, as such clustering patterns have been found in both social and solitary animals alike. The probability that non-dispersing individuals remain near to relatives through the lack of movement, rather than an evolutionary benefit, is an equally plausible explanation. Further research combining the spatial structuring of relatives with the degree to which individuals’ home ranges overlap could clarify the degree of tolerance between kin and non-kin. Males found in MW site were more philopatric than those found in the IoW site, as indicated by the higher levels of relatedness for MW males and was likely due to the site being more isolated than the IoW site, providing fewer opportunities for dispersal, with an increased cost of dispersal attached. The result being, that nearest male pairs in MW were significantly more related than the IoW nearest male pairs, which could imply a flexibility in social structuring, whereby individuals will tolerate kin when environmental conditions prevent dispersal. However, the lack of site replicates limits the conclusions that can be drawn on the role differing habitat conditions play in dormouse social structuring. Opposite sex individuals do not appear to cluster or to actively avoid one another, as both related and unrelated pairs were in close proximity. Such situations would be expected to increase the chance of inbreeding; however, the lack of inbreeding within the MW site indicates that inbreeding avoidance mechanisms other than sex-biased dispersal are in operation. Though inbreeding was seen within the IoW site, this was ascribed to the presence of an unknown, permeable barrier, disrupting natural movements and gene flow, resulting in the levels of inbreeding detected. If inbreeding avoidance mechanisms are in operation, it may explain why genetic diversity levels are not low enough to cause immediate conservation concern (Chapter 5).
Chapter 5
Genetic Variation and Population Structure of the British Hazel Dormouse (*Muscardinus avellanarius*)

Abstract
This chapter approaches the social structure of the Hazel Dormouse by viewing it from a higher hierarchical level, which is that of the dispersal between populations. Dispersal is an important determinant of the population dynamics of any species, where patterns of gene flow influence population genetic structure. Dispersal is also essential for maintaining the genetic viability of species. Dispersal can be measured at various spatial scales. Local dispersal; for example, can be inhibited by barriers such as geographic distance (as individuals are more likely to randomly mate with individuals in close proximity), in addition to physical barriers, such as a reservoir or railway line. If these barriers are sufficient to block dispersal, and hence gene flow, local genetic structuring is expected, as the genetic diversity of the isolated breeding groups would be subjected to genetic drift. At the landscape scale, dispersal is considerably influenced by environmental characteristics. Habitat fragmentation can disrupt species’ natural movements. In animals with naturally low dispersal ability, like the hazel dormouse, it is likely that fragmentation further decreases dispersal and, as such, may prevent gene flow. Male dormice have been found to move further than females, thus indicating the possibility of sex-biased dispersal, thus fragmentation may disrupt male-biased dispersal, further increasing the risk of inbreeding. It is yet unknown what
Genetic Variation and Population Structure of the British Hazel Dormouse

(constitutes as a barrier to dispersal for the hazel dormouse, but previous research has found them to be reluctant to cross open ground. Therefore, both local and landscape levels of differentiation may be expected. The chosen sample sites include potential barriers to local and landscape level gene flow: local in the form of water reservoirs (Wakehurst place), railway lines and geographic distance (Isle of Wight). Landscape level in the form of farmland (Isle of Wight and Mallydams), coastline (Mallydams) and the sea (Isle of Wight). The results of this chapter indicate a lack of genetic structuring in habitats with potential barriers to dispersal and as such, imply that the hazel dormouse can navigate a number of different physical structures that may be assumed to pose some difficulty to a small, arboreal mammal. However, these results are in line with recent studies indicating that dormice are more mobile, travel greater distances and traverse both natural and man-made structures, such as roads, that were previously deemed impassable, much more frequently than expected. Genetic structuring was detected within the Isle of Wight populations, but no firm conclusions could be drawn on the cause of the genetic structuring. In light of both general findings, it would appear that the hazel dormouse is less affected by anthropogenic barriers than previously thought.

5.1 Introduction

The evolutionary potential of a species is believed to be a function of the level of genetic variation contained within a population [25][24]. Populations with lower levels of variation are less likely to have the genetic potential to respond to natural and anthropogenic change through the process of evolutionary adaptation [14][24]. Thus, the maintenance of genetic variation is a general conservation concern for endangered species [259].

5.1.1 Measuring the structuring of genetic variation

Natural populations are structured into separate mating units and as such, most of the genetic variation can be partitioned at two levels: within population genetic variation and between population genetic diversity [14]. The mating and dispersal dynamics of individuals, when known, can lead to predictable patterns of genetic variation. For example, social behaviour, such as cooperation, and low dispersal ability leads to increased relatedness within groups [71][215] and barriers to dispersal reduces gene flow, resulting in non-random mating, increasing the likelihood of inbreeding.
that manifests as excess homozygosity at the individual level ($F_{IS}$) or between subpopulations ($F_{ST}$) [32][14]. Therefore, measuring genetic variation, as well as behavioural traits, within and between populations can offer insights into the behaviour of a species, such as the hazel dormouse, that is difficult to observe in the wild. Such an approach can also identify populations that are most under threat from changing environmental conditions, such as habitat loss and habitat modification [25].

Traditionally, the monitoring of genetic diversity in natural populations under threat from anthropogenic environmental change, employs the use of neutral markers. Neutral markers are so named as, being inter-genic, they are not influenced by natural selection and are more sensitive to the forces that erode diversity in small, isolated populations, such as genetic drift. Both the number of alleles and the level of heterozygosity at such markers are used as a measure of the evolutionary potential of a species, albeit with an upwards bias in favour of high polymorphism due to the increased information gleaned from markers found to be polymorphic [260].

Alleles, particularly rare ones (frequency <0.1), are lost rapidly from populations that have experienced a pronounced reduction in size, for example the African cheetah (*Acinonyx jubatus*)[261], which can lead to anthropogenically induced population bottlenecks. Because rare alleles are more likely to be lost from a population of reduced size, heterozygosity is not initially affected. This is because rare alleles have negligible effect on heterozygosity levels and, therefore, there is a transient excess of heterozygosity relative to that predicted by the number of alleles soon after a bottleneck ensues. The relationship between number of alleles and levels of heterozygosity can therefore offer clues as to the past occurrence of a population bottleneck [262]. Higher levels of heterozygosity are theoretically linked to greater population fitness [56] and this relationship has been found in some natural populations, such as zebra finches (*Taeniopygia guttata*) [65], however evidence is rare and conflicting [62]. Neutral markers, particularly microsatellites, have been criticised for not accurately reflecting genome-wide variation, in part due to the small numbers used and thus, do not truly reflect adaptive potential [41]. However, as microsatellites are still the most cost effective way of monitoring passive forms of evolutionary change, which are most influential in endangered species, they remain a valuable resource in population genetics.

The most established measures of genetic variation are Wright’s $F$ - statistics [32]. $F$ -statistics are a mathematical framework, originally based upon the correlation of alleles within and between individuals, describing the distribution of genetic variation within individuals and between individuals and populations using a series of inbreeding coefficients: $F_{IS}$, $F_{ST}$ and $F_{IT}$. $F_{IS}$ corresponds to the increased correlation between
alleles within individuals (the 'I' of the subscript) relative to that of the subpopulation (the 'S' of the subscript). Typically, it is estimated as a departure of genotype proportions from Hardy-Weinberg expectations (see Methods) that can result due to inbreeding at the level of the individual. $F_{ST}$ similarly corresponds to the increased correlation between alleles, but here it is within subpopulations ('S') relative to the total population ('T'), but is typically measured as allele frequency divergence among populations. Finally, $F_{IT}$ is the overall departure from HW of the entire base population due to non-random mating within subpopulations ($F_{IS}$) and divergence among subpopulations ($F_{ST}$) [14] and hence the F-statistics are related to each other as $(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$.

5.1.2 The genetic impacts of habitat fragmentation

Large, continuous habitats have high resource availability and, thus, can support more individuals in a population than small, fragmented populations with low resource availability; the theoretical model of which was developed in the 1960s by MacArthur and Wilson and termed the Island Biogeography [263]. In general, and in the context of semi-isolated subpopulations within larger populations rather than islands, large meta-populations house more genetic diversity, both when measuring the number of alleles and the amount of heterozygosity. Diversity is generated through mutation and more individuals represent more opportunities for beneficial mutations to enter the population, increasing the overall number of alleles. The loss of heterozygosity through genetic drift is inversely proportional to population number, which is indicated by the equation [38]:

$$\Delta H = \left(1 - \frac{1}{2N}\right)H$$

(5.1)

In addition to anthropogenic influence, biotic (e.g. presence of food supply or hosts, vegetation, predators, etc.) and abiotic (e.g. climate, geology) preconditions naturally structure populations; those populations existing over long periods of time in naturally fragmented habitats often show low genetic diversity within and high differentiation among populations [264]. This is due to the effects of genetic drift acting upon each subpopulation separately, whereby an allele initially present in all demes can tend toward fixation in one and loss in another deme, leading to differentiation [14]. The same process is in operation in naturally and anthropogenically fragmented habitats as both create barriers to gene flow and hence generate independent evolution between them. However, species persisting in naturally fragmented habitats are expected to
have existed at low population numbers for many generations and have a reduced genetic load (presence of detrimental mutations) through the purging of deleterious mutations over many generations and, thus, further fragmentation and loss of habitat are unlikely to severely impact the population. In addition to this, naturally fragmented habitats are more likely to be successfully colonised by species with higher mobility [265] and, thus, are able to move through an inhospitable habitat matrix. In contrast, species that have evolved in naturally interconnected habitats will not have these prerequisites for surviving in modified, newly fragmented habitats and are likely to be more affected by anthropogenic changes to their environments.

British woodland habitats are more frequently being subdivided by human modifications, such as urban development, roads and agricultural land, leading to increasingly fragmented populations; as has been observed in Denmark with regard to the bank vole (Clethrionomys glareolus)[71] and small mammal populations in the USA [266]. Loss and fragmentation of habitats leads to the overall reduction of available habitat and subsequently smaller patches, with a lower carrying capacity and, hence, small population numbers, where specialist species are particularly affected [71][24][267][268]. Anthropogenic habitat loss has been associated with rapid reductions in genetic variation [269]. For example, the Australian mountain pygmy possum (Burramys parvus) has experienced habitat loss through natural forest fire, but only the human modification of ski resort development resulted in reduced heterozygosity of the population [269]. Patch size of remnant habitats has also been found to be influential in the levels of genetic diversity in populations, with genetic variation decreasing with the size of the habitat area [14]. It is not only the size of the remaining patches, but also the connectivity between patches that can have an effect on levels of genetic diversity. The increasing isolation of remaining patches and the presence of inhospitable environments between patches, alters the successful movement of species [27] and restricts gene flow through reduced dispersal [24]; in short, a fine-scale illustration of MacArthur and Wilson’s Island Biogeography theory [263].

Gene flow reduces genetic differentiation among subpopulations and increases the genetic variation within populations, thus gene flow is the main driver that holds geographically separated populations into a single evolutionary unit [14]. Consequently, gene flow homogenises subpopulations, whereas drift acts to genetically differentiate them, therefore, the actual amount of divergence between subpopulations is a balance between gene flow and drift. The simplest model explaining this relationship is the Island model of migration [32], where it is assumed that each individual from each generation has an equal probability of breeding in a deme other than that of their
Genetic Variation and Population Structure of the British Hazel Dormouse

\( \text{(Muscardinus avellanarius)} \)

birth. It is also assumed that an individual is equally likely to immigrate into any of the other demes. Under these assumptions, population number becomes negligible, as drift is slower in larger populations, thus fewer migrants are needed to counteract drift, whereas small subpopulations diverge rapidly through drift, thus proportionally more migrants are needed to counteract drift and thus, \( F_{ST} \) can be approximated using the equation:

\[
F_{ST} \approx \frac{1}{4Nm + 1} \quad (5.2)
\]

Where \( N \) is the effective population size and \( m \) is the migration rate per generation; although see [270]. However, in natural populations dispersal between geographically close populations usually occurs more frequently, which violates the assumptions of the Island model. The stepping stone model of migration takes into account both short and long distance migrations [271]. As a result with this model, there will be greater differentiation among subpopulations that a further apart from one another, explaining the patterns of isolation-by-distance often found in natural populations [14].

The disruption of gene flow through a decrease in dispersal can lead to increased relatedness within a population [67], hence increased risk of inbreeding and heightened rates of genetic erosion, potentially lowering genetic variation, become likely outcomes [71]. Both geographic distance and landscape features, such as roads, railways and bodies of water, as well as habitat type, have been reported to reduce dispersal ability, causing the eventual divergence of populations and the genetic loss associated with reduced gene flow and population numbers [272][273][274].

Forests are one of the most heavily disrupted ecosystems, suffering loss and fragmentation through timber exploitation and clear cutting for agriculture [275]. As a result, arboreal mammals, such as the hazel dormouse, can become increasingly subdivided into small, isolated populations at greatest risk from genetic erosion [256][276]. Hence, forest habitat specialists are deemed particularly vulnerable to fragmentation and the compounding effects of inbreeding, which may lead to reduced individual fitness and adaptability [268].

5.1.3 The threat of habitat loss and fragmentation to the hazel dormouse

The effects of habitat fragmentation depend upon the focal species, as it will vary with the behaviour, distribution and habitat requirements of said species [71]. Species most at risk of extinction through the fragmentation effects of decreasing fragment area,
increasing isolation and increasing matrix contrast are inversely proportional to their size and they tend to be habitat specialists with low dispersal ability [256]. Populations of arboreal mammals with low dispersal ability (e.g. the common ringtail possum (*Pseudocheirus peregrinus*), Australian squirrel glider (*Petaurus norfolcensis*) and the fat/edible dormouse (*Glis glis*)) have been found to house less genetic diversity when inhabiting small, isolated patches compared to larger patches [277][276][268]. However, high levels of gene flow have been reported in another small bodied arboreal mammal, the red squirrel (*Sciurus vulgaris*), across strongly fragmented landscapes, due to their ability to cross open areas and longer distance dispersal [278]; highlighting the fact that the natural behaviour of a species influences the overall effects of fragmentation on genetic diversity.

The hazel dormouse is a deciduous woodland habitat specialist, occurring in low densities. When comparing them to other rodent species, the average population densities of the hazel dormouse vary between 1-10 individuals per hectare in different parts of their range and differ with habitat type [10][11]. A maximum of 15.6 adults per hectare have been recorded in Britain [279] and 18.7 adults per hectare recorded in Northern Germany [280]. Other rodents, such as woodmice (*Apodemus sylvaticus*), yellow-necked mice (*Apodemus flavicollis*) and bank voles (*Myodes glareolus*) have fluctuating population densities, where in peak years densities may exceed 100 individuals per hectare and can fall to one individual per hectare in crash years [281]. Thus, the highest dormouse densities reported are only 10 - 19% of peak densities of other common European rodent species.

The hazel dormouse also has comparatively lower reproductive potential than other small mammals in the same habitat, as in most cases only one litter is produced per year, with between 1 - 7 young, with occasional second broods if conditions are optimal [21]. Young become independent between 6 - 8 weeks [7], which is partly why so few litters can be produced. By comparison woodmice, which prefer similar habitats, can produce between 4-6 litters per year of up to 9 young, becoming fully independent in 2-3 weeks [22].

Adult hazel dormice are relatively sedentary, with fixed home ranges throughout the breeding season (females ≈ 1ha; males much larger and typically incorporating several female ranges) and generally travel less than 500m during this season [11]. Natal dispersal has been observed both through direct and genetic means [10][75]. Distance travelled varies with the time of birth, with those born earlier in the year travelling greater distances [15]. Dispersing male juveniles generally travel further than female juveniles, however the distance travelled is generally under 2km, with occasional longer
distance movements, with the longest distance recorded to date being 7km [11], but generally only short distances are covered.

These behavioural traits make the hazel dormouse particularly vulnerable to habitat loss and fragment isolation [23][282][117] and as such, the species provides a valuable opportunity to investigate the impact of variable habitats on their social and population structure.

A number of studies have highlighted the importance of patch size on dormouse presence [283][284][285]. The carrying capacity for a species can be defined as the maximum population size that a particular habitat can support indefinitely given the resource (food, water, space etc.) availability of the environment [286]. It is possible that such limitations are the reason that the size of the patch influences whether dormice occupy certain woodland fragments. Mortelliti et al. [117] found that although patch quality was the most important factor when considering individual survival (regardless of patch size), patch size was the most significant factor for population persistence. The authors concluded that limitations of space in smaller habitats enforced a low absolute population number, which habitat quality could not offset, thus contributing to the higher extinction risk of dormouse populations in small fragments. It has also been demonstrated that populations of hazel dormice found in patchy habitats have lower genetic diversity compared with those in a continuous habitat [75]. However, the levels of genetic diversity were not significantly different between habitats and no bottlenecks were detected in any site studied, which suggested enough gene flow to maintain diversity. For gene flow to be maintained, there must be some level of dispersal. The same study found overall levels of dispersal were limited to fine geographic scale (<1km) for the dormouse, however rare inter-patch movement was detected (>1km).

What constitutes a barrier to dormouse dispersal is not clear. Field experiments by Bright [287] indicated that even small gaps in hedgerows (i.e. 6m) constitute a barrier to dormouse movement. There is much evidence to suggest that the presence of hedgerows increases the likelihood of dormouse occupancy [288][282][289] as they are essential dispersal routes for dormice [12] due to their reluctance to leave the canopy and cross open ground [7]. However, dormice do inhabit unconnected patches, completely surrounded by an agricultural matrix [116][284], suggesting that such crossings may not be uncommon. Various studies have demonstrated their ability to navigate open fields, both grass and mown [15][116][117]. From these studies it appears that dormice are able to cross open ground, but will circumvent fields when hedgerows are present, most probably to reduce risk of predation and, thus, fields are permeable to dormouse
movements in the absence of hedgerows, but hedgerows may increase the likelihood of dispersal when present.

Past research into dormouse movements assumed that even small forest pathways function as a barrier to movement [7] and therefore it is generally accepted that all roads, and by extension railways, would serve as complete barriers to movement [280]. However, the hazel dormouse has been reported crossing from the central reservations of a dual-carriageways to the side in South West England [118]. A study by Kelm et al. [280] found that dormice could cross both minor roads and motorways, with one individual crossing multiple times. Density - dependant dispersal is a widely accepted phenomenon in mammals [290] and this particular dormouse study focussed on populations with extremely high densities (e.g., 18.7 adults/ha) and hence may be found to be the exception rather than the rule. It is also important to note that increased dispersal can be triggered by fragmentation, due to resource limitation, for example, resulting in increased risk taking behaviour in the form of longer distance movement in order to find suitable habitat [291][292]. However, Mills [50] found that dormice on the edge of their range (Cornwall) have reduced genetic diversity and lower gene flow between subpopulations, compared with subpopulations within a dormouse stronghold area (Devon); indicating that perhaps less than optimal conditions (e.g., geographic barriers to suitable habitat) do not trigger increased dispersal generally (e.g., between subpopulations) in dormice. It is therefore important to assess the role of fragmentation and barriers to gene flow in patchy populations of hazel dormice found within a known species stronghold, like Sussex.

The sample sites used in the current study varied in size, levels of isolation and connectivity to other woodland patches and within site structure (see site descriptions in section 5.2.2). Inferences can be made from patterns of genetic variation on the effects that such habitat characteristics impose on levels of dispersal and movement.

5.1.4 Aims and predictions

The main aims of this study are to 1) assess levels of genetic diversity at multiple scales (local, patch level and regional level) in order to identify populations that show signs of reduced population viability (a proxy being increased levels of inbreeding) and 2) use the observed patterns of genetic variation to detect the within habitat characteristics (e.g., barriers to gene flow) that may interrupt dormouse movements and pose a threat to the species persistence.

A number of predictions have been made in light of these aims:
1. Mainland populations and those from the IoW will be significantly differentiated. The IoW is expected to be genetically depauparate compared to mainland populations, as it is isolated and will have a relatively small overall population size, due to the size of the island compared to the mainland.

2. Each sampling site within Sussex will represent a separate subpopulation with significant levels of genetic differentiation expected between all sites due to geographic distance.

3. A significant positive relationship between patch size, levels of connectivity and levels of genetic variation are predicted. Thus, larger, more continuous habitats with good levels of connectivity to other woodland patches would be expected to show the highest diversity and lower levels of inbreeding, with the Isle of Wight showing the most inbreeding and lowest levels of variation, overall.

4. Barriers within sites, such as reservoirs, railways and open fields are expected to cause population substructuring and increased levels of inbreeding within these subpopulations.

5.2 Methods and materials

5.2.1 Study populations

Initially, nine sample sites were identified for study, however only five sites generated sample sizes considered sufficient for further analysis; four from Sussex with differing patch size, varying degrees of fragmentation (from continuous to patchy) and with varying potential barriers to gene flow and levels of connectivity. A site on the Isle of Wight (IoW) is also included, as a long isolated population, providing a comparison for the levels of diversity by being a population considered completely isolated from the mainland by sea, quite unlike populations of the mainland, surrounded by more permeable barriers to gene flow, such as agricultural land. As each population is separated by a minimum of 14.5km (between Tilgate and Wakehurst), each population will be treated as an independent site with regards to regular gene flow and, as such, the IoW can be considered without special reference to Island processes.

5.2.2 Sampling sites

All five study sites are considered in this chapter (see chapter 2 for a general site description and selection criteria), see Figure 5.1.
Fig. 5.1 Map of the positions of all four sites in Sussex (BW: Binsted Wood; TP:Tilgate Park; WH: Wakehurst; MD: Mallydams) and the one on the Isle of Wight (IoW: Briddlesford).
5.2.3 Sampling procedure

DNA extraction and amplification and microsatellite genotyping are described in the general methods, Chapter 2.

5.2.4 Sample selection

Samples that failed to amplify at more than three of the seventeen loci (failure $\geq 15\%$) were removed, in order to improve robustness by minimising missing data, as poorly amplifying samples are more likely to present null alleles and allelic drop-out. Individuals were not marked throughout the study, thus repeat samples from the same individual were possible. Cervus v.3.0.7 [129] was used to calculate identity (Identity Analysis) in order to identify and allow the removal of repeated individuals (individuals that matched at all loci).

Pairwise $F_{ST}$ between sites was calculated using FSTAT v. 2.9.3.2. [293] to quantify the magnitude of isolation between each sampled woodland. As allele frequencies of natural populations can vary over generations [294], samples were taken from the same sites over a number of years. Samples were collected from Mallydams and the Isle of Wight over four years (2012 - 2015); samples were collected from the Wakehurst Place in 2011, 2012 and 2014, from Binsted in 2012, 2013 and 2014 and from Tilgate in 2012 and 2013. For each year, each sample would contain at least two overlapping generations, potentially more. A significant $F_{ST}$ between two successive generations was taken to indicate meaningful genetic differentiation (allele frequency change) and such samples were considered as separate breeding populations.

5.2.5 Clustering analysis

Sample sites that were found to be genetically isolated from each other (significant $F_{ST}$) are still able to comprise cryptic substructure where subpopulations exist due to unknown barriers to gene flow. To investigate this type of genetic structure, Bayesian model-based algorithms were employed using STRUCTURE 2.3.4 [295]. STRUCTURE analysis of all sites was initially carried out with no a priori information on sample location. An admixture model with correlated allele frequencies was used with varying numbers of possible clusters (k=1-10) for 10 runs for each value and 1000,000 MCMC iterations following a burn-in of 100,000. STRUCTURE HARVESTER v0.6.94 [296] was used to generate plots of K against a) estimated log-probability of data value and b) delta K [297]. Within-site STRUCTURE analysis was performed with an a priori population number for the IoW, based on spatial information of where individuals were
sampled, and Mallydams and Wakehurst Place based on the year the samples were taken using the same specifications as above (admixture model, k=1-10 for 10 runs for each value and 1000,000 MCMC iterations following a burn-in of 100,000). The analysis was not performed on Tilgate Park or Binsted wood, due to the low sample number (n<18). The results were visualised using STRUCTURE PLOT [298].

5.2.6 Genetic structure

Weir & Cockerham’s $F_{ST}$ [299] was used as a measure of genetic divergence, along with 95% confidence intervals obtained by bootstrapping across all loci (1000 permutations) and tested for statistical significance using FSTAT (ver. 2.9.3.2) by permuting the genotypes among samples, which does not assume sampled populations conform to Hardy-Weinberg proportions [300]. A sequential Bonferroni correction was applied to each probability estimate.

5.2.7 Isolation - by - distance measures

Isolation - by - distance was calculated by correlating Euclidean distance measures and $F_{ST}$ scores between sites. Distance was measured using the straight line measuring tool in QGIS and was taken from and to the centre of the sampling points of each site. Isolation - by - distance was first calculated for all sites, then for those sites in Sussex and then for sites within West Sussex. A Mantel test [301] was performed in the programme R [203] using the ADE4 package [302], which produced simulated P-values based on 9999 replicates.

5.2.8 Detection of first generation immigrants

The number of first generation immigrants into each sample site was estimated using frequency - based assignment tests in Geneclass2 v.2 [303]. The frequency method of Paetkau et al. [304] was employed to calculate the likelihood of an individual belonging to the population where it was sampled. This method is deemed appropriate when all potential source populations have not been sampled [305] and, given the relatively large geographic distance between sampling sites, it was considered to be the most suitable approach. The probability of an individual being resident to a particular site was then assessed using a resampling procedure [305] and individuals with a probability of less than 0.05 were excluded as a resident.
The program POWSIM (v.4.1) [306] was used to evaluate the statistical power of the microsatellite markers to detect genetic differentiation at various levels of $F_{ST}$. This analysis simulates sampling from a specified number of populations that have reached pre-defined levels of divergence and estimates the probability of false negatives for population differentiation at the expected degree of divergence. Population drift was simulated to $F_{ST} = 0.001, 0.0025, 0.005, 0.01, 0.02, 0.025$ and $0.05$ by using an effective population size ($N_e$) of 2,000 and varying the number of generations accordingly for 1000 replications. The statistical power was ascertained for 17 microsatellite loci in order to detect differentiation among five populations based on sample sizes ranging from 4 - 205. At each locus the $H_0$ of identical allele frequencies was tested using Pearson's contingency chi-squared. An estimate of type I error was obtained by setting $N_e$ to 0 ($F_{ST} = 0$). Power is expressed as the proportion of significant outcomes (1,000 replicates, rejecting the null hypothesis ($H_0$) of no allele frequency difference, or $F_{ST} = 0$ at $p = 0.05$).

5.2.9 Genetic variation

Genetic variation based on polymorphisms at 17 microsatellite loci was measured (within each sample across all five populations) as the mean number of alleles across loci ($A$), the mean allelic richness across loci ($AR$), mean observed heterozygosity ($H_O$) and mean expected heterozygosity ($H_E$). Genetic variation was estimated for all sites, separating into years and generations and for samples chosen for STRUCTURE analysis. FSTAT V. 2.9.3.2 [293] was used to estimate $A$ and $AR$, as it uses a rarefaction based on the smallest population size to calculate allelic richness, to overcome the bias produced by variable sample sizes: the larger the sample size, the more alleles are expected. In addition, the frequency of rare alleles, $F_{IS}$ and significant heterozygosity deficit/excess, thus deviations from Hardy-Weinberg equilibrium, were also performed in FSTAT. Cervus [129] was used to estimate $H_O$ and $H_E$. GenAlEx 6.502 [307] was used to estimate the number of private alleles. Single factor analysis of variance (ANOVA) was performed in Excel to establish whether differences between the genetic diversity measures of sites were significant.

In order to investigate the relative contributions of inbreeding and population substructure within each site, the R script ConStruct [246] was employed for each site separately, where the algorithm estimates the joint likelihood of the percentage of the population with inbred parents and $F_{ST}$ within an existing multiocus data-set, using the method of Overall & Nichols [247]. Settings advised by the author were used,
setting the value of the inbreeding coefficient \( g \) to 0.0625, equivalent to parents being related as first cousins.

### 5.2.10 Relatedness coefficient

The method described by Wang [237] was employed to estimate relatedness coefficients between all individuals within each site, within each sub-population and between sub-populations within sites. Wang’s method was employed as this estimator correlated best with individuals of known pedigree from captive populations (see Chapter 3). Point estimates of relatedness were produced using the programme COANCESTRY (v.1.01.5)[76]. Genotyping error data was used to account for any mistyping, applying 100 threads for each run.

### 5.2.11 Sample pooling

Allele frequencies can vary both spatially and temporally in populations of a finite size, thus ideally samples taken within a population should belong to a single generation or collected at the same time (representing a single cohort of overlapping generations) [308]. However, obtaining an adequate sample size within a single point in time can be particularly difficult with the hazel dormouse as they are seasonal hibernators, are long-lived with overlapping generations and naturally found in low densities [12]. Therefore, it is necessary to identify any significant temporal genetic structuring including structuring between generations [294]. To test for this, pairwise \( F_{ST} \) within populations, between years and generations of each site were generated and are shown in table 5.1. First, the statistical power of the microsatellites to detect genetic differentiation was determined. POWSIM analysis indicated 100% chance of detecting \( F_{ST} \geq 0.005 \) (see Figure 5.2).

Once generations had been pooled for the remaining sites, another pairwise \( F_{ST} \) was performed between years for each site, in order to determine whether the samples could be pooled further (see table 5.2).

For most sites, \( F_{ST} \) values were not significantly different between generations, making it possible to pool different generations of the same year together [294]. All samples were pooled for Binsted, as only adults were sampled and there was no significant differentiation between years (see table 5.2).

The year 2013 was removed from the IoW samples; 2012 was significantly differentiated from 2013, but 2013 contained only three samples. Tilgate did not have significant differentiation between years, and so could be pooled together and treated
as one population. Mallydams showed significant genetic differentiation between years, therefore it was decided to pool generations of each year together and treat them as separate populations. For Wakehurst place, 2011 and 2013 had significant pairwise $F_{ST}$ and so would be treated as two populations; 2014 did not significantly differ from either year, but as the sample size was so low (W-14 $n = 2$), it was decided to remove 2014. Where different years in the same location were treated as separate populations, each pool was labelled with the first letter of the site name and then the year in which they were sampled; i.e. genotypes collected from Mallydams in 2012, would be M-12. Details of sample sizes for each site and year are given in table 5.3.
5.3 Results

5.3.1 Microsatellite markers

Microsatellite amplification success averaged across all 17 loci was 99.3%. All 17 loci tested were found to be polymorphic in all populations, accept Mav034, which was found to be monomorphic in the Mallydams population. A total of 155 alleles were identified for 17 loci in 288 individuals, sampled across years in five populations (Table 5.3). High estimates of null allele frequencies (> 0.05) were detected in 11 of the 17 loci, however the estimates were inconsistent across populations and thus, did not provide evidence that any one locus would produce consistent genotyping errors.

![Graph](image-url)

Fig. 5.2 Results of the POWSIM analyses based on 17 microsatellites. The y-axis represents the power of the markers to successfully recover the value of $F_{ST}$ indicated on the x-axis, expressed as the proportion of 1000 simulations.
### 5.3.2 Sample pooling

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2013A</td>
<td>0.0336</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.004</td>
<td>0.0484</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014A</td>
<td>0.019</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td>0.0333</td>
<td>0.0365</td>
<td>0.0237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014J</td>
<td>0.0062</td>
<td>0.05</td>
<td>-0.0001</td>
<td>0.0039</td>
<td>0.0063</td>
<td>0.0333</td>
<td>0.0368</td>
<td>0.0237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015A</td>
<td>0.0153</td>
<td>0.0148</td>
<td>0.0039</td>
<td>0.0063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015J</td>
<td>0.0377</td>
<td>0.0366</td>
<td>0.0532*</td>
<td>0.0532*</td>
<td>0.045</td>
<td>0.2157*</td>
<td>0.2581</td>
<td>0.2582*</td>
<td>0.1966</td>
<td>0.0438</td>
</tr>
</tbody>
</table>

Table 5.1 Pairwise $F_{ST}$ between years and generations for each site: a) IoW b) Wakehurst Place c) Mallydams Wood d) Tilgate Park e) Binsted. * indicate significant values after Bonferroni correction.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2012J</td>
<td>0.0222</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0222</td>
<td></td>
</tr>
<tr>
<td>2013A</td>
<td>0.0122</td>
<td>0.0101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0122</td>
<td>-0.0044</td>
</tr>
<tr>
<td>2013J</td>
<td>0.0373*</td>
<td>0.0406*</td>
<td>0.0059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0373*</td>
<td>0.0406*</td>
</tr>
<tr>
<td>2014A</td>
<td>0.0017*</td>
<td>0.0017*</td>
<td>0.0063</td>
<td>0.0108*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0017*</td>
<td>0.0017*</td>
</tr>
<tr>
<td>2014J</td>
<td>0.0494*</td>
<td>0.0574*</td>
<td>0.0329*</td>
<td>0.0447*</td>
<td>0.0119</td>
<td></td>
<td></td>
<td></td>
<td>0.0494*</td>
<td>0.0574*</td>
</tr>
<tr>
<td>2015A</td>
<td>0.0116</td>
<td>0.016</td>
<td>-0.0419</td>
<td>0.0041</td>
<td>-0.0448</td>
<td>-0.0312</td>
<td></td>
<td></td>
<td>0.0116</td>
<td>-0.0066</td>
</tr>
<tr>
<td>2015J</td>
<td>0.0516*</td>
<td>0.0573*</td>
<td>0.0233</td>
<td>0.0481*</td>
<td>0.0229</td>
<td>0.0208</td>
<td></td>
<td></td>
<td>0.0516*</td>
<td>0.0573*</td>
</tr>
</tbody>
</table>

Table 5.2 Pairwise $F_{ST}$ between years for each site: a) IoW b) Mallydams Wood c) Wakehurst Place d) Binsted e) Tilgate Park. * indicate significant values after Bonferroni correction.
5.3 Results

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>A</th>
<th>AR</th>
<th>$H_O$</th>
<th>$H_E$</th>
<th>$F_{IS}$</th>
<th>$N_{AP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOW</td>
<td>90</td>
<td>5.47</td>
<td>3.69</td>
<td>0.49</td>
<td>0.53</td>
<td>0.089*</td>
<td>13</td>
</tr>
<tr>
<td>M-12</td>
<td>32</td>
<td>5.12</td>
<td>4.14</td>
<td>0.66</td>
<td>0.63</td>
<td>-0.05†</td>
<td>1</td>
</tr>
<tr>
<td>M-13</td>
<td>46</td>
<td>5.24</td>
<td>4.09</td>
<td>0.64</td>
<td>0.63</td>
<td>-0.01</td>
<td>0</td>
</tr>
<tr>
<td>M-14</td>
<td>40</td>
<td>5.53</td>
<td>4.31</td>
<td>0.66</td>
<td>0.65</td>
<td>-0.019</td>
<td>2</td>
</tr>
<tr>
<td>M-15</td>
<td>9</td>
<td>3.41</td>
<td>3.41</td>
<td>0.65</td>
<td>0.60</td>
<td>-0.102†</td>
<td>0</td>
</tr>
<tr>
<td>Binsted</td>
<td>18</td>
<td>4.65</td>
<td>3.99</td>
<td>0.63</td>
<td>0.60</td>
<td>-0.049</td>
<td>8</td>
</tr>
<tr>
<td>Tilgate</td>
<td>18</td>
<td>4.41</td>
<td>4.02</td>
<td>0.68</td>
<td>0.66</td>
<td>-0.042</td>
<td>4</td>
</tr>
<tr>
<td>W-11</td>
<td>12</td>
<td>4.47</td>
<td>4.24</td>
<td>0.63</td>
<td>0.66</td>
<td>0.043</td>
<td>0</td>
</tr>
<tr>
<td>W-13</td>
<td>18</td>
<td>4.41</td>
<td>3.93</td>
<td>0.61</td>
<td>0.59</td>
<td>-0.036</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 5.3 Genetic characteristics of pooled samples hazel dormouse populations including sample number ($n$), mean number of alleles across 17 microsatellite loci ($A$) mean allelic richness across 17 microsatellite loci and standardised for a sample size of 9 loci ($AR$), mean observed heterozygosity ($H_O$), mean expected heterozygosity ($H_E$), multi locus estimates for $F_{IS}$ and the number of private alleles per population ($N_{AP}$). Populations with significant heterozygosity deficiencies are indicated by * and those with significant excess heterozygosity are indicated by † according to an adjusted (5%) nominal level calculated per group.

5.3.3 Isolation - by - distance

Isolation - by - distance analysis revealed a significant positive correlation between geographical distance and genetic distance ($F_{ST}$), indicating that genetic distances increases with increasing geographic distance increases (see Figure 5.3).

5.3.4 Among population genetic structure

The Isle of Wight and Sussex populations

The program STRUCTURE was employed for all clustering analysis with no a priori spatial information. At the highest hierarchical level, the Evanno method [297] suggests that the data consists of two main clusters: the Isle of Wight and East Sussex with a West Sussex populations showing mixed ancestry (see Figure 5.4).

Pairwise $F_{ST}$ was estimated between the IoW, East Sussex (Mallydams Wood) and pooled West Sussex populations. The IoW and East Sussex showed very great differentiation ($F_{ST} = 0.26$), the IoW and West Sussex showed great differentiation ($F_{ST} = 0.19$) and East and West Sussex showed moderate levels of differentiation ($F_{ST} = 0.12$). All significant after Bonferroni correction (see Table 5.4).

Pairwise $F_{ST}$ between each Sussex population and the IoW were also significant after Bonferroni correction and ranged between 0.25 and 0.27, showing considerable levels of genetic differentiation (see Table 5.5).
Fig. 5.3 Isolation-by-distance relationship between geographical distance (km) and genetic distance (pairwise $F_{ST}$) for all sites.

<table>
<thead>
<tr>
<th></th>
<th>IoW</th>
<th>East Sussex</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Sussex</td>
<td>0.26*</td>
<td></td>
</tr>
<tr>
<td>West Sussex</td>
<td>0.19*</td>
<td>0.12*</td>
</tr>
</tbody>
</table>

Table 5.4 Pairwise $F_{ST}$ for IoW, East Sussex and pooled West Sussex populations. * indicate significant values after Bonferroni correction.

<table>
<thead>
<tr>
<th></th>
<th>IoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-13</td>
<td>0.27*</td>
</tr>
<tr>
<td>Binsted</td>
<td>0.25*</td>
</tr>
<tr>
<td>Tilgate</td>
<td>0.25*</td>
</tr>
<tr>
<td>W-13</td>
<td>0.26*</td>
</tr>
</tbody>
</table>

Table 5.5 Pairwise $F_{ST}$ for all sites. * indicate significant values after Bonferroni correction.
Fig. 5.4 Bar plot showing the estimated membership coefficient (Q) for each individual. Each vertical bar corresponds to one individual and each cluster is represented by a different colour. Groups of individuals are labelled by site (M: Mallydams; W: Wakehurst Place; B: Binsted Wood; T: Tilgate and I: IoW).
Measures of $AR$ and $H_E$ were significantly higher in the pooled West Sussex population than those found in the IoW population (ANOVA: $AR$ $p = 0.022$; $H_E$ $p = 0.02$) (see Figure 5.5). However, when each Sussex site is considered separately together with the IoW, there is no significant difference for the diversity measures $AR$ and $H_E$ ($p > 0.05$).

![Figure 5.5](image)

**Fig. 5.5** Genetic diversity measures for IoW, East Sussex and pooled West Sussex populations a) shows mean allelic richness ($AR$) and b) shows mean expected heterozygosity ($H_E$) with standard error bars.

**Sussex populations**

For the Sussex populations, the Evanno method [297] identified two genetically distinct clusters corresponding to the sites in West Sussex and the site in East Sussex (Mallydams Wood), with little to no admixture (see Figure 5.6). The East Sussex group had one individual that was different and was identified as a dormouse found at Seaford train station, opportunistically sampled, and which was later released into Mallydams Wood.
5.3 Results

The West Sussex group was further clustered into three genetically distinct clusters, which correspond to the geographical sites (Binsted, Tilgate and Wakehurst Place) (see Figure 5.7).

Fig. 5.6 Bar plot showing the estimated membership coefficient ($Q$) for each individual. Each vertical bar corresponds to one individual and each cluster is represented by a different colour, West Sussex is in blue and East Sussex is in green. Groups of individuals are labelled by site (M: Mallydams Wood; W: Wakehurst Place; B: Binsted Wood; T: Tilgate).

All pairwise $F_{ST}$ values among sample sites were high (>0.1) and significant, ranging from 0.13 - 0.21, confirming that there was limited gene flow between the sites (see Table 5.6).

Detection of immigrants

Assignment tests revealed a total of six individual migrants across all populations, meaning that 97.92% of individuals were residents ($n = 288$) with over a 95% likelihood, indicating low level immigration. Out of the six likely immigrants, three were detected
Fig. 5.7 Bar plot showing the estimated membership coefficient ($Q$) for each individual. Each vertical bar corresponds to one individual and each of the clusters is represented by a different colour, Binsted is in green, Tilgate Park in blue and Wakehurst Place is in light blue. Groups of individuals are labelled by site (B: Binsted Wood; T: Tilgate and W: Wakehurst Place).

<table>
<thead>
<tr>
<th></th>
<th>M-13</th>
<th>Binsted</th>
<th>Tilgate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binsted</td>
<td>0.21*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tilgate</td>
<td>0.17*</td>
<td>0.18*</td>
<td></td>
</tr>
<tr>
<td>W-13</td>
<td>0.19*</td>
<td>0.2*</td>
<td>0.13*</td>
</tr>
</tbody>
</table>

Table 5.6 Pairwise $F_{ST}$ for all Sussex sites. * indicate significant values after Bonferroni correction.
in the IoW population (1.04% of total \( n = 101 \)), two in Mallydams (1.6% of total \( n = 127 \)) and one in Wakehurst Place (0.35% of total \( n = 33 \)). No immigrants were detected in Binsted \( n = 18 \) or Tilgate \( n = 18 \), but this could be due to the small sample sizes. Those detected in the IoW populations were all adult males and represented 3% of the population. The two detected in Mallydams consisted of one adult male and one juvenile female, the latter being a known immigrant, introduced into the population after being found at Seafront train station. Thus, only 0.8% of the population were identified as migrants. At Wakehurst Place, the one migrant was an adult female and represented 3.03% of the population.

5.3.5 Genetic variation among populations

Genetic variation

Genetic diversity measures for each site are shown in Table 5.3 The mean number of alleles ranged from 3.41 ± 0.32 (M-15) and 5.53 ± 0.38 (IoW). Mean allelic richness estimates varied from 3.41 ± 0.32 (M-15) to 4.24 ± 0.29 (W-11), but were not consistent with allelic number in identifying populations with the highest and lowest diversity, showing the importance of adjusting allelic richness for population size. Observed and expected heterozygosities were moderate for all sites, ranging from 0.49 (IoW) to 0.68 (Tilgate) and 0.53 and 0.66 respectively (Tilgate and W-11). Expected heterozygosities were higher than observed for the Isle of Wight and W-11 and lower for all other populations. Departures from Hardy - Weinberg equilibrium after Bonferroni correction was observed in three of the nine populations (see Table 5.3). Significant heterozygote deficiency was detected for the IoW, with five of the 17 loci showing heterozygote excess and 12 showing heterozygote deficiency. Significant heterozygosity excess was found for M-12 and M-15, due to the differences in allele frequencies of the sexes (see Figure 5.8).

Six of the nine populations contained private alleles (see Table 5.3). Five of the private alleles found in the Sussex populations were also found in samples taken from populations in Sussex, but not included in this study due to extremely small sample sizes \( n \leq 3 \). These populations being Beckley, Bexhill, Wadhurst and Worth.

The majority of alleles were at a low frequency \(< 0.1\) for all populations, except Tilgate and M-15. For Tilgate and M-15, the majority of alleles fell into the intermediate frequency \((0.1 - 0.2)\). Populations that have suffered bottlenecks are expected to show a reduction in low frequency alleles, \(< 0.1\) [262]. Figure 5.9 shows the proportion of alleles, across markers, that fall into low, medium and high allele frequency categories.
Fig. 5.8 Allele frequencies for 17 loci that differed between males and females in the populations a) M-12 and b) M-15.
Fig. 5.9 Proportion of alleles across 17 microsatellite loci that occur at different frequencies, calculated per each sample pool. White bars represent samples from the IoW, patterned white and black bars indicate samples pooled into years for Mallydams, East Sussex. Coloured solid filled bars represent sampled populations from Binsted and Tilgate West Sussex and grey and black patterned bars represent samples pooled into years for Wakehurst Place, West Sussex.
Fig. 5.10 Genetic diversity measures for each site a) shows mean allelic richness (AR) and b) shows mean expected heterozygosity ($H_E$) with standard error bars.
5.3 Results

5.3.6 Within site genetic variation and structure

Only Mallydams, Wakehurst Place and the IoW had enough sampled individuals in order to investigate population structure at the local level. In order to investigate the possibility of genetic structuring and or inbreeding within each site, all samples per site were analysed individually using ConStruct (Overall 2016).

**Mallydams**

The results of the ConStruct analysis are presented in Figure 5.11. The plot shows the joint likelihood of the causes of excess homozygosity: cryptic substructure ($F_{ST}$) and/or consanguinity, where $C_g$ is the proportion of the sample inbred to degree $g$, set at 0.0625, which is equivalent to first cousin mating and the level at which inbreeding is expected to impact fitness. The support limit for the maximum likelihood value is given by $G$. The maximum likelihood value of $F_{ST}$ and $C_g = 0$ ($G = 0.004$).

Because the maximum likelihood estimate of inbreeding was zero (Figure 5.11) clustering analysis (which assumes zero inbreeding) was employed to investigate the possibility of structuring further, employing *a priori* population number of four, reflecting the significant differentiation between years. The Evanno method [297] indicated two genetically distinct clusters with little admixture between them (see Figure 5.12).

Each individual was assigned to the subpopulation it was designated to by STRUCTURE (A and B). Population A contained 112 individuals, whereas population B contained just 14 individuals. Relatedness was estimated between all individuals and mean relatedness was estimated within and between the two clusters (see Figure 5.13). Individuals within cluster A were not related on average (mean $r = 0.002$), but those in cluster B were highly related on average (mean $r = 0.47$), to the equivalent of parent/offspring or full sib relationships; indicating that population B was in fact a family group, confirming that there was no genetic structuring within Mallydams.

**Wakehurst Place**

ConStruct analysis revealed maximum likelihood values of $F_{ST} = 0.042$ and $C_g = 0$. However, the support limits indicate that other scenarios (e.g., $F_{ST} = 0$, $C_g = 0.8$) are not ruled out (see Figure 5.14).

Again, due to the maximum likelihood estimate of inbreeding being zero ($C_g = 0$) clustering analysis was employed to investigate low level of structuring further employing *a priori* population number of three, reflecting the significant differentiation
Fig. 5.11 Contours representing the joint likelihood of the causes of excess homozygosity: cryptic substructure ($F_{ST}$) and consanguinity, where $C_g$ is the proportion of the sample inbred to degree $g$, which in this case is 0.0625 for the Mallydams Wood sample.
Fig. 5.12 Bar plot showing the estimated membership coefficient ($Q$) for each individual from Mallydams wood. Each vertical bar corresponds to one individual and based on a priori of four populations.
Fig. 5.13 Mean relatedness ($r$) within and between subpopulations designated by STRUCTURE for Mallydams wood.
Fig. 5.14 ConStruct results for all individuals within Wakehurst Place, with \( g \) set at 0.0625
Genetic Variation and Population Structure of the British Hazel Dormouse
(Muscardinus avellanarius)

between years. The Evanno method [297] indicated three genetically distinct clusters with little admixture between them (see Figure 5.15). Each individual was assigned to the cluster it was designated to (Cluster A, B and C). Cluster A mainly consisted of individuals sampled in 2011, Group B from those sampled in 2013 and cluster C consisted of juveniles from 2013 and adults from 2014, thus largely agreeing with the pooling by year.

![Fig. 5.15](image)

**Fig. 5.15** Bar plot showing the estimated membership coefficient ($Q$) for each individual for the Wakehurst Place site. Each vertical bar corresponds to one individual based on three a priori populations.

Relatedness estimates revealed that, on average, cluster A (blue) was unrelated (mean $r = -0.076$), cluster B (green) was related on average to the degree of half sibs (mean $r = 0.27$) and individuals within cluster C (light green) were related on average by $r = 0.17$, slightly more than first cousins (see Figure 5.16). When considering only adults for each cluster, cluster C has only two adult individuals and the relatedness between them drops to $r = 0.0648$ (see Figure 5.17), indicating that the structuring could be due to the presence of multiple generations. Thus, juveniles were removed from
the sample pool and evaluated again using ConStruct, and all evidence of structuring disappeared (Maximum likelihood $F_{ST} = 0$, $C_g = 0$, $G = 0.0004$) (see Figure 5.18).

![Fig. 5.16 Relatedness within and between assigned STRUCTURE groups including all individuals for Wakehurst Place.](image)

Isle of Wight

ConStruct analysis revealed maximum likelihood values of $F_{ST} = 0.025$, $C_g = 0.4$ and $G = 0.00035$ (Figure 5.19)

Samples were taken from seven designated woodlands, making up four woodland patches with a railway separating the largest patch from the rest. In order to investigate whether the railway was responsible for the genetic structuring, samples were first pooled by woodland name and a pairwise $F_{ST}$ was performed to determine if there was any genetic differentiation (see Table 5.7). Small, but significant differentiation was found between the designated woodlands: Sandpit and Briddlesford north, mid and
Fig. 5.17 Relatedness within and between assigned STRUCTURE clusters only including adults.
Fig. 5.18 ConStruct results for adults only within Wakehurst Place, with \( g \) set at 0.0625.
Fig. 5.19 ConStruct results for Isle of Wight, with $g$ set at 0.0625.
5.3 Results

<table>
<thead>
<tr>
<th></th>
<th>BM</th>
<th>BN</th>
<th>BS</th>
<th>BW</th>
<th>D</th>
<th>MU</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBN</td>
<td>0.0128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS</td>
<td>0.0542</td>
<td>0.0314</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BW</td>
<td>0.0163</td>
<td>0.033</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.0185</td>
<td>0.0289</td>
<td>0.0339</td>
<td>-0.0036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MU</td>
<td>0.0237</td>
<td>0.0445*</td>
<td>0.0348</td>
<td>0.0057</td>
<td>0.0165</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>0.0477*</td>
<td>0.0748*</td>
<td>0.0436</td>
<td>0.0093</td>
<td>0.0362</td>
<td>0.0561*</td>
</tr>
</tbody>
</table>

Table 5.7 Pairwise $F_{ST}$ between all designated woodlands present in the Isle of Wight site. BM: Briddlesford copse mid; BN: Briddlesford copse north; BS: Briddlesford copse south; BW: Big Wood; D: Dunnage; MU: Moorwood West/Uppersheepwash and S: Sandpit Copse. * indicates $p \leq 0.05$.

<table>
<thead>
<tr>
<th></th>
<th>Brid</th>
<th>PBWS</th>
<th>D</th>
<th>b)</th>
<th>Brid</th>
<th>BWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BWS</td>
<td>0.038*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.017</td>
<td>0.019*</td>
<td></td>
<td></td>
<td>DMU</td>
<td>0.018*</td>
</tr>
<tr>
<td>MU</td>
<td>0.029*</td>
<td>0.039*</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.8 Pairwise $F_{ST}$ between grouped woodlands within the IoW site. a) depicts samples pooled by woodland patch (Brid: Briddlesford Copse; BWS: Big Wood and Sandpit Copse; D: Dunnage, MU: Moorwood West and Uppersheepwash) b) depicts the smallest number of possible pools (DMU: Dunnage with Moorwood West and Uppersheepwash).* indicates significant after Bonferroni correction.

Moorwood West / Uppersheepwash and between Moorwood West / Uppersheepwash and Briddlesford North.

Once it was determined which patches showed no significant differentiation between each other, samples were pooled by woodland patch (see Table 5.8 a)). All patches showed significant differentiation, accept for Dunnage and Moorwood West / Uppersheepwash and Dunnage and Briddlesford. It was decided to pool Dunnage and Moorwood West / Uppersheepwash together, as the two woodland patches were directly connected. The pairwise $F_{ST}$ showed small, but significant differentiation between the three sample pools (see Table 5.8 b)).

This information was then used as a priori population number for clustering analysis using STRUCTURE. The Evanno method [297] suggests that the data consists of two genetically distinct clusters with some admixture, see Figure 5.20. Each individual was assigned to the cluster in which it shared 60% or more of its ancestry (A and B). Those without a clear cluster were assigned to a third group of mixed ancestry (M). Pairwise $F_{ST}$ revealed significant, moderate levels of genetic divergence between clusters A and B and significant but little differentiation between the main clusters and those of mixed ancestry (Table 5.9).
Genetic Variation and Population Structure of the British Hazel Dormouse

(Muscardinus avellanarius)

Fig. 5.20 Bar plot showing the estimated membership coefficient ($Q$) for each individual based on a priori three populations for the IoW site. Each vertical bar corresponds to one individual and each cluster has a different colour (subpopulation A: light green, B: dark green, M: mixed ancestry)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td></td>
<td>0.09*</td>
</tr>
<tr>
<td>M</td>
<td>0.04*</td>
<td>0.03*</td>
</tr>
</tbody>
</table>

Table 5.9 Pairwise $F_{ST}$ between subpopulations assigned by STRUCTURE and those with mixed ancestry.* indicates significant after Bonferroni correction.
The GPS locations of the individuals of each category were then plotted onto a map of the site to determine whether any obvious landscape features could have caused the apparent sub-populations (see Figure 5.21). Cluster A is found throughout the woodland patches, whereas cluster B is confined to Briddlesford patch, one side of the railway. Those of mixed ancestry were sampled within Briddlesford, Big Wood and Dunnage.

Fig. 5.21 Map displaying the location of each subpopulation as defined by the program STRUCTURE (\textit{a priori} $K = 2$). Each woodland patch comprises: Big Wood; Sandpit Copse; Dunnage and Moorwood West and Upper Sheepwash. The black dots represent cluster A, the red dots cluster B and those of mixed decent are represented by split black and red.

Mean relatedness scores were estimated both within (A; B and M) and between subpopulations (A\_B; A\_M and B\_M). Mean relatedness was generally higher within clusters than between clusters, however relatedness estimates between clusters A and B (A\_B) had a slightly higher mean relatedness score, but not significantly. Individuals within clusters were not on average highly related to one another (cluster A mean $r = -0.049$; cluster B mean $r = 0.0067$; cluster M mean $r = 0.012$), however individuals
Genetic Variation and Population Structure of the British Hazel Dormouse

*Genetic Variation and Population Structure of the British Hazel Dormouse

within cluster A were significantly less related to one another on average than individuals
within cluster B or cluster M, indicating that the genetic structuring is real and not
due to the inclusion of family groups (see Figure 5.23).

Fig. 5.22 Mean relatedness scores for within (e.g. only including individuals cluster A) and between
(e.g. including individuals from cluster A and B) the two subpopulations and the group of mixed
ancestry.

Genetic diversity measures for each subpopulation identified by STRUCTURE
are shown in Table 5.10. Both the mean number of alleles and allelic richness are
highest in Cluster A. Observed heterozygosity is lower than expected heterozygosity in
Cluster A and higher than expected in Cluster B. Significant departure from HWE
after Bonferroni correction was found in Cluster A, showing heterozygosity deficiency.

Genetic diversity measures were then generated for the adults only of each population
to see if the results were similar, as allele frequencies can change between generations
[294] and, thus, effect the diversity measures $A_R$ and $H_O$. Again, Cluster A showed
significant deviations from HWE, with a significantly larger $F_{IS}$ than expected due to
5.3 Results

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>A</th>
<th>A_R</th>
<th>H_O</th>
<th>H_E</th>
<th>F_{IS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>69</td>
<td>5.35</td>
<td>3.54</td>
<td>0.51</td>
<td>0.55</td>
<td>0.069*</td>
</tr>
<tr>
<td>B</td>
<td>21</td>
<td>3.65</td>
<td>2.85</td>
<td>0.44</td>
<td>0.42</td>
<td>-0.041</td>
</tr>
<tr>
<td>Mix (M)</td>
<td>8</td>
<td>3.59</td>
<td>3.48</td>
<td>0.53</td>
<td>0.54</td>
<td>0.095</td>
</tr>
</tbody>
</table>

Table 5.10 Genetic characteristics of subpopulations designated by STRUCTURE: number of samples (n), mean number of alleles across 17 microsatellite loci (A) mean allelic richness across 17 microsatellite loci and standardised for a sample size of 9 genes (A_R), mean observed heterozygosity (H_O), mean expected heterozygosity (H_E), multilocus estimates for F_{IS} and the number of private alleles per population (N_{AP}). Populations with significant heterozygosity deficiencies are indicated by * according to an adjusted (5%) nominal level calculated per group.

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>A</th>
<th>A_R</th>
<th>H_O</th>
<th>H_E</th>
<th>F_{IS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>39</td>
<td>5.18</td>
<td>3.40</td>
<td>0.50</td>
<td>0.55</td>
<td>0.094*</td>
</tr>
<tr>
<td>B</td>
<td>12</td>
<td>3.18</td>
<td>2.65</td>
<td>0.46</td>
<td>0.42</td>
<td>-0.104†</td>
</tr>
<tr>
<td>Mix (M)</td>
<td>7</td>
<td>3.41</td>
<td>3.28</td>
<td>0.45</td>
<td>0.51</td>
<td>0.133</td>
</tr>
</tbody>
</table>

Table 5.11 Genetic characteristics of adults within subpopulations designated by STRUCTURE: number of samples (n), mean number of alleles across 17 microsatellite loci (A) mean allelic richness across 17 microsatellite loci and standardised for a sample size of 9 genes (A_R), mean observed heterozygosity (H_O), mean expected heterozygosity (H_E), multilocus estimates for F_{IS} and the number of private alleles per population (N_{AP}). Populations with significant heterozygosity deficiencies are indicated by * and those with significant excess heterozygosity are indicated by † according to an adjusted (5%) nominal level calculated per group.

Cluster B adults show a higher $H_O$ than $H_E$, indicating heterozygosity excess (see Table 5.11).

$F_{IS}$ estimates were then generated for each sex within Cluster A. Only males showed a statistically significantly larger $F_{IS}$ than expected (males $F_{IS} = 0.172$, p = 0.0015).

To investigate this result further, individual $F_{IS}$ values were then estimated using the maximum likelihood option in the R program ConStruct for Cluster A and mean relatedness was estimated including all individuals and only the adults. Males had a significantly higher mean $F_{IS}$ score than females, both when including all individuals and when only including adults (see Figure 5.24 a) and b)). On average neither sex was highly related. Males were significantly less related to each other than females, both when all individuals were included and when only including the adults.

The heterozygosity excess seen in adults from Cluster B was likely to be produced by differences in allele frequencies of the sexes, as shown in Figure 5.25.
Fig. 5.23 Mean relatedness scores for within (e.g. only including individuals cluster A) and between (e.g. including individuals from cluster A and B) the two subpopulations and the group of mixed ancestry.
5.3 Results

Fig. 5.24 a) Mean $F_{IS}$ for males and females, both including and not including juveniles within Cluster A; b) Mean relatedness estimates for males and females, both including and not including juveniles within Cluster A.
Fig. 5.25 a) Allele frequencies for 17 microsatellite loci that differed between males and females in Cluster A.
5.4 Discussion

The current study explored patterns of genetic variation of an endangered arboreal mammal, at multiple scales within five sites with different habitat characteristics. The overall purpose was, firstly, to identify whether any of the sampled populations had diversity levels low enough to warrant immediate concern for the population’s persistence (i.e., low diversity would predict poor health) and, secondly, to use the observed patterns of variation to identify which habitat characteristics pose the greatest threat to the continued survival of the hazel dormouse. The levels of expected heterozygosity (HE) are often used as a measure of genetic variation, which is a measure from 0 (no variation) to 1 (complete variation). A measure of HE of 0.3 or lower for microsatellite markers is considered to merit concern for a population [309], whereas populations with HE of 0.7 and above are considered to have high genetic variation [310].

5.4.1 Among population genetic structuring and variation

Sussex and the Isle of Wight

The Isle of Wight is situated 6km off the coast of Southern England with the Solent Straight separating the island from the mainland and representing a complete barrier to dormouse dispersal. The land bridge that connected the Isle of Wight to the mainland survived until around 7.5kya when it became inaccessible from the mainland [311]. Island populations are generally believed to have less genetic variation than their equivalent mainland populations [312] and this is attributed to founder events, increased inbreeding and the stronger effects of drift caused by smaller population numbers and decreased immigration [313]. Taking these two issues into account, it was predicted, firstly, that the Isle of Wight would be highly differentiated from all the mainland populations situated in the species stronghold of Sussex and, secondly, the Isle of Wight population would have significantly less genetic variation than all Sussex populations. Clustering analysis using Bayesian assignment tests found two main populations at the highest hierarchical level when including all samples, splitting the samples into an Isle of Wight and an East Sussex subpopulation with West Sussex showing mixed ancestry (see Figure 5.4); coinciding with the lengthy physical separation and confirming little to no gene flow between the island and Sussex populations.

Levels of genetic variation were significantly lower for the IoW when compared to the pooled West Sussex samples, both when measured by $A_R$ and $H_E$, but this significance disappeared when considering each sampled woodland separately. The genetic variation found within the IoW population was actually higher than that found in peripheral
mainland populations (IoW, $A_R = 3.69$; $H_E = 0.53$; East Cornwall, $A_R = 2.08 - 2.67$, $H_E = 0.40 - 0.51$) [50]. This result contrasts with a study comparing populations of marsupials found in pristine habitats on islands, compared to remnant mainland populations affected by habitat loss and fragmentation. These highly fragmented mainland populations still harboured higher levels of diversity than populations found on surrounding islands, even when the mainland population showed signs of a recent bottleneck [314]. The Isle of Wight is a large island of 380 km$^2$ with 1602.43 ha of ancient woodland, thus ample habitat for the hazel dormouse. The sites sampled by Mills [50] were highly fragmented with non-optimal conditions, whereas Briddlesford, IoW, is managed in order to promote dormouse numbers; therefore it is highly likely that the Briddlesford site has conditions more conducive to larger populations of dormice than those sampled in Eastern Cornwall and, thus, have higher levels of variation.

The external environmental conditions, may also play a part in the differences seen. Although islands are generally considered to have more variable and harsher conditions [313], this may not be the case for the Isle of Wight in comparison with Cornwall, in the context of the hazel dormouse.

Cornwall is at the very edge of the dormouse range and has an oceanic climate, typified by cool winters and warm summers, with precipitation all year, whereas the Isle of Wight has a milder sub-climate than other areas of Great Britain. Winters are generally warmer and wetter in Cornwall compared with those on the Isle of Wight (Average winter temp: Cornwall = 1 - 4°C, IoW = -1.6 - 4°C; Average winter rainfall: Cornwall = 341 - 928mm, IoW = 48 - 125mm [211], which are associated with high winter mortality [12]. Cornwall is also positioned at the tip of the country and is subjected to strong westerly winds that bring changeable weather conditions. The positioning of the Isle of Wight close to the mainland within the English Channel makes it relatively sheltered, with conditions not unlike those found within the dormouse stronghold of Sussex and therefore provides more stable conditions than those found in Cornwall. These differences could mean that the founding populations of the Isle of Wight might have been larger than those founding populations in East Cornwall, contributing to the lower level of diversity found in the present day.

The number of private alleles, another common measure of diversity, was found to be higher in the Isle of Wight population compared to any of the Sussex populations (see Table 5.3), this is a common feature of island populations and could be due to the random effect of drift increasing the frequency of alleles that are rare or absent on the mainland [315]. However, it may be that the alleles which are deemed private
are in fact present within nearer mainland populations of dormice, such as those in Hampshire, which were not sampled in this study.

The Isle of Wight population significantly deviated from HWE with an $F_{IS}$ of 0.089 ($p < 0.05$). The inbreeding coefficient was significantly higher in the Isle of Wight population compared with the Sussex populations M-12, M-13, M-15 and Binsted ($p = <0.05$), but not when compared with the populations W-11, W-12 and Tilgate. Binsted and Mallydams do not contain any obvious within site barriers to gene flow, whereas Wakehurst Place, Tilgate and the Isle of Wight do, which could explain this result and will be discussed in further detail below.

**Sussex population structuring**

Clustering analysis performed when including all Sussex samples identified two clusters: East Sussex and West Sussex, indicating a divergence between the two counties with a pairwise $F_{ST} = 0.12$ ($p < 0.05$). This result was predictable given that the East Sussex sample site (Mallydams) is situated towards the far East of the County (see Figure 5.1) with a minimum distance of 88km (Euclidean) between the sites Wakehurst Place and Mallydams. An AMOVA revealed that even though the effect of region (East and West Sussex) was significant, only 3% of the total variance was accounted for by regional differences, whereas 14% was accounted for by among population variation within regions. West Sussex samples were assigned to three clusters by STRUCTURE and these clusters corresponded to sample site. Pairwise $F_{ST}$ results between all Sussex sites were significant after Bonferroni correction and ranged between moderate (W-13 - Tilgate: $F_{ST} = 0.13$) and large (Binsted - M-13: $F_{ST} = 0.21$), indicating limited gene flow, which is unsurprising given the distance between the sites ($>10km$). A significant ($p = 0.0154$) isolation-by-distance relationship was found between pairwise $F_{ST}$ and geographical distance (km), confirming that distance between sites accounts for the regional substructuring found.

**Detection of first generation immigrants**

First generation immigrants were detected in three of the five sample locations, 3% of the Isle of Wight and Wakehurst Place populations were immigrants and 0.8% of Mallydams, showing low level gene flow into the woodlands, which is consistent with the findings of Naim [75]. The absence of immigrants detected from Tilgate and Binsted (East Sussex) could be due to the surrounding habitat. The area sampled within Tilgate park is bordered on all sides by potential barriers to dispersal and is separated from the nearest woodland by the M23 motorway. Binsted is separated from
the largest woodland patch by the A27 duel carriageway, thus the results could indicate that dormice from surrounding areas are not able to access woodlands. However, only 18 individuals were found in each of these two woodlands, therefore the likelihood of sampling a migrant is small (migration would need to be above 5.5%), as such the apparent lack of detection of immigration is likely attributable to the low sample number, rather than any habitat effects.

Natal dispersal has been reported in both sexes of juvenile hazel dormouse via radio-tracking and capture mark recapture studies, with males generally dispersing further than females [10][11]. It has been suggested that distances travelled are not only dependent on sex, age and time of birth, but are also density-dependent [10][11]. Dispersing male dormice have been reported to typically travel between 1.5 and 1.8km, although a dispersal distance of 3.3km has been reported in Germany and as far as 7km within a forest in Southwest Germany [11]. Females mostly disperse less than 1km. These studies concur with genetic studies, where rare inter-patch movement greater than 1km has been detected [75]. The same study also found evidence of male biased natal dispersal. Out of the five natural immigrants detected in this study, four of them were adult males and one was an adult female. Thus, the results seem to coincide with both tracking and genetic evidence that males are the main dispersers, but females do also sometimes disperse.

**Among population genetic variation**

The proportion of alleles present at each frequency category (Figure 5.9), where the largest proportion of alleles fell within the <0.1 frequency, indicated that most of the sampled subpopulations had not been through a recent bottleneck, with the exception of M-15 and Tilgate. For M-15, the sample size of \( n = 9 \) is likely to be the cause for not sampling rare alleles. Tilgate is surrounded on all sides by possible anthropogenic barriers to hazel dormouse dispersal and the part of the woodland that is sampled is patchy. Thus, it is possible that habitat loss and fragmentation through urban development has reduced the population size enough to cause the loss of alleles at low frequencies, as the first effect of a population decline is seen in the loss of rare and low frequency alleles [262]. The Tilgate sample size of \( n = 18 \) is also on the low side, although Binsted \( (n = 18) \), W11 \( (n = 12) \) and W13 \( (n = 18) \) all had rare alleles present in high proportions.

Levels of genetic variation as measured by expected heterozygosity and allelic richness, were moderate \( (A_R = 3.41 - 4.31; H_E = 0.53 - 0.66) \) in all sampled subpopulations (see Table 5.3). Similar levels have been reported in Devon by Mills...
[50] using 22 microsatellites, 17 of which were used in the present study \( (A_R = 3 - 4.02; H_E = 0.54 - 0.7) \); and by Naim [75] in reintroduced populations in Wales \( (A_R = 5.16 - 6.46, H_E = 0.64 - 0.69) \) using 10 microsatellites, 4 of which were used in this study; however, slightly higher levels were found in natural populations from the same study \( (A_R = 8.56 - 9.27, H_E = 0.68 - 0.7) \). Allelic richness in the natural population studied in Wales was much higher than the levels found within Sussex. This could be explained by a number of factors, firstly the sample size was much larger \( (n=296) \) than any of the Sussex populations, potentially allowing for more alleles within the population to be sampled. Secondly allelic richness accounts for the lowest population number, which in Naim’s 2010 study was \( n=138 \) [75], whereas in the current study the lowest was \( n = 9 \) and, lastly, a different suite of microsatellites were used, making direct comparisons difficult. The expected heterozygosity levels found were only marginally higher than those found in Sussex, therefore indicating that the differences seen in allelic richness are likely to be due to smaller sample sizes.

Levels of diversity \( (A_R \) and \( H_E \)) between sites did not differ significantly, indicating that the differences in woodland structure and levels of connectivity does not effect the amount of genetic diversity found in these hazel dormouse subpopulations. This does not agree with the findings of Naim [75], where significantly lower levels of \( H_E \) and \( A_R \) were found within fragmented dormouse populations compared with continuous habitats. However, the finding of the present study is consistent with a study by Fietz et al. [268] on the edible dormouse \( (Glis glis) \), where lower levels of genetic diversity \( (A_R, H_E \) and \( A_P \)) were not detected in fragmented woodlands compared to continuous woodlands. The findings of the current study could indicate a number of things: firstly that the moderate levels of genetic diversity within fragmented populations reflect high effective population sizes \( (N_e) \) for all the sites. Secondly, that the low, but present, level immigration detected at most sites (IoW, Mallydams and Wakehurst Place) is sufficient to prevent any significant loss of genetic diversity and, finally, that the initial levels of diversity within the populations were so high, that the subsequent human induced habitat loss and fragmentation did not happen long enough ago to have taken effect on allelic richness and heterozygosity.

### 5.4.2 Within population structure

#### Sussex populations

The subpopulation sampled at Mallydams presented no evidence of genetic structuring or inbreeding (see Figure 5.11), but clustering analysis was employed due to the signifi-
Genetic Variation and Population Structure of the British Hazel Dormouse

(Muscardinus avellanarius)

 significant $F_{ST}$ results between the years, therefore \textit{a priori} of four populations was considered for STRUCTURE analysis. Two clusters were identified, with some admixture; however, most of the population was assigned to one cluster, whereas only 14 individuals were assigned to the second cluster. This is likely the reason the ConStruct analysis failed to identify evidence of significant substructure, as this analysis considers the \textit{average} degree of population substructure, which can be negligible when the total population is composed largely of just a single subpopulation [246]. The inclusion of relatives in data analysis using STRUCTURE can cause the appearance of separate clusters and thus, genetic structuring, so relatedness coefficients were estimated in order to investigate this further [316][317]. The average relatedness within the smallest cluster revealed that it was the inclusion of a large family group that caused the apparent genetic structuring, as the cluster containing 14 individuals were related on average by $r = 0.47$, which is the equivalent of these individuals being, on average, related as closely as parent - offspring or full sibs. Therefore it was concluded that there was no significant population structure and Mallydams represented a single subpopulation unit, which was unsurprising due to the size of the wood (i.e., small) and the lack of any obvious habitat characteristics that could inhibit dormouse movement. The small size of the wood did suggest increased opportunities for inbreeding; however, evidence for this was absent. Across the four years of sampling, only one migrant was detected, clearly indicating very low levels of immigration, which again was unsurprising given the poor connectivity between the woodland patches present. This scenario of high relatedness but no inbreeding suggests the possibility that dormice display some form of kin discrimination in order to avoid mating with close relatives (see Chapter 3). However, it may be relevant that the sampling protocol, which employed nest boxes, biases the sampling towards residents, as migrants may be less inclined to use the nest boxes, though there has been no research into this subject to date. Another alternative is that due to the limited resource availability in such a small woodland, individuals are more likely to relocate from Mallydams to other woodlands not sampled in this study; i.e., disperse. Further investigation would be required that included the sampling of individuals from woodlands in close proximity to Mallydams in order to confirm this.

Wakehurst place is a small woodland situated in a habitat matrix of agricultural land with small to large woodland patches situated within a 5km radius, with good levels of connectivity between them. A reservoir cuts through the middle of the sampled areas within Wakehurst Place, which is expected to act as a complete barrier to dormouse dispersal and therefore gene flow. In light of these habitat characteristics it was predicted that significant genetic structuring would be detected within Wakehurst
Place, with separate subpopulations each side of the reservoir. No significant inbreeding was expected due to the presence of woodland patches on all sides, providing potential migrant sources. Although, significant within population genetic structuring was identified, inbreeding levels detected by ConStruct were ambiguous (Figure 5.14), though the most likely scenario was that no inbreeding at the level of first cousin mating was occurring (maximum likelihood $C_g = 0$), agreeing with initial predictions. Further investigation was needed in order to establish whether the water body had caused genetic structuring. STRUCTURE analysis assigned individuals to three groups that showed some admixture between them. The groups largely agreed with the year of sampling and individuals on each side of the reservoir were assigned to a single population, indicating that the reservoir was not the reason for the pattern in genetic variation observed. This implies that the foot bridges that traverse the reservoir at a number of places in addition to the tree canopy that spans the reservoir where the water narrows allows the dormice to cross, the latter being a known dispersal aid [229]. Tree lined roads, of which one crosses the reservoir south of Wakehurst Place, have also been recorded as habitat corridors for dormice [11][229] and could explain the reservoir not being the cause of the genetic structuring. To investigate further, relatedness estimates were calculated for each group. The average relatedness was high within two of the three groups (Figure 5.15). Both of these groups contained high numbers of juveniles, with many likely to be related to one another as sibs or half sibs, as they were found within the same nest box. When the juveniles were removed from the analysis, no genetic structuring was detected and thus it was concluded that the presence of relatives and overlapping generations influenced the STRUCTURE results and that gene flow appears to be uninhibited throughout Wakehurst Place.

Isle of Wight

The Briddlesford site on the Isle of Wight consists of a series of woodland patches, separated by agricultural fields, glades and rides with a railway line running through the middle. Due to this composition, it was predicted that the site would show significant genetic structuring; specifically the railway would present a barrier to dispersal for the hazel dormouse and due to the small size of the woods either side, positive evidence of inbreeding was expected. Significant, but low levels of structuring and high levels of inbreeding were detected (maximum likelihood values of $F_{ST} = 0.025$, $C_g = 0.4$, where $g = 0.0625$, Figure 5.19), which corresponds to 40% of the sample having parents related as closely as first cousins. Pairwise $F_{ST}$ estimates were used to establish the smallest number of woodlands that could be considered as connected ($= 3$), which was
used to inform the number of *a priori* subpopulations in the programme STRUCTURE. Two genetic clusters with moderate levels of admixture (see Figure 5.20). Pairwise $F_{ST}$ revealed moderate and significant levels of genetic differentiation after Bonferroni correction between each cluster ($F_{ST} = 0.09$). It was hypothesised that these two clusters represented the separate subpopulations (A and B) corresponding to those found either side of the railway. To test this, each sample was mapped according to the cluster in which it was assigned, with those of mixed descent being highlighted separately (see Figure 5.21). Individuals from both subpopulation A and those of mixed decent were found throughout the woodland patches, and thus both sides of the railway, whereas subpopulation B was only found within Briddlesford Copse (north of the railway). Thus, the hypothesis that the structuring detected was due to the railway representing a significant barrier to gene flow was rejected. Mean relatedness estimates within each group showed that, on average, individuals within groups were more related to each other than those between groups, but this was not always significant (see Figure 5.25). Average relatedness was low in all groups, indicating that the structuring could not be attributed to the sampling of family groups, thus warranting further investigation. Both $A_R$ and $H_E$ were higher in subpopulation A than B, but not significantly. Subpopulation A displayed significant heterozygosity deficiency indicating higher than average levels of inbreeding ($F_{IS} = 0.069$). Similar results were generated when only including adults, except levels of inbreeding within adults was moderately higher ($F_{IS} = 0.094$) and adults in subpopulation B displayed significant heterozygosity excess ($F_{IS} = -0.1$). Individual $F_{IS}$ and relatedness estimates were calculated for males and females in subpopulation A separately, both including all individuals and just the adults. The results indicated that average relatedness within males was significantly lower than females, but inbreeding levels were significantly higher in males than females, both when all individuals were included in the analysis and when only the adults were included. This observation indicates that, within population A, the dispersing males are showing a significant heterozygosity deficit. This suggests evidence of the Wahlund effect [69]. The Wahlund effect occurs when the heterozygosity deficiency observed is caused by the inadvertent sampling of multiple subpopulations. This effect is common within the dispersing sex, as both resident and migrant individuals are both sampled [14]. Therefore, as the Wahlund effect is the most likely explanation of the pattern detected, it would appear that in fact subpopulation A is not a population undergoing inbreeding.

The excess heterozygosity detected in adults from subpopulation B was most likely due to one of two things; 1) a difference in allele frequencies between the sexes
5.5 General conclusions

The results demonstrate that all the populations sampled from Sussex and the Isle of Wight show reasonable levels of genetic diversity, that are not significantly different from each other and show limited indication of inbreeding levels. Thus, these populations do not appear to be in immediate danger from loss of fitness and numbers, under the

and/or 2) the sampling of multiple family groups [14], which can manifest as excess heterozygosity. In this example the former was deemed most likely, because average relatedness was only $r = 0.0067$ (see Figure 5.23). The Isle of Wight site clearly shows genetic substructuring and, although this was to some extent predicted, the hypothesis that identifiable barriers within the site would cause the substructuring cannot be accepted, as there was no evidence that any of the physical barriers significantly disrupted gene flow. With subpopulation B, seemingly restricted to the north side of the railway and the presence of individuals of mixed ancestry present on both sides of the railway, leads to the conclusion that the railway can in fact be traversed. The lack of significant differentiation between the woods in Briddlesford Copse with Big Wood, Dunnage and Moorwood West/ Uppersheepswash on opposite sides of the railway lends further support to the notion that the railway did not impede dormouse movements. No other obvious physical barrier was apparent and thus, the hypothesis that barriers prevent dispersal and interrupt gene flow must be rejected.

The subpopulations of the Isle of Wight sample are significantly differentiated however, suggesting that something is still impeding gene flow. The subpopulation differentiation could be due to subtle differences in resource use, habitat preference or behaviour, which could mean that members from each subpopulation very rarely come into contact, or that mates are selected for based upon their similarities, keeping them distant with some admixture. Another possibility is that subpopulation A is the resident populations and subpopulation B represents juvenile dispersers entering the site from the large woodland to the north of Briddlesford Copse and, on finding suitable habitat, do not venture beyond it, whereas the offspring from unions between members of the separate sub-populations travel beyond the railway to establish a new territory. Further research, which tracked the movements of juveniles found within the site and around the site would be needed in order to determine whether this could be a plausible explanation for the substructuring detected. There could of course be a physical barrier that would not necessarily be perceived by humans, but which causes enough resistance to keep each subpopulation significantly differentiated.
assumption that levels of diversity measured at neutral loci reflect levels of diversity across the genome. Interestingly, none of the habitat features present in the study appeared to have any significant effects on heterozygosity levels or allelic richness as levels were not significantly different among sites. Many studies have highlighted that habitat size is an important factor for species persistence [117][291], as larger areas generally have more resources and therefore can house higher numbers of individuals, increasing the amount of genetic variation [14]. Thus it may be the lack of repeat sample sites with very similar conditions that caused any effects of habitat features on diversity levels to be missed. It is also important to state that, though there appears to be no immediate danger for these particular populations, any further habitat fragmentation and loss could lead to a decrease in genetic diversity and increased risk of extinction. Habitat quality is known to effect hazel dormouse survival rates [117], therefore continued efforts to preserve the remaining dormouse habitats is essential for preservation of the species in Britain. Another point would be that the populations sampled within the study are in an area known to be the species’ stronghold, partly due to high levels of woodland coverage and due to more optimal environmental conditions. It could be possible that populations with similar habitat characteristics, but which are situated within an area with environmental conditions less conducive to hazel dormouse survival, would show signs of reduced diversity. Thus, when applying these findings to other populations, such knowledge must be considered. Also, it is unknown whether neutral diversity correlates with fitness levels in this species and therefore further research using markers known to contribute to individual survival (i.e. major histocompatibility genes) would be justified.

The lack of genetic structuring in habitats with potential barriers to dispersal indicate that the hazel dormouse can navigate a number of different physical structures that may be assumed to pose some difficulty to a small, arboreal mammal. However, these results are in line with recent studies indicating that dormice are more mobile, travel greater distances [11][75] and traverse both natural and man-made structures, such as roads, that were previously deemed impassable, much more frequently than expected [117][280][118]. Genetic structuring was detected within the Isle of Wight populations, but no firm conclusions could be drawn on the cause of the genetic structuring. In light of both general findings, it would appear that the hazel dormouse is less affected by anthropogenic barriers than previously thought.
Chapter 6

The Phylogenetic Scale of British Hazel Dormouse Diversity: The Story So Far

Abstract

Phylogeography is concerned with the distribution of genetic lineages in a spatiotemporal context. Although historical connections of populations may be very different to current patterns of gene flow, both are relevant to the contemporary distribution of genetic variation. A number of studies to date have investigated the phylogeographic structure of the hazel dormouse, however the magnitude of differentiation among British populations remains unclear. Some authors suggest the levels are enough to warrant specialised conservation measures, whereas others have found little evidence of differentiation. The Cytochrome B and D-loop regions were sequenced for novel samples and combined with those generated by previous research in order to construct phylogenetic trees to clarify the levels of differentiation in hazel dormouse populations in Great Britain. No clear British lineages were detected and as such, previous recommendations of maintaining regional purity when considering individuals for reintroductions were not supported. Caution is advised to avoid the over-interpretation of genetic analysis. No isolation-by-distance relationship was found, implying that genetic-drift - gene flow equilibrium has not been achieved among British populations.
6.1 Introduction

Molecular markers can be used to reconstruct the genealogical history of a species (phylogeny), which can be used to inform on historical patterns of gene flow [37]. Although historical connections of populations may be very different to current patterns of gene flow, both are relevant to the contemporary distribution of genetic variation [37]. Throughout Europe, a species’ current distribution would have been greatly influenced by ice age events, the current interglacial having started around 10,000 years ago. Phylogenies, by establishing the evolutionary history of a genetic sequence can, in conjunction with geographic distributions, be used to identify the most likely refugial locations and infer post-glacial colonisation events. The detection of population structure at the highest hierarchical level (e.g., species) can be important for revealing long isolated populations that may have distinct gene pools and be subject to local adaptations [4] and lead to the assignment of evolutionary significant unit (ESU) status.

An ESU is a population unit that has high priority for conservation and merits separate management [318]. This is of particular relevance to species considered to be of conservation concern, such as the hazel dormouse, where maintaining genetic diversity is of prime importance to the long term survival of a species. However, the utilisation of ESUs is controversial, partly due to the inconsistent definitions of an ESU and thus, the difficulty in effectively identifying appropriate population units to conserve [318][4].

There have been a number of studies to date utilising hazel dormouse phylogenies, which have led to certain recommendations for their conservation. However, past narratives have been conflicting and thus, clarification is in order. The current study aims to amalgamate sequence data generated from previous research with novel data in order to clarify the level of structure within the British dormouse at the population scale.

6.1.1 The current phylogenetic status of the British hazel dormouse

The hazel dormouse (Muscardinus avellanarius) is a nocturnal arboreal mammal distributed across Europe, from the Mediterranean to Southern Sweden, eastwards towards Russia excluding Iberia and extends its range to parts of Asia Minor [8]. The hazel dormouse reaches its northwest limit of its European range in the UK, where it is found over much of Southern England, with isolated populations in northern Wales [7]. In parts of the *M. avellanarius* Northern range (UK, Netherlands, Sweden, Germany
and Denmark) populations are declining [1] and, as such, they are strictly protected in Europe under the Habitat Directive Annex IV and the Bern Convention Annex III. In Britain, populations of *M. avellanarius* were once widespread, but over the last 100 years they have become extinct in up to seven English counties, highlighting the detrimental effects of habitat fragmentation and loss [12]. The hazel dormouse then suffered a further 64% decline in numbers since the late 1970s, mainly attributed to unsympathetic woodland management, a series of bad breeding years and the increasing isolation of populations through fragmentation [12][7]. As a result, they are considered to be of conservation concern and are protected under the Wildlife and Countryside Act 1981, the conservation (Natural Habitats) regulations 1994 and are considered a priority under the UKBAP.

There have been a number of studies to date which have investigated the phylogeographic structure of the hazel dormouse both within its British and European range, the first of which was presented by Naim [75]. The study used two primer sets for the D-loop and cytochrome c oxidase subunit 1 (COI) regions of mitochondrial genome, which had been used in previous studies for the bank vole (*Myodes glareolus*) (D-loop) and field voles (*Microtus agrestis*) and tundra shrews (*Sorex tundrensis*) (COI); both markers gave corresponding results for the hazel dormouse for samples obtained throughout the British range (excluding the south east) as well as from Lithuania. Naim [75] presented the first evidence of both strong divergence between British and Mainland European dormice (approximately 0.7-1.5 MYA) and geographical partitioning of genetic variation for the British dormouse. Three British lineages were reported (bootstrap values > 70%) for 32 haplotypes, clustering into northern (Wales and Welsh English border populations), central (Worcestershire, Bedfordshire and Suffolk) and southern lineages (Somerset, Dorset, Devon and Cornwall). The northern lineage was further subdivided into three sub-lineages containing north Wales, central Wales and the Welsh/English border individuals. The southern lineage was further subdivided, clustering Somerset and Dorset together and Devon and Cornwall together. It was suggested that the observed geographic partitioning of the three main lineages could indicated the presence of evolutionary significant units (ESU). As such, it was advised that the observed genetic partitioning should be taken into account when planning and executing reintroduction schemes. It is worth noting however, that a simple clustering algorithm (neighbour-joining) was used to produce the phylogenies, as opposed to more computationally demanding Bayesian algorithms which are expected to better resolve the evolutionary lineages.
The most extensive phylogeographical study to date was based upon Cytochrome B (CytB) primers designed specifically for the hazel dormouse with samples taken from 28 locations throughout its Mainland European range (but excluding Britain) [4]. A complicated genetic structure was found for the hazel dormouse, with two highly divergent lineages in Europe, which were further subdivided into genetically and geographical determined sublineages. Lineage 1 contained individuals spread throughout Western Europe and Italy, whereas lineage two included those from Northern and central Europe, the Balkans and Turkey. The deep divergence between the two main lineages (approximately 7.7%) fell within the range of interspecific and intraspecific CytB distances observed in mammals and, as a result, the authors suggested the possibility that the two lineages represented possible subspecies, though this could not be confirmed, due to the use of only one mitochondrial marker. The high genetic divergence of all the lineages found were regarded to reflect the species’ low mobility.

Glass et al. [124] reanalysed the haplotypes published by Mouton et al. [4] with the inclusion of novel sequences sampled from Great Britain (Sussex, Isle of Wight, Devon and Wales), which placed the British dormouse in the central-northern European clade (CNE) and found that all British samples shared the same common haplotype (Hap15) which was found in Poland by Mouton et al. [4].

Combe et al. [5] concatenated the mtDNA markers (D-loop and Cytochrome B) and a nuclear marker to determine the likely post-glacial colonisation route and to further investigate the genetic structure of the British dormouse. Genetic variation was found at both mtDNA markers, with the most variation being reported for D-loop, whereas the nuclear marker was monomorphic for the UK. The results suggested a single colonisation event, most likely via the landbridge present up until approximately 7.5kya, which connected the UK to mainland Europe (Doggerland). On the basis of this, the authors believed it was reasonable to consider that the dormouse continued to enter Britain from mainland Europe across this land bridge until around 7.5kya, when Britain finally became inaccessible [311]. There probably followed a considerable period of optimal dormouse habitat throughout Britain until intensive farming and woodland destruction started to impact upon numbers and dispersal opportunities in the last century. Regional clustering similar to that reported by Naim [75] was reported and considered to be due to geographical features such as major rivers and that each cluster could represent an ESU. However, the statistical support for this claim is weak (most bootstrap values fall well below 70%). The authors suggested the possibility of local adaptation and suggested that patterns of genetic variation should
be taken into account for reintroduction schemes to prevent outbreeding depression and the loss of local adaptation.

Mouton et al. [126] expanded upon their previous work [4] by including more samples, further locations (including England) and by using two nuclear markers as well as the CytB marker. The two main lineages described in their previous paper was confirmed by the nuclear markers, with strong differentiation and no sharing of nuclear alleles between lineages. The further subdivision of the lineages differed when compared to CytB, which was attributed to the nuclear markers falling within genes, having a slower evolutionary rate and a longer coalescent time when compared with mtDNA markers. The two lineages could no longer be considered a single entity and represented partial ESUs, thus future conservation and management plans, such as reintroduction or breeding programs, were recommended to take into account the presence of these two genetic lineages.

When only considering the CytB marker, two haplotypes were obtained for the British dormouse (Hap15 and Hap51 (not currently available on Genbank)), with haplotype 15 being very widespread (Slovakia, Poland, Sweden, Denmark, eastern, central and northern Germany and England), implying a recent expansion of the hazel dormouse population in central-northern Europe. Mouton et al. [126] also suggested that post-glacial colonisation of *M. avellanarius* to Denmark and northern Germany occurred around 12,000BP following the extension of deciduous woodland promoted by warmer climate. They further supported the hypothesis of a single colonisation route via Doggerland from Europe to mainland Britain with the most likely origin of the British population being Denmark, given that the land bridge is believed to have connected Britain to Europe up to the Scandinavian regions during and shortly after the last ice age.

The post-glacial colonisation route into Britain via Doggerland and the suggestion of a single colonisation event appears to be well supported by both Combe et al. [5] and Mouton et al. [126]. The presence of genetic structure and differentiation throughout Europe is also clear, though the level of differentiation is dependent upon whether mtDNA or nuclear markers are assessed [126]. However, the level of differentiation between populations within Britain appears to be less well-defined, with conflicting conclusions between these two studies: Combe et al. [5] suggest the magnitude of differentiation within Britain is meaningful and should influence conservation strategy.
6.1.2  Genetic drift - gene flow equilibrium

The genetic diversity levels of populations are balanced through processes of allelic loss (through genetic drift) and allelic gain (through gene flow). Gene flow is expected to be greater between geographically proximate populations and, as a result, populations in close proximity should be more similar at neutral loci [319]. This relationship is known as isolation-by-distance and assumes a stepping stone model of gene flow, together with the sufficient passage of time for populations to have reached genetic drift - gene flow equilibrium [319]. As mitochondrial DNA is maternally inherited, it is assumed to represent only one quarter of the effective population size ($N_e$) of the nuclear genome. Due to the random sampling of gametes from this assumed smaller gene pool, mtDNA would be expected to reach drift - gene flow equilibrium over a shorter time period than that of nuclear markers and a significant isolation-by-distance relationship would be predicted to become established sooner for mtDNA markers. Under equilibrium conditions, gene flow offsets genetic drift at a rate proportional to the migration rate ($m$) and $N_e$ and hence pairwise $F_{ST}$ values are expected to increase with increasing distance. However, when $N_e$ is large, drift - gene flow equilibrium can take a long time to become established and measures of $F_{ST}$ are expected to correlate less well with geographic distance such that an isolation-by-distance relationship would not be expected. The isolation-by-distance relationship has been proposed as a means to assess whether a population is in equilibrium [319]: a population of randomly dispersing individuals will be expected to display isolation-by-distance at the mtDNA markers first and subsequently isolation-by-distance at the nuclear markers. Deviations from this would therefore be indicative of non-random migration patterns influencing the $N_e$ of the mtDNA markers. Previous research on the hazel dormouse has found isolation-by-distance patterns for microsatellite markers [75][50] but to our knowledge, the comparison with mtDNA markers has not been conducted.

6.1.3  Aims and objectives

The main purpose of this chapter is to reanalyse the current British dormouse phylogeny with the inclusion of newly acquired samples from the South East. Given that these new samples have been obtained from a region previously believed to have been colonised early on during the dormouse migration across Doggerland, it is anticipated that the result will throw some light onto the question of regional differentiation resulting from ancestral migration routes. Secondly, mtDNA sequences will be used to estimate $F_{ST}$ in order to assess whether there is an isolation-by-distance relationship for these markers.
In light of the previous research it is predicted that: 1) Britain will be significantly differentiated from mainland Europe 2) There will be some differentiation within the British populations i.e. it is expected that the Isle of Wight has been separated from the mainland for long enough to cause significant genetic differentiation. 3) There will be a significant and positive isolation - by - distance relationship found.

6.2 Methods and materials

Samples were collected from Sussex, the Isle of Wight (IoW) and the Wildwood Trust, Kent. All tissue samples were provided from animals that died of natural causes. Please see general methods for details of sample collection and DNA extraction.

6.2.1 Mitochondrial DNA

Mitochondrial DNA (mtDNA) is commonly used to establish population and species relationships by estimating the temporal sequence of shared ancestry [37]. The most extensive phylogeny for *M. avellanarius* was based on a 704 base pair (bp) region of the CytB gene in a study by Mouton et al. [4]. This data consists of 33 haplotypes located throughout the *M. avellanarius* range (see Table 1), excluding Britain. Previous work provided sequence data for CytB for areas within Sussex, IoW, Devon, Wales and Ireland [124]. Further sequence data for the British range was available from Combe et al. [5] for both CytB and D-loop, consisting of four and seven haplotypes respectively. New sequences were obtained from seven sites in Sussex, one site on the IoW and two samples were provided by Wildwood Trust, Herne Bay, Kent.

Mitochondrial DNA amplification

A 704 bp fragment of the CytB region of the mtDNA was amplified by PCR using primers: LMA14255, 5' -TGG TGG AAT TTC GGT TCT CT - 3'; RMA15192, 5' -GTT GCC TCC AAT TCA TGT T - 3' [4]. In addition a 449 bp fragment of the D-loop region was amplified by PCR using primers: M15997: 5' -TCC CCA CCA TCA GCA CCC AAA GC - 3' and H16401: 5' -TGG GCG GGT TGT TGG TTT CAC GG - 3' designed by Stacey et al. [320].

Amplifications were carried out in 10µl reactions containing < 10ng of DNA template. 0.25µM of each primer and 4µl of QIAGEN multiplex PCR mix. Cycling was performed using a DNA Engine Tetrad PTC-225 thermal cycler (MJ Research, Bio-Rad, Hemel Hempstead, Herts., UK), employing an initial denaturing step of 95°C
for 15 minutes, followed by 40 cycles of 30 - 45 s at 94°C; 30 - 45 s at 50°C and 45 - 90 s at 72°C), with a final extension at 72°C for 10 minutes for CytB. For the D-loop: 6 cycles of 95°C for 30 s; 54°C for 30 s; 72°C for 45 s; 36 cycles of 92°C for 30 s; 54°C for 30 s, 72°C for 55 s with a final extension of 10 minutes at 72°C. PCR products were then purified using ExoSAP (ThermoFisher) following the standard protocol provided by the manufacturers. DNA sequencing was performed using a BigDye V3.1 (Applied Biosystems) sequencing kit with approximately 20ng of cleaned PCR products with 3.3µM of primer (forward and reverse separately) in each reaction. Amplified products were then cleaned and loaded on an ABI 3730 48-well capillary DNA Analyser (Applied Biosystems, California, USA) following standard protocols.

6.2.2 Statistical analysis

Sequences were checked and aligned using CLC Sequence Viewer (Version 7.6.1, QIAGEN), where it was discovered that CytB haplotype 39 published in Combe et al. [5] was in fact identical to haplotype 15 published by Mouton et al. [4]. Checking also revealed that Combe et al. [5] had included both the forward and reverse primer sequence within the published D-loop sequences and, consequently, all seven sequences were trimmed from 506 bp to 449 bp to remove primer sequences and ensure quality for analysis.

MrBayes (version 3.2.2, [321]) was used to construct Bayesian Markov chain Monte Carlo (MCMC) phylogenetic trees in order to establish the evolutionary relationships among all haplotypes, for which we ran 50,000 generations, but otherwise the default priors. The edible dormouse (Glis glis) was used as an out-group, in keeping with the Mouton et al. study [4] and the Gliridae molecular phylogeny [322]. Minimum spanning networks were generated using Arlequin (ver. 3.5, [323] to calculate connection values, which were then input into HapStar (Teacher Griffiths, 2011).

Isolation-by-distance was calculated by correlating Euclidean distance measures and $F_{ST}$ scores between sites for all UK D-loop haplotypes. D-loop haplotypes were used as CytB was not variable enough. D-loop pairwise $F_{ST}$ values were calculated between haplotypes applying Tamura’s [324] model as implemented in Arlequin. Distance was measured using the straight line measuring tool in QGIS and was taken from and to the centre of the sampling points of each site. A Mantel test [301] was performed in the programme R [203] using the ADE4 package [302], which produced simulated p-values based on 9999 replicates.
6.3 Results

6.3.1 Phylogenetic trees

Most of the samples collected from Sussex and the Isle of Wight for the purpose of this study had identical CytB sequences (Hap15) which is widespread in Northern and central Europe (Poland, Slovakia, Sweden, Denmark, eastern, central and northern Germany) and previously reported to be present in England by both Glass et al. [124] and Mouton et al. [126]. Two novel haplotypes were established from an individual from Binsted Wood, West Sussex, and an individual from Worth Wood, West Sussex (see Table 6.1).

<table>
<thead>
<tr>
<th>Sample site</th>
<th>Geographic origin</th>
<th>N</th>
<th>CytB</th>
<th>D-Loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binsted</td>
<td>Sussex</td>
<td>5</td>
<td>Hap15, Binsted1</td>
<td>FN796767* Hap14,17 Pending</td>
</tr>
<tr>
<td>Wakehurst Place</td>
<td>Sussex</td>
<td>5</td>
<td>Hap15</td>
<td>FN 796767* Hap12,13,15,16 Pending</td>
</tr>
<tr>
<td>Tilgate Park</td>
<td>Sussex</td>
<td>5</td>
<td>Hap15</td>
<td>FN 796767* Hap13,14,18 Pending</td>
</tr>
<tr>
<td>Mallydams</td>
<td>Sussex</td>
<td>11</td>
<td>Hap15</td>
<td>FN 796767* Hap2,12 KU312933†</td>
</tr>
<tr>
<td>Worth</td>
<td>Sussex</td>
<td>2</td>
<td>Hap15, Worth1</td>
<td>FN 796767* Hap2 KU312933†</td>
</tr>
<tr>
<td>Bechill</td>
<td>Sussex</td>
<td>2</td>
<td>Hap15</td>
<td>FN 796767* Hap12,19 Pending</td>
</tr>
<tr>
<td>Flatropers</td>
<td>Sussex</td>
<td>1</td>
<td>Hap15</td>
<td>FN 796767* Hap2 KU312933†</td>
</tr>
<tr>
<td>Wadhurst</td>
<td>Sussex</td>
<td>1</td>
<td>Hap15</td>
<td>FN 796767* Hap2,12 KU312933†</td>
</tr>
<tr>
<td>Bridlesford</td>
<td>IoW</td>
<td>18</td>
<td>Hap15</td>
<td>FN 796767* Hap2,20 Pending</td>
</tr>
<tr>
<td>Hillbrook</td>
<td>Devon</td>
<td>2</td>
<td>Hap15</td>
<td>FN 796767* Hap2 KU312933†</td>
</tr>
<tr>
<td>Carmarthan</td>
<td>Wales</td>
<td>1</td>
<td>Hap15</td>
<td>FN 796767* Hap2 KU312933†</td>
</tr>
</tbody>
</table>

Table 6.1 Geographic location of sample locations, with number of samples obtained, haplotype number for each location and Genbank accession number, where relevant for both Cyt b and D-loop.
*Accession number from Mouton et al. [4]. †Accession number from Combe et al. [5].

Figure 6.1(a & b) shows that the Binsted Wood haplotype showed greater similarity to Hap15 than Worth Wood, with 69% support, as a result of the Worth haplotype showing greater similarity to the Irish haplotype (Ireland34) than either Hap15 or Binsted. Both Worth and Binsted are one mutational step from Hap 15. The CytB tree containing all haplotypes (Figure 6.2) had the same topography as the trees reported by Mouton et al. [4][126]. All dormice sampled from Britain fell within the Central-north European lineage with 84% support. For the Bridlesford site on the Isle of Wight, Hap15 was the only variant found by both Glass et al. [124] (n=5) and the current study (n=18). However, Combe et al. [5] reported finding a single novel haplotype (Hap40) within the five sampled individuals at the same site, which was two mutational steps from Hap15 (see Figure 6.3). The minimum spanning network (Figure 6.3) displays a star-shaped formation radiating out from Hap15.

A 449 bp fragment of the D-loop section of the mtDNA was amplified for 58 samples, which contained 14 polymorphic sites with a total of 11 haplotypes, ten of which were novel (see Table 6.1). Haplotype 2 (Hap2) (described by Combe et al. [5]) was found
Fig. 6.1 (a) Bayesian phylogenetic analysis of CytB haplotypes with support values. Geographic origins are followed by haplotype number; (b) detail of minimum spanning network showing the number of mutational steps between haplotypes.
Fig. 6.2 Bayesian phylogenetic analysis of CytB haplotypes with support values, including those from Mouton et al. [4] and Combe et al. [5]. Geographic origins are followed by haplotype number.
Fig. 6.3 Detail of minimum spanning network for the Central-Northern Europe clade and Turkey [4] showing the number of mutational steps between haplotypes.
in both East and West Sussex, on the Isle of Wight and in Carmarthen, Wales. Three of the ten novel haplotypes were found at more than one site (see Figure 6.4) and seven haplotypes were found within one area (Wakehurst Place, Binsted, Tilgate, Bexhill, IoW and Herne Bay) (see Figure 6.8). There does not appear to be a clear geographical partitioning at this scale (see Figure 6.4). The minimum spanning network (Figure 6.5) again revealed a star formation, where all haplotypes stemmed from Hap2 with between 1 and 3 mutational steps radiating out from the central haplotype.

The phylogeny for all available D-loop haplotypes (see Figure 6.6) grouped Ireland closer to Belgium, however no clear geographic partitioning at this scale could be found for the British and Lithuanian samples. The minimum spanning network (see Figure 6.7) also revealed a star formation, with all haplotypes radiating out from Hap2 with between 1 and 5 (Hap8 Suffolk2) mutational differences. The haplotypes found for Briddlesford Copse on the Isle of Wight were again different from Combe et al. [5], with those found for this study being Hap2 (two mutational steps from Hap4) and Hap20 (three mutational steps from Hap4).
Fig. 6.4 Bayesian phylogenetic analysis of D-loop haplotypes with support values. Geographic origins are followed by haplotype number, initials of the wood are used to represent the data collected for the current study: Wakehurst Place (WH), Tilgate Park (TP), Binsted (B) Mallydams (MW), Worth (W), Bexhill (BH) and Flatropers (FR) in Sussex, Isle of Wight (IoW), Herne Bay in Kent (HB) and Carmathen, Wales (CW).
Fig. 6.5 Detail of minimum spanning network for current study D-loop haplotypes showing the number mutational steps between haplotypes.
Fig. 6.6 Bayesian phylogenetic analysis of D-loop haplotypes with support values, including haplotypes from Combe et al. [5]. Haplotype number are followed by geographic origins, initials of the wood are used to represent the data collected for the current study: Wakehurst Place (WH), Tilgate Park (TP), Binsted (B) Mallydams (MW), Worth (W), Bexhill (BH) and Flatropers (FR) in Sussex, Isle of Wight (IoW), Herne Bay in Kent (HB) and Carmathan, Wales (CW)
Fig. 6.7 Detail of minimum spanning network for British and the Lithuanian (Hap5) D-loop haplotypes, showing the number of mutational steps between haplotypes.
Fig. 6.8 Regional distribution of D-loop haplotypes depicting the number of haplotypes found within each site. Initials represent site names: Briddlesford, Isle of Wight (IoW); Binsted wood, West Sussex (B); Tilgate park, West Sussex; Worth wood, west Sussex (W); Wakehurst Place, West Sussex (WH); Bexhill, East Sussex (BH); Flatropers, Rye, East Sussex (FR); Mallydams wood, East Sussex (MW) and Herne Bay, Kent (HB)
6.3 Results

6.3.2 Genetic Diversity

Low levels of haplotype diversity were found for the UK samples (see Table 6.2).

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>H</th>
<th>Hd (±SD)</th>
<th>π (±SD)</th>
<th>Taj D</th>
<th>Fu’s F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CytB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Sussex</td>
<td>25</td>
<td>3</td>
<td>0.2747 ± 0.1484</td>
<td>0.000406 ± 0.000515</td>
<td>-1.48074</td>
<td>-1.47534</td>
</tr>
<tr>
<td>East Sussex</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IoW</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UK 2017</td>
<td>52</td>
<td>3</td>
<td>0.1402 ± 0.0871</td>
<td>0.000203 ± 0.000336</td>
<td>-1.5106*</td>
<td>-2.26798*</td>
</tr>
<tr>
<td>UK all</td>
<td>109</td>
<td>7</td>
<td>0.581 ± 0.0354</td>
<td>0.001057 ± 0.000871</td>
<td>-0.64816</td>
<td>-0.95942</td>
</tr>
<tr>
<td><strong>D-loop</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Sussex</td>
<td>25</td>
<td>9</td>
<td>0.9 ± 0.0273</td>
<td>0.005822 ± 0.003589</td>
<td>-0.0686</td>
<td>-1.641</td>
</tr>
<tr>
<td>East Sussex</td>
<td>14</td>
<td>3</td>
<td>0.4725 ± 0.1358</td>
<td>0.00145 ± 0.001338</td>
<td>-0.9573</td>
<td>-0.0075</td>
</tr>
<tr>
<td>IoW</td>
<td>18</td>
<td>2</td>
<td>0.366 ± 0.1124</td>
<td>0.000817 ± 0.000920</td>
<td>0.488</td>
<td>0.796</td>
</tr>
<tr>
<td>UK 2017</td>
<td>58</td>
<td>11</td>
<td>0.8391 ± 0.0287</td>
<td>0.004369 ± 0.002784</td>
<td>-1.054</td>
<td>-2.580</td>
</tr>
<tr>
<td>UK all</td>
<td>109</td>
<td>19</td>
<td>0.8940 ± 0.0086</td>
<td>0.007265 ± 0.004161</td>
<td>-0.79361</td>
<td>-2.69053</td>
</tr>
</tbody>
</table>

Table 6.2 Genetic diversity measures for Cytochrome Bb and D-loop for the novel data for each of the main regions. UK 2017: all the novel samples; UK all: all sample from this study with the addition of those reported by Combe et al. [5]. Number of samples (n), the number of haplotypes (H), haplotype diversity (Hd) and nucleotide diversity (π), measures including standard deviations (SD), and neutrality test statistics (Tajima’s D (Taj D) and Fu’s F). Significant results (p<0.05) are depicted with *.

Hierarchical analysis of genetic variation (AMOVA) revealed more within population diversity than among population diversity for both CytB (% variation among = -0.01; % within = 100.01) and D-loop (% variation among = 35.69; % within = 64.31) (Tables 6.3 and 6.4).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>% Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>2</td>
<td>0.095</td>
<td>-0.00000 Vα</td>
<td>-0.01</td>
</tr>
<tr>
<td>Within populations</td>
<td>39</td>
<td>1.860</td>
<td>0.04768 Vb</td>
<td>100.01</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>1.955</td>
<td>0.04768</td>
<td></td>
</tr>
</tbody>
</table>

Fixation Index $F_{ST}$ : -0.00005
Significance tests (1023 permutations)


When including all UK haplotypes for the D-loop region of the mtDNA, a slight positive, but non-significant (p = 0.072), isolation-by-distance relationship was found between $F_{ST}$ and geographic distance (see Figure 6.9).
Fig. 6.9 Isolation-by-distance relationship between geographical distance (km) and pairwise $F_{ST}$ between haplotypes for all UK D-loop samples.


<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>% Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>2</td>
<td>7.070</td>
<td>0.17450 Va</td>
<td>35.69</td>
</tr>
<tr>
<td>Within populations</td>
<td>54</td>
<td>16.983</td>
<td>0.31449 Vb</td>
<td>64.31</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>24.053</td>
<td>0.48899</td>
<td></td>
</tr>
</tbody>
</table>

Fixation Index $F_{ST}: 0.35685$

Significance tests (1023 permutations)
$V_a$ and $F_{ST}$: $P(\text{rand. value} > \text{obs. value}) = 0.00000$
$P(\text{rand. value} = \text{obs. value}) = 0.00000$
p-value = 0.00000±0.00000

Table 6.4 D-loop AMOVA for Sussex and IoW

6.4 Discussion

The purpose of the current study was, in part, to clarify the ambiguous results of Combe et al. [5] through the inclusion of additional data in an attempt to better establish the phylogenetic structure of the UK dormouse. Contrary to the conclusions of previous studies [75] [5], the phylogenetic trees produced here did not show any clear UK lineages and thus geographic partitioning was not supported (see Figures 6.1; 6.2; 6.4; 6.6).

The Cytochrome B phylogeny showed a division between Turkey and the central-northern European lineage with a bootstrap support of 89%, but no clear division between haplotypes found in Britain, Germany, Poland, Czech Republic and Slovakia. The lack of any differentiation and the star formation of the minimum spanning network, with haplotypes radiating out from Hap15 (see Figure 6.3) backs up the finding of Mouton et al. [126] of a recent expansion in northern Europe and suggests a similar recent expansion occurred in Britain.

A similar result to that of the CytB was found for the D-loop data, showing a lineage between Belgium and all other haplotypes, with a bootstrap value of 98% (see Figure 6.6). Naim [75] reported clear phylogeographic partitioning of UK haplotypes with bootstrap support greater than 70% for the main clades of northwestern, central and southern lineages. However, the phylogeny presented by Naim [75] was a neighbour-joining tree, which uses a simple clustering algorithm rather than the computationally intensive Bayesian algorithms used to produce the phylogenies for the current study and may explain the differences between the two phylogenies.

Combe et al. [5] produced a concatenated phylogeny including CytB, D-loop and the $bfibr$ (nuclear marker) employing a maximum likelihood method, which, though
more powerful than the clustering algorithm used by Naim [75], it is not as powerful as the Bayesian algorithms utilised here [321]. Combe et al. [5] reported regional lineages likely due to geographic features (e.g. rivers) separating the regions in question, however the bootstrap support for these lineages was low (<70%) indicating that the claim for meaningful differentiation was not unambiguously supported. Both haplotype and nucleotide diversity was higher for the D-loop than CytB and although the overarching story provided by the markers was consistent, these differences again highlight the need to consider differing mutation rates across markers used. The D-loop region of the mtDNA is known to evolve particularly rapidly within some taxa [37], therefore it would be prudent not to concatenate the two mtDNA markers, lest subtle differences in population structure are masked - the rate at which markers reach drift-migration equilibrium will be influenced by the differing effective population sizes of these markers, which will impact isolation-by-distance estimates. The star formation for the minimum spanning networks for all UK haplotypes was observed for the D-loop (Figures 6.5 and 6.7) and provides further evidence for a recent expansion within the UK, rather than continuous migration or multiple colonisation events.

Of all the novel haplotypes discovered, two for CytB (Worth1; Binsted1) and seven for D-loop (Hap15 and Hap16 (Wakehurst Place); Hap17 (Binsted); Hap18 (Tilgate); Hap19 (Bexhill), Hap20 (IoW) and Hap21 (Herne Bay)) were each found within a single site. This could be due to limited sampling, however, as each phylogenetic study conducted on the hazel dormouse has detected new variants. Thus, the sampling of haplotypes within Britain appears by no means exhausted and further sampling would be required.

When considering genetic diversity for all UK samples, haplotype diversity was found to be moderate for CytB (Hd = 0.581 ± 0.0354) and high for D-loop (Hd = 0.8940 ± 0.0086), whereas nucleotide diversity was low for CytB (\(\pi = 0.0011 \pm 0.0009\)) and relatively high for D-loop (\(\pi = 0.0073 \pm 0.0042\)). Combe et al. [5] reported much higher levels of haplotype diversity for CytB (Hd = 0.728 ± SE 0.052) and higher (though still low) levels of nucleotide diversity (\(\pi = 0.0028 \pm SE 0.0002\)) than found here, with similar levels to those reported for the whole central northern European clade (Hd = 0.786 ± SD 0.096; \(\pi = 0.0034 \pm SD 0.007\)) [4].

Combe et al. [5] also found lower levels of haplotype diversity (Hd = 0.874 ± SE 0.035) but higher levels of nucleotide diversity (\(\pi = 0.0616 \pm SE 0.0071\)) with the D-loop than found in the current study, whereas Naim [75] found higher levels of haplotype and nucleotide diversity the Southern region described (Hd = 0.926 ± SD 0.022; \(\pi = 0.014 \pm SD 0\)), higher haplotype and similar nucleotide diversity for the
central region (Hd = 0.945 ± SD 0.013, π = 0.006 ± SD 0.001) and lower diversity levels in the northern region (Hd = 0.78 ± SE 0.081; π = 0.004 ± SD 0).

The rate at which a population loses genetic diversity through genetic drift depends on the effective population size and the level of gene flow [38]. A significant isolation-by-distance relationship would only be expected when a population has reached genetic drift - gene flow equilibrium [319]. A positive, but not significant pattern of isolation-by-distance was found when including all UK populations with a sample number greater than n = five (see Figure 6.9). The lack of significant relationship indicates that genetic drift - gene flow (migration) equilibrium has not yet been reached, which signifies a large effective population size. As mtDNA is uniparentally inherited (matrilinearly) and haploid, it is expected to represent only one quarter of the effective population size of nuclear markers [319]. A number of demographic factors can affect the relationship between genetic differentiation and geographic distance, including sex biased dispersal and mating strategy, which may affect population structure differently, depending on the mode of inheritance of the marker in question (i.e. uniparental or biparentally inherited) [325]. Thus, it would be prudent to compare the result for the D-loop to an isolation-by-distance relationship with a biparentally inherited marker (e.g. microsatellites) in order to establish whether the same pattern can be found, so that a more conclusive picture of movement patterns and population structure can be determined.

Previous research has suggested the assignment of evolutionary significant units to regions within the UK based upon mtDNA phylogeny trees; advising care when choosing genetic strains for reintroduction programmes, so as to prevent the loss of local adaption through outbreeding depression [75][5]. The current study does not support this claim, due to the lack of significant genetic partitioning for both markers employed. The presence of common variants found in multiple regions of the UK for both markers substantiates the migration of individuals, as the parallel evolution of these haplotype mutations is extremely improbable, if not impossible [326]. The lack of any significant isolation-by-distance relationship indicates genetic drift - gene flow equilibrium has not been reached, which signifies one of two things: firstly, that insufficient time has passed for the population to reach equilibrium, due to a large $N_e$; secondly that one reason for this is that levels of matrilineal gene flow are ample enough to inflate the $N_e$ of the mtDNA. Further comparisons with biparentally inherited markers (e.g. microsatellites) are required before firm conclusions can be made, but if the latter is true, the results further support the lack of evidence for distinctive gene pools within the British population.
6.4.1 Conclusions and implications

Given the lack of clear regional haplotypes, there is no real justification for attempting to maintain regional purity when considering reintroduction schemes, as suggested by both Naim [75] and Combe et al. [5]. These claims appear to be based upon a misunderstanding of how genetic variation is used to recognise evolutionary significant units. Mitochondrial DNA is useful for molecular phylogenies, as sequences are passed from one generation to the next unperturbed, due to the lack of recombination and, as such, the phylogenetic history is less noisy than with nuclear markers. Combining this with the relatively low mutation rate (when compared to neutral nuclear markers, such as microsatellites) makes teasing apart the evolutionary history of a population relatively straightforward.

For the appropriate designation of ESUs, the combination of both ecological data and genetic variation of adaptive significance are considered more relevant for conservation purposes than the partitioning of neutral variation [318]. Thus, to identify ESUs appropriately, local adaptations would first need to be identified. If such local adaptations are present in a population, then genetic divergence would be expected for the genes responsible for the adaptation. However, genes responsible for adaptations are rarely known, therefore proxy genetic markers are used. The use of such markers is fraught with difficulty, however, due to the degree of variation in terms of mutation rate and strength of selection. Consequently, the results obtained from phylogenetic analysis will differ from marker to marker. For example, microsatellites mutate more rapidly ($\approx 3 \times 10^{-4}$ for dinucleotides [327]) than nuclear genome on average ($\approx 1 \times 10^{-8}$), or the differences seen in this study between the results obtained from CytB and D-loop markers, where the mtDNA mutation rate is considered, on average, to be about ten times that of nuclear DNA [328]. Where local adaption is considered to be driven by nuclear genes, it may be reasonable to use a selection of nuclear genes as a proxy. However, if the result is that there is no diversity at all, as was found by Combe et al. [5], then in the absence of any ecological, behavioural or morphological data to the contrary, it must be concluded that there is no evidence for local adaptation and therefore no evidence for the existence of evolutionary significant units. When results obtained from variable markers are considered in favour of those from monomorphic markers, ascertainment bias arises and overestimate the magnitude of genetic differentiation [329].

The difference of opinion resulting from similar studies of the same population is a cause for reflection of the need to avoid the over-interpretation of genetic analysis. On the basis of this we consider that it would be more helpful if authors stated at
the outset of the study what the data supporting conservation measures and special management should look like (e.g., phylogenetic branches with bootstrap values > 0.8). Genetic data, although hard earned, does not always offer clear resolution and should be stated as such so as to avoid the misplacing of valuable resources.
Chapter 7

General Discussion

7.1 Introduction

The hazel dormouse is a nocturnal, habitat specialist and the recent decline in numbers means that they are considered a threatened species in Great Britain. The decline in numbers has led to a drive by conservation organisations to monitor, study and manage dormice populations (see section 1.1.2). A decline in numbers, though a worrying trend, does not necessarily inform on the evolutionary potential of a population and, as such, says little about their survival capabilities. If genetic variation observed at neutral markers can be considered to represent genome-wide variation, including genes influencing fitness, levels of neutral variation can highlight those populations most at risk of extinction. Maintaining genetic variation levels is, therefore, of prime importance to the conservation of species.

Due to the threatened status of the hazel dormouse, understanding life history traits and behaviours that may impact upon the amount of genetic variation present is paramount to the effective conservation of the species and the extent of social structuring and dispersal are intimately linked when it comes to the distribution and the maintenance of population genetic diversity (see Chapters 4 and 5). Sociality includes both cooperative and mating behaviours, but the degree to which sociality operates was largely unknown for the hazel dormouse and the available information could be conflicting.

The main aims and questions of the thesis were as follows: Establish whether the fermentation hypothesis is a plausible mechanism for kin discrimination. Do relatives share more similar bacterial communities than non-relatives? Establish whether dormice populations show social structuring based upon the spatial distribution of relatives and any inbreeding consequences. Do same sex relatives cluster together and opposite sex
relatives actively avoid one another? Quantify the magnitude of within and between population differentiation and estimate the impact of different habitat conditions on gene flow. Does the make-up of the environment influence the spatial distribution of genetic variation? Reanalyse dormouse phylogeny to determine the level of UK wide and European differentiation and assess whether the British population is in genetic drift-migration equilibrium. Is there any evidence for British population differentiation or an isolation-by-distance relationship?

7.2 Inbreeding avoidance and social structuring in the hazel dormouse

Inbreeding results in the increase of homozygosity, through the increase in alleles being identical by descent, with the consequent decrease in heterozygosity, which leads to an overall reduction in both population and individual levels of genetic variation. The social organisation of relatives into separate breeding demes can restrict gene flow, and lead to a reduction in the effective population size, allowing the increased importance of genetic drift and thus, a loss of genetic variation over time. Therefore, understanding the degree to which sociality operates and assessing the levels of inbreeding in endangered species, such as the hazel dormouse, are of prime importance to their continued survival. Both sexes were found to cluster spatially with relatives, but to a higher degree in females (see Chapter 4) and this relationship between geographic proximity and relatedness was found to be driven by the presence of close relatives \((r = 0.5)\), which indicated a high degree of tolerance between highly related individuals. The presence of both related and unrelated same sex individuals in close proximity is expected to encourage the evolution of kin directed behaviours, thus such clustering could be driven by kin cooperation. Proportions of adult relatives and non-relatives were reasonably high and relatively consistent both across years (Mallydams Wood) and sites (Mallydams and IoW) (see Chapter 4). The consistency of these population proportions at different sites with varying conditions implies that it may be the normal make up of populations and, as such, could be stable across populations of dormice. Degrees of philopatry varied between males, which lends support to the flexibility of dormouse social structuring being dependent on environmental conditions. The level of tolerance between same sex relatives could not be confirmed without further research into the degree of overlap between home ranges and as such, the likelihood that relatives cluster near to each other due to a lack of movement is equally plausible. However, if kin cooperation is in operation in the hazel dormouse, which has been suggested by other authors [75][13],
then the ability to discern kin from non-kin would be necessary when both related and
unrelated individuals live in close proximity to one another.

Inbreeding avoidance mechanisms include sex-biased dispersal, polyandry and
kin recognition (see Chapter 3 for detailed discussion). A general pattern of female
philopatry and male biased dispersal was found for the hazel dormouse, but this pattern
was not uniformly observed (see Chapter 4). Opposite sex relatives did not appear to
actively avoid or cluster together and nearest neighbour analysis revealed the presence
of opposite sex relatives ($r \geq 0.06$) at both sites studied (see Chapter 4). If neither
sex exclusively disperses, or remains in their natal site, and opposite sex relatives are
in close proximity to one another, then the likelihood of encountering relatives in a
random mating scenario would be high. The high probability of mating with a relative
by chance would result in inbreeding, leading to excess homozygosity and the loss of
genetic variation. Thus high levels of inbreeding (positive $F_{IS}$) and low levels of genetic
variation would be expected in dormouse populations. Significant excess homozygosity
was found at Briddlesford Copse on the Isle of Wight, however when investigated further,
the excess was attributed to the presence of multiple populations within the dispersing
sex (IoW) (see Chapter 5). Moderate levels of genetic variation were found throughout
the sample sites (see Chapter 5), which were similar to previous studies on both hazel
dormice [50][75] and edible dormice ($Glis glis$) [268]. The moderate levels of genetic
variation coupled with the lack of inbreeding at all sites in the presence of opposite sex
relatives implies the use of other inbreeding avoidance mechanisms, such as polyandry
and kin recognition. Previous research has revealed multiple paternity for litters of
dormice [53][50], however some of the sites showed zero inbreeding (Mallydams), thus
polyandry was determined not to be a strong enough mechanism to completely prevent
inbreeding.

The fermentation hypothesis provides a model of how bacteria contribute to recog-
nition cues in mammals, firstly through the production of volatile odourants within
the scent gland and secondly that the composition and structure of these scent gland
bacterial communities provides the inter-individual variation necessary for recognition.
If bacterial diversity were to provide such a mechanism, then bacterial community
profiles would be more similar in relatives than non-relatives. A positive and significant
correlation between genetic relatedness and anal bacterial community composition was
found in the hazel dormouse (see Chapter 3), suggesting that relatives share more
similar bacterial profiles than non-relatives. Therefore, the fermentation hypothesis
was regarded as a plausible mechanism for kin discrimination in the hazel dormouse
and to explain the lack of inbreeding and maintenance of genetic variation. For the
fermentation hypothesis to be employed, the bacteria would need to produce scent. The current study did not identify the specific bacterial species, which would be necessary to assess whether the bacteria responsible for the pattern were known odour producers. Future work employing next generation sequencing techniques on metagenomic samples, such as scent-gland swabs, would allow for the identification of the bacterial species and thus, whether or not they are likely responsible for odour.

7.3 The influences of potential barriers on levels of genetic variation and gene flow

The sites sampled for this study varied in patch size, structure (continuous to patchy), presence of potential barriers to gene flow (within and surrounding sites) and levels of connectivity to other woodlands (see Chapters 2 and 5). The population sampled on the IoW, provided a comparison for the levels of variation by being completely isolated from the mainland by the Solent Straight, an impermeable barrier to gene flow, compared to mainland populations surrounded by more permeable barriers, such as agricultural land. Due to the minimum distance between sites being 14.5km (between Tilgate and Wakehurst), each population could be treated as an independent site with regards to gene flow. Levels of genetic variation were expected to vary with the environmental conditions, with higher levels in large, continuous woodlands and lower levels in small, isolated woodlands, with the lowest diversity found within the island population. However, levels of expected heterozygosity and allelic richness did not differ significantly between sites (see Chapter 5), indicating that the differing conditions found at each site had no effect on these measures of diversity. Low levels of immigration were found in Mallydams wood, Wakehurst Place and Briddlesford Copse, which may have been sufficient to maintain diversity levels. Only the number of private alleles were variable enough to analyse and a significant, positive correlation was found between woodland size and the number of private alleles (when the IoW site was excluded); caution needed to be taken not to over emphasize this result however, as some of the alleles deemed private were found in other woodlands with too low sample numbers to include in the analysis.

Potential within site barriers, such as the reservoir at Wakehurst Place and the railway running through Briddlesford Copse, IoW, were expected to disrupt dormouse movements enough to impede gene flow and cause genetic sub-structuring. The low levels of sub-structuring found within Wakehurst Place could be attributed to the sampling of multiple generations and thus, the reservoir did not appear to impeded
gene flow. Genetic substructuring was detected within Briddlesford Copse, where two sub-populations were revealed. However, the sub-structuring could not be attributed to the railway line and in light of the findings at both sites it was concluded that the hazel dormouse is able to traverse a number of different structures that were hitherto considered to be potential barriers to gene flow (see Chapter 5).

The results presented in Chapter 5 indicate that dormice are less affected by fragmentation than previously thought. However, higher levels of male philopatry (as indicated by the higher levels of male relatedness) found in the more isolated site of Mallydams wood when compared to the more connected site on the IoW (see Chapter 4). The result was ascribed to Mallydams wood being more locally isolated than the IoW site, with fewer opportunities for dispersal and the increased cost of any such movement suggests disruption to social structuring. Fewer immigrants were also detected at Mallydams (0.8% of the population) compared to the IoW (3% of the population) which may also be due to the differing levels of local isolation (see Chapter 5). Dispersal from and to more isolated woodlands would have increased costs to the individuals, this may mean that only risk taking individuals migrate, whereas non-risk taking individuals may remain in their natal habitat. Thus, the consequences of fragmentation may only be apparent in changes to social structuring, which have not yet impacted upon diversity levels. The disruptions may not pose an immediate threat to the populations’ viability, but if the pattern was to be exacerbated by further fragmentation and isolation, then it is entirely possible that alterations on social structuring could lead to reduced variation in the future and potentially increase the vulnerability of the population.

The major drawback to this part of the study was the lack of site repeats, as each site had differing conditions and thus could not be considered as replicates. Future work including replicate sites would be needed to confirm whether habitat conditions have any effect on the levels of genetic variation and to determine if social structuring is disrupted due to habitat differences.

Microsatellite markers revealed high levels of differentiation between all the sites sampled for the study, indicating little gene flow between them (see Chapter 5). Mitochondrial markers (Cytochrome B and D-loop) on the other hand revealed a total lack of differentiation, both within Britain and mainland Europe (see Chapter 6). A common haplotype (Hap15) was found for cytochrome B and was present at every British sample site. This common haplotype is also found in northern Europe (Germany, Denmark and Poland). For the D-loop marker, phylogenetic trees revealed no significant differentiation between any of the British sites or from Lithuania. The
result was surprising, especially in light of the long separation between mainland Europe and Britain and between the mainland Britain and the IoW. Thus differentiation and gene flow appears to heavily depend on the scale that is being studied for the hazel dormouse.

7.4 Differences between microsatellite and mitochondrial markers

The loss of genetic variation through genetic drift will depend upon the effective population size ($N_e$) and the amount of gene flow. Levels of gene flow are expected to lessen with increasing distance, thus at neutral loci allele frequencies of populations in closer proximity will be more similar when compared to those at a greater distance. This isolation-by-distance relationship assumes a stepping stone model of migration (gene flow) and the passing of sufficient time to allow the levels of gene flow and genetic drift to reach equilibrium conditions (see Equation 7.1, where $N$ is the effective population size and $m$ is the migration rate). The isolation-by-distance relationship can, as a result, be used to assess whether populations are in equilibrium and whether they violate any of the assumptions of the model.

\[ F_{ST} = \frac{1}{1 + 4Nm} \]  

(7.1)

A significant positive isolation-by-distance relationship was found when using microsatellite markers for the Sussex and IoW populations (see Chapter 5), whereas no relationship was established when employing the D-loop mitochondrial marker for all sites (including those used by Combe et al [5] and those in this study) (see Chapter 6) and for the Sussex and IoW populations alone. Thus, when using microsatellite markers, it would appear that the populations are in drift-migration equilibrium, but not when using mitochondrial markers. Mitochondrial DNA (mtDNA) is maternally inherited (uniparentally) and haploid, thus only represents one quarter of total $N_e$ observed at diparental markers such as microsatellites [37]. At any given magnitude of migration, drift should be occurring at a greater rate for mtDNA than nuclear markers and we should therefore expect drift-migration equilibrium to be reached at mtDNA and to see isolation-by-distance before we do at nuclear markers (see Equation 7.2). As this was not the case, one or more of the underlying assumptions of the isolation-by-distance model must have been violated.
7.4 Differences between microsatellite and mitochondrial markers

\[ F_{ST} = \frac{1}{1 + Nm} \]  

(7.2)

Violations to the isolation-by-distance assumptions could explain the discrepancies between the two markers such as: differing levels of selective force on each marker; differing mutation rates and sex-biased dispersal. Both mitochondrial markers and microsatellites are selectively neutral and as the microsatellites used were believed to be unlinked to genes under selection, differential selective effects could be ruled out. The differences in mutation rate could account for the pattern seen, as the lower mutation rate for mtDNA compared to microsatellite DNA (see Chapter 1) would mean it would take longer for mutation-drift equilibrium to be reached (Equation 7.3, where \( \mu \) is the mutation rate), thus explaining the pattern. However, this scenario is considered unlikely given the fact that dormice have been present in England since the last ice age [5][126], thus experiencing several thousand generations.

\[ F_{ST} = \frac{1}{1 + 4N\mu} \]  

(7.3)

Polygynous mating can result in an inflated \( N_e \) for mtDNA compared to nuclear markers, which could give rise to drift-migration equilibrium being reached for nuclear markers before mtDNA markers. However, evidence of both polygyny and polyandry have been established for the hazel dormouse [53][34][13][11] and as such, dormice appear to display promiscuous mating strategies, thus an inflated \( N_e \) for mtDNA can be ruled out on this basis.

The most likely scenario to account for the discordance between microsatellites and mtDNA with respect to patterns of isolation-by-distance could be ascribed to differences in dispersal between males and females. Female-biased dispersal would have a disproportionate influence on mtDNA relative to microsatellites (due to the mode of inheritance). Thus, female migration would be expected to prolong the time it takes for the population to reach equilibrium, especially with respect to mtDNA, potentially explaining the isolation-by-distance relationship found for microsatellites and not for mtDNA [319]. Further support for this scenario comes from the fact that microsatellites revealed a strong pattern of genetic structure among sample sites (see Chapter 5), whereas phylogenetic analysis of both mtDNA markers indicated no significant differentiation between any British or European sample sites (see Chapter 6). Several haplotypes were found in multiple sample sites and as parallel mutation is all but impossible [326], this would also suggest that female-mediated gene flow was considerably higher than male dispersal, resulting in mtDNA displaying less structure.
than microsatellites. However, female migration would be expected to homogenise both mtDNA and microsatellite allele frequencies simultaneously. Piertney et al. [326] found that when male movement was minimal, but not absent (approximately 5%) the level of structuring for mtDNA would be less than nuclear markers and that this difference was largest when female movement was incomplete (approximately 25%). A similar situation could be happening with the dormouse, as we know that neither female nor male dispersal is dominant (see Chapters 4 and 5). Thus, it could be possible that female dispersal is sufficiently large enough to cause the differential levels of population structure found here. However, further investigation using a higher sample number and further sites, together with employing the models used in Piertney et al. [326] to identify the ecological factors that cause the observed differences in mtDNA and microsatellites would be necessary to confirm this. The disparities between mitochondrial and nuclear markers emphasises the caution that needs to be taken when inferring gene flow and population differentiation from only one type of marker, as the levels of structure and migration may be scale dependent.

7.5 Conclusions and implications

Despite the current conservation status of the dormouse in Britain, the results of this study provided little genetic evidence of an immediate threat to the persistence of the sampled populations. It must be taken into account however, that all the sites sampled for the present investigation were within the species stronghold and as such, past levels of genetic variation may have been higher than the moderate levels obtained here. It could be that loss of habitat, habitat fragmentation and the loss of connectivity may be having a negative effect on the genetics of dormice, but not enough time has past to detect it genetically. However, similar levels of heterozygosity were reported in other natural populations of hazel dormice in Wales (edge of range) [75] and Devon (species stronghold) [50]. All the sites included in the study are also part of the National Dormouse Monitoring Scheme and as such, they are either managed specifically to promote dormouse numbers or biodiversity in general, thus the levels of genetic variation found may be a reflection of the effectiveness of the conservation measures already in place.

The evidence presented in Chapters 3 and 4 allude to the possibility of sociality and kin discrimination behaviours in the hazel dormouse, thus potential cooperation between individuals. Same-sex clustering of relatives implies some level of tolerance for highly related individuals, especially in females. Coupling this with past documentation
7.5 Conclusions and implications

of crèching behaviour points to potential cooperative breeding behaviours. Such information could be important for captive breeding programmes and reintroductions. For instance, housing related females together in captivity may increase the number of offspring produced and litter success through the cooperative raising of young. Releasing female relatives together into reintroduction sites may also increase the likelihood of the population becoming established and persisting. Although female philopatry and male-biased dispersal was found in this study, the pattern was not always consistent, leading to the presence of opposite sex relatives in close proximity. As opposite sex relatives did not appear to actively avoid one another and there was no evidence for significant levels of inbreeding, the evolution of other inbreeding avoidance mechanisms, such as kin discrimination, is implied. Evidence of polyandry has also been found in wild populations of dormice. Thus, giving captive dormice access to multiple mating opportunities could improve the fitness of the offspring through mate selection and/or multiple paternity.

The population proportions of relatives were found to be similar at both Mallydams and the IoW sites (Tables 4.3 and 4.4), which indicates that such structuring is normal for dormouse populations and may be important for the maintenance of certain behaviours. Thus, releasing multiple related and unrelated individuals into reintroduction sites may help to mimic natural population conditions, which may also increase the probability of population permanency.

The evidence presented here for kin discrimination through bacterially mediated odour production might need to be taken into account when applying any potentially bacteria altering medications to captive dormice pre-release (such as antibiotics), to ensure normal behaviours are operating. The use of such mechanisms could also have wider implications in natural populations. Pollutants, such as pharmaceuticals and hormone mimicking chemicals entering the water system could potentially alter natural bacterial compositions in the scent glands and guts of species employing discriminatory behaviours through scent, thus modifying species social behaviours and mate selection abilities.

The results presented in Chapter 5 indicate the ability of dormice to navigate a number of environments and obstacles, thus they appear less vulnerable to fragmentation than previously thought. However, the higher levels of male philopatry at a more isolated site compared to a well-connected site could be a signal of population disruption through fragmentation. Although further work needs to be performed using replicate sample sites, efforts to increase the connectivity of woodlands through the
promotion of hedgerows and artificial bridges would be recommended to ensure the persistence of dormice populations.

As the reconstructed phylogeny in Chapter 6 provided no evidence for differentiation between British populations at mtDNA markers, there is no reason to assume local adaptation and the dormice in Britain should be considered one effective population. Until evidence to the contrary is presented, there is no strong argument to maintain regional purity when reintroducing dormice to the wild.

The findings of this study provide insights into social, inbreeding avoidance and dispersal behaviours of the hazel dormouse. These behaviours are intimately linked to the maintenance of genetic variation by influencing effective population size, as well as levels of inbreeding and gene flow. To our knowledge the research presented here is the first to investigate sociality through genetic methods, the outcomes of which could enhance our understanding of the elusive, threatened species and help inform their conservation.
References


References


References


References


Anderson EC and Dunham KK. The influence of family groups on inferences made with the program structure. *Molecular Ecology Resources*, 8:1219–1229, 2008.


