LANDSCAPE GENETICS OF HIGHLY DISTURBED ARABLE SYSTEMS: INSIGHTS GAINED FROM INVESTIGATING A SMALL MAMMAL SPECIES

Amanda Wilson

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2014

Landscape genetics of highly disturbed arable systems: insights gained from investigating a small mammal species

Amanda Wilson

University of St Andrews

A thesis submitted for the degree of PhD at the University of St Andrews

September 2014

Abstract

A large proportion of the earth's surface is dedicated to food production, and agriculture is widely acknowledged to influence local biodiversity via habitat loss and degradation. Landscape genetics is an emerging field which can provide detailed understanding of how wildlife populations are influenced by landscape configuration and composition but the approach is yet to be fully integrated with agroecology. When addressing landscape genetics questions, small mammals may provide insight; they may act as model organisms, they are abundant, they are relatively easy to sample and they may have important ecological roles within arable ecosystems. This thesis merged the study of arable landscapes, landscape genetics and small mammals, to develop what is known about the landscape genetics of wild species in this dynamic habitat type. To decide upon a study organism, small mammals were surveyed at an example arable field site. Wood mice (*Apodemus sylvaticus*) were found to be the most abundant species and a microsatellite marker multiplex was developed for genotyping individuals. Two aspects of their landscape genetics in arable habitat were investigated. First, the possibility of temporal patterns in fine scale genetic structure of arable populations was explored, since this had not been investigated previously. Next, inter-population genetic differentiation was examined to determine whether arable habitat acted as a barrier to gene flow for this species. At the fine scale, three genetically distinct clusters of wood mice were identified and temporal variation in the spatial pattern was confirmed. There was no evidence that arable habitat acted as a barrier to gene flow for this species in comparison to populations in urban habitat, which showed significant differentiation. It is hoped that the landscape genetic insights provided by this thesis will encourage greater momentum for conducting landscape genetics studies in agricultural habitat.

1. Candidate's declarations:

I, Amanda Wilson, hereby certify that this thesis, which is approximately 43,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2011 and as a candidate for the degree of Ph.D in Biology in September 2011; the higher study for which this is a record was carried out in the University of St Andrews between September 2011 and September 2014.

Date …… signature of candidate ……

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D in Biology in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …… signature of supervisor ……

3. Permission for publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date …… Signature of candidate ……

Date …… Signature of supervisor ……

Acknowledgements

I wish to thank...

My primary PhD supervisors: Graham Begg, Brian Fenton and Steve Hubbard for all they have done to make my PhD project possible; from writing the initial PhD project proposal, to obtaining funding and for supporting me throughout the project.

My secondary PhD supervisors: Brian Boag, Gaynor Malloch, Anne Magurran and Dave Parish for providing additional assistance.

The organisations that funded my PhD project: The Natural Environment Research Council (NERC), The James Hutton Institute, The Game and Wildlife Conservation Trust (GWCT) and NFU Mutual.

The student helpers for assistance with field work: Shona Jack, Rebecca Sargeant, Jennifer Rowland, Karim Hussain, Sarah King, Nick Argiropoulos, Hazel Brown, Vicky Stavropoulou and Bethany Westwood.

The staff managing and maintaining the Centre for Sustainable Cropping (CSC), where a large portion of my field work was carried out: Cathy Hawes for managing the CSC with the assistance of Euan Caldwell, John Bennett and Stuart Wale, and Linda Nell, Paul Neave and Mark Young who carried out the baseline crop and weed sampling.

James Hutton Institute staff who contributed: Craig Simpson and Pete Goddard from my JHI thesis committee, the sequencing service staff Clare Booth, Louise Donnelly who operated the ABI genotyper throughout my PhD project and the IT, HR staff, workshop and reception staff.

My thesis committee at The University of St Andrews: Pat Wilmer and Neil Hazon.

All the many others who provided feedback at conferences, at work and outside of work.

CONTENTS

Chapter 1

General introduction

1.1 Agricultural wildlife and biodiversity

For several centuries humans have been interacting with wildlife to the extent that they have been considered the agents of the sixth mass extinction event (Brook *et al.*, 2008). Of the various ways in which humans negatively interact with wildlife, habitat loss and degradation due to agriculture may rival the effect of climate change (Tilman *et al.,* 2001). A total of 40% of the land surface is now dedicated to agricultural production and, in future, this is likely to rise alongside the predicted human population increase (Foley *et al.,* 2005; Seto *et al.,* 2011).

Global objectives have been set to halt biodiversity decline for a variety of reasons (Convention on Biological Diversity, 2014). The most fundamental motivation suggested for the conservation of biodiversity is that each species has some intrinsic value which should be preserved (Justus *et al.,* 2009). Alongside this is the suggestion that many species could have unidentified but critical roles within ecosystems or future ecosystems and therefore, species should be protected as a precaution to mitigate against ecosystem collapse (Aronson and Precht, 2006). There are also some convincing examples of particular species or ecosystems which provide a service to humanity that can be valued in monetary terms, alongside a few scarce examples where biodiversity itself correlates with the level of service provision (James *et al.,* 2001; Isbell *et al.,* 2011). To achieve goals to conserve biodiversity, the human population is required to trade-off the need to provide food, water and shelter against the preservation of biodiversity and the services it can provide (Foley *et al.,* 2005).

The Convention on Biological Diversity (CBD) is the global organisation responsible for the conservation of biodiversity, sustainable use of biodiversity, and fair and equitable sharing of the benefits that biodiversity brings (Convention on Biological Diversity, 2014). The CBD was founded in 1992, and its most recent update was made in Nagoya in 2010 (Convention on Biological Diversity, 2014). The organisation relies on worldwide parties ratifying its treaty, which then becomes legally binding for the signatories. The European Union is one party that has approved the treaty, and mandatory elements of developed protocols are presently being implemented by the European Commission (European Commission, 2014a) and throughout its member states, including the United Kingdom. There are therefore legal obligations to consider and accommodate biodiversity within agricultural habitat.

The European Commission reforms to the Common Agricultural Policy are the main agent by which CBD objectives are being met in the agricultural sector. The CAP reform, adopted in

2013, will result in a move away from paying land owners for crop yield to paying for environmentally-sensitive practices that are intended to benefit the environment (European Commission, 2014b). In England this is likely to result in 30% of landowners' direct payments from government depending upon greening measures, which will include crop diversification, retaining permanent grassland and the creation of ecological focus areas for wildlife (DEFRA, 2014). In Scotland, similar greening payments will be offered (The Scottish Government, 2014).

Prior to the CAP reforms, landowners throughout Britain were subsidised for voluntarily implementing agri-environment schemes. There was a wide range of possible management options that would be subsidised, including management of hedgerows and ditches, buffer strips along water courses, enhanced grass buffer strips, maintenance of species rich grassland and reduced fertiliser application (Natural England, 2013). The success of these measures for accommodating wildlife was not always tested but reviews of studies that appraised a selection of the schemes suggested that although schemes sometimes improved some biodiversity indices, benefits were not all-inclusive (Kleijn and Sutherland, 2003; Kleijn *et al.,* 2006; Whittingham, 2007; Pretty *et al.*, 2010).

Since the term 'agroecology' was coined in the 1930s, and reinforced more recently in the 1970s, the accumulation of the knowledge about the ecology of food production systems has been increasing (Harper, 1974; Dalgaard *et al.,* 2003). More recently the driving force for agroecology has been the need to develop and appraise sustainable farming methods, which is best achieved when the interaction between agriculture and wildlife has been understood comprehensively (Dalgaard *et al.,* 2003; Chappell and LaValle, 2011).

1.2 Landscape genetics of agricultural systems

Landscape genetics is a more recently emerging field than agroecology, with the term being coined by Manel *et al.* in 2003. This research area seeks to address how landscape configuration shapes population genetic structure (Manel *et al.,* 2003; Storfer *et al.,* 2007). Most often landscape genetic studies aim to capture a genetic snapshot across a landscape, and correlate discontinuities in genetic variation with environmental data, the aim being to identify how a species is influenced by features within the landscape (Manel *et al.,* 2003). There is scope to combine landscape genetics and agroecology more fully, in order to improve understanding of the ecology of wild species within an agricultural matrix.

Landscape genetics has gained various definitions. At the broadest end of the spectrum, the term has been used for studies which consider species' genetics within landscapes, commonly at finer spatial and temporal scales than phylogeography studies (Manel *et al.,* 2003). A stricter definition was suggested by Storfer *et al.* (2007) who restricted the use of the term to apply to studies that considered and quantified the influence of landscape composition, configuration and matrix quality on spatial genetic variation. Landscape genetics can be distinguished from phylogeography because it focuses on contemporary processes influencing genetic variation, whereas phylogeography primarily considers historical processes that acted on variation (Wang, 2010).

The insight gained from landscape genetics has been used for a variety of purposes. For example, genetic information has helped to delineate management units, to investigate landscape connectivity, to provide insight into metapopulation dynamics, in studies of speciation, to investigate the efficacy of biological wildlife corridors, to investigate barriers to gene flow and more generally to improve the understanding of wildlife populations within changing landscapes (Manel *et al.,* 2003; Storfer *et al.,* 2007). Storfer *et al.* (2007) categorised landscape genetics studies from the past decade into five main research areas: the influence of landscape variables and configuration on genetic variation, identifying barriers to gene flow, examining source–sink dynamics, understanding the spatio-temporal scale of ecological processes and testing species-specific hypotheses.

In general, few landscape genetic studies have considered agricultural habitat (Gauffre *et al.,* 2008; Chambers and Garant, 2010). A review recently reported that only 11% of landscape genetic studies included agricultural habitats, and that studies in freshwater, meadow/shrub and temperate forest were more common (Storfer *et al.*, 2010). From traditional ecological studies in agricultural habitat, it is possible to make predictions about how genetic variation might vary across space and time in this habitat type, but examples are scarce. For reference, Table 1.1 details examples of landscape genetic studies that did investigate genetic variation of a wild species, and made reference to the influence of agricultural habitat. The majority of studies appraised quantitatively or, more often, qualitatively, the extent to which agricultural habitat acts as a barrier to gene flow, but results were mixed. Some studies reported that agricultural habitat acted as a barrier to gene flow (for example, Cegelski *et al.,* 2003; Coulon *et al.,* 2004; Lindsay *et al.,* 2008), but for other species and studies, gene flow did not appear to be impeded (for example, Johansson *et al.,* 2005; Purrenhage *et al.,* 2009; Schmidt *et al.,*

2009). Arable habitat has been suggested to be the most extreme example of a habitat that varies in space and time, and this could have potentially interesting consequences for the genetic variation of wild species (Gauffre *et al.,* 2008). No published studies have examined the dynamic aspect of a possible agricultural influence, which would perhaps be evident if genetic variation was examined over time (Storfer *et al.*, 2010).

There is substantial scope to develop the agricultural landscape genetics literature in terms of all five of the key research areas identified by Storfer *et al.* (2007) to obtain a clearer understanding of the landscape genetic patterns of wild species in this habitat type.

Table 1.1 Table detailing landscape genetics studies within an agricultural context.

Species	Inference	Citation
Greater Rhea, *Rhea americana*	Four populations in agricultural settings showed greater genetic differentiation and lower genetic diversity than predicted considering population sizes and reproductive success.	Bouzat, 2001
A land snail, *Helix aspersa*	Landscape metrics that best explained genetic distances between sites included those that suggested migration occurred preferentially along roadside verges, canal embankments and hedges in the agricultural matrix.	Arnaud, 2003
Montana wolverines, *Gulo gulo*	Significant population substructure detected, which was suggested to be caused by intolerance of roads, urban settlements and agricultural habitat.	Cegelski *et al.*, 2003
European roe deer, *Capreolus capreolus*	Gene flow occurred preferentially along wooded corridors within agricultural matrix.	Coulon *et al.*, 2004
Common frog, *Rana temporaria*	Effect of agricultural intensity on genetic diversity was positive and negative depending on the region. Roads had a negative effect on genetic diversity but ditches had a positive effect.	Johansson *et al.*, 2005
A brachypterous carabid beetle, *Carabus auratus*	Demonstrated genetic differentiation within agricultural landscape and reported higher genetic diversity in areas with more grassland.	Sander *et al.*, 2006
Common voles, *Microtus arvalis*	No genetic differentiation between populations sampled over a 500 km^2 agricultural landscape.	Gauffre *et al.*, 2008

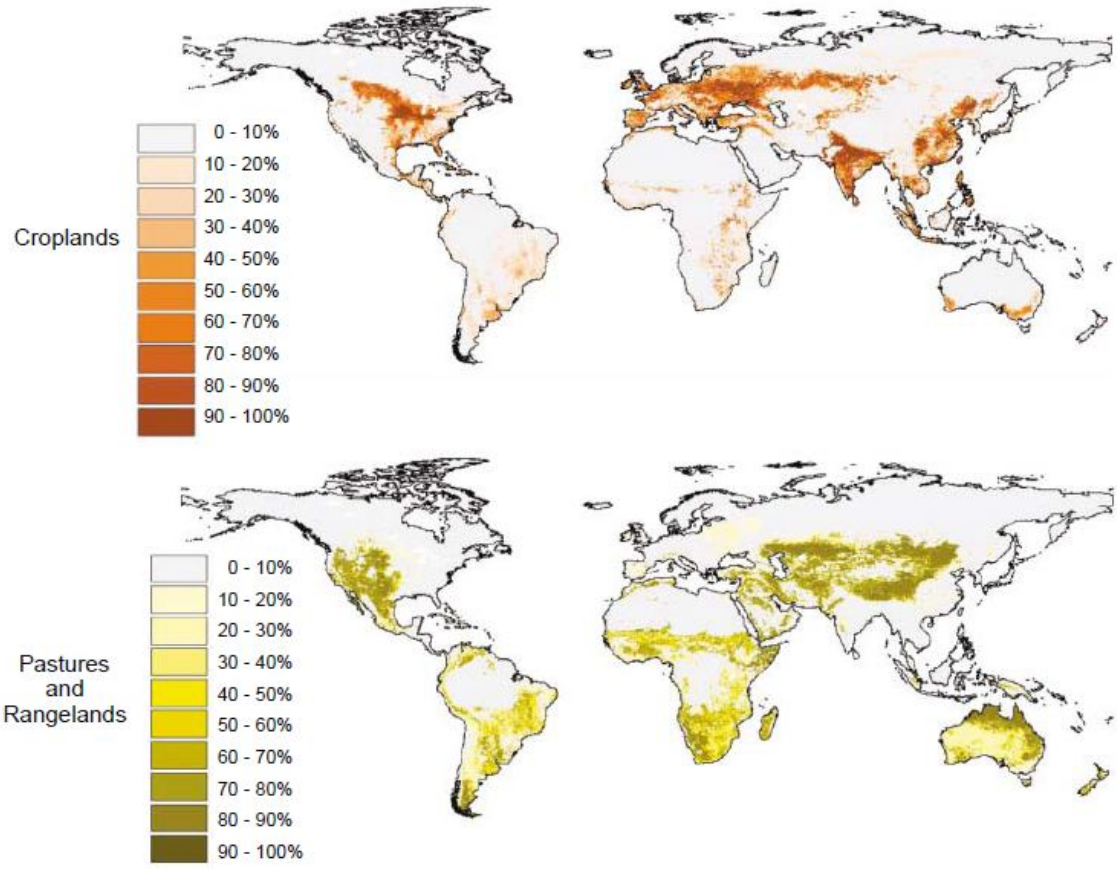
Yellow-footed antechinus, *Antechinus flavipes*	Individuals sampled from adjacent woodland and farmland locations fell within the same genetic cluster but cleared land, mainly for farming, was suggested to reduce connectivity due to results of a cost-analysis.	Lada *et al.*, 2008
Golden-cheeked warbler, *Dendroica chrysoparia*	Genetic differentiation between seven sites was positively associated with the amount of agricultural habitat, which appeared to hinder gene flow.	Lindsay *et al.,* 2008
Black-capped vireos, *Vireo atricapilla*	Significant genetic differentiation between 12 sites. No correlation between genetic differentiation and landscape attributes (agricultural habitat within unsuitable habitat category).	Barr *et al.,* 2008
Spotted salamanders, *Ambystoma maculatum*	Connectivity of the landscape appeared to be high in area hypothesised to be fragmented due to agricultural and urban habitat.	Purrenhage *et al.,* 2009
Wood avens, *Geum urbanum*	Inter-population differentiation did not differ for three landscapes differing in agricultural intensity.	Schmidt *et al.,* 2009
Columbia spotted frogs, *Rana luteiventri*; long-toed salamanders, *Ambystoma macrodactylum*	Urban and rural developed land provided greatest resistance to gene flow for both species. For Columbia spotted frogs, scrubland and agricultural land provided least resistance and for long-toed salamanders, forest habitat provided least resistance.	Goldberg and Waits, 2010
A quillwort sp., *Isoëtes malinverniana*	Genetic differentiation between seven sites suggested to be recent and possibly due to agricultural intensification.	Gentili *et al.,* 2010
Alpine newt, *Mesotriton alpestris*	Genetic differentiation was negatively related to amount of agricultural habitat but other landscape variables were more important (amount of forest, urban and orchard).	Emaresi *et al.,* 2011

Drylands vesper mouse, *Calomys musculinus*	Lower genetic differentiation between individuals sampled within arable field site and within an adjacent continuous secondary road compared to urban habitat.	Chiappero *et al.,* 2011
Pygmy mouse, *Baiomys musculus*	Genetic differentiation between four tropical dry forests sampled but no relationship between genetic diversity and amount of surrounding cultivated land.	Vargas *et al.,* 2012
Shepherd's purse, *Capsella bursa-pastoris*	Genetic diversity correlated with cropping intensity.	Begg *et al.,* 2012
A damselfly sp., *Coenagrion mercuriale*	Open agricultural land improved gene flow. In contrast, elevational change, forest habitat and water-bodies impeded gene flow.	Keller *et al.,* 2012
Sierra Madre Sparrow, *Xenospiza baileyi*	Gene flow restricted perhaps due to the inability to cross unsuitable habitat, which included agricultural habitat.	Oliveras de Ita *et al.,* 2012
Wet grassland plant, *Lychnis flos-cuculi*	Gene flow between natural populations in agricultural matrix found to be moderate and higher than gene flow between natural populations and those sown using wildflower seed mixes.	Aavik *et al.,* 2012
Field vole, *Microtus agrestis*	Landscape structure (especially amount of unmanaged habitat) has a greater influence on gene flow and genetic diversity, than organic vs conventional farming.	Marchi *et al.*, 2013
White-footed mice, *Peromyscus leucopus*	Agricultural habitat provided least resistance to gene flow between 11 forest samples sites.	Marrotte *et al.*, 2014
Five bumblebee sp., *Bombus terrestris, B. lapidaries, B. pascuorum, B. hortorum, B. ruderatus*	Rarest species has low heterozygosity and allele richness. Low levels of spatial genetic structure, hypothesised to be due to widespread queen dispersal.	Dreier *et al.*, 2014

1.3 Arable and pastoral agriculture

Global agricultural habitat can be divided into two broad habitat types: arable and pastoral farmland. Arable farmland is habitat dedicated to the production of crops for human food or food for livestock. Pastoral farmland is used by grazing animals for milk or food production. Together they make up approximately 40% of the land surface (Foley *et al.,* 2005). Figure 1.1 shows the global distribution of arable and pastoral land during the 1990s (Foley *et al.,* 2005). The total global area used for arable farmland has increased by 12% during the past four decades (Foley *et al.,* 2005).

Figure 1.1 Figure taken from Foley *et al.* (2005), showing the distribution of arable and pastoral habitat during the 1990s.



Agricultural habitat is the majority habitat throughout Britain and in east Scotland, where the research for this thesis was carried out. In 2000, 75% of the land available in Britain was used for agriculture, and 35% was dedicated to arable agriculture (Robinson *et al.,* 2002). In Scotland, the majority of arable farming occurs in the east (in 1997, 94.4% of all arable

farmland in Scotland) and the main crops sown are wheat, barley, grass for silage or grazing, potatoes, oilseed rape and a small amount of oats, vegetables and soft fruit for human consumption (Scottish Agricultural Science Agency, 2000). In more recent times the uptake of winter crops (Autumn-sown) has increased and this has been implicated in the decline of several wildlife species that make use of overwinter stubble (Scottish Agricultural Science Agency, 2000). Since agricultural habitat makes up such a large proportion of the British landscape, any aims to manage, conserve or understand British biodiversity should consider this habitat type.

Arable and pastoral habitat can be compared qualitatively in terms of the level of stability offered to wildlife. Arable agriculture has been suggested to provide greater levels of habitat instability compared to pastoral because land must be ploughed, drilled and sprayed, and harvesting also alters the habitat (Fitzgibbon, 1997). For reference, the timeline in Figure 1.2 illustrates the management processes involved in arable agriculture alongside their approximate timings. The potentially dramatic changes in resource availability and crop cover provided by arable habitat within a single cropping season are illustrated by Figure 1.3.

Figure 1.2 Approximate timeline of processes involved in arable farming. Months are shown along the x-axis and processes by labelled lines. Timings are crop and weather dependent and will vary between farms and regions.

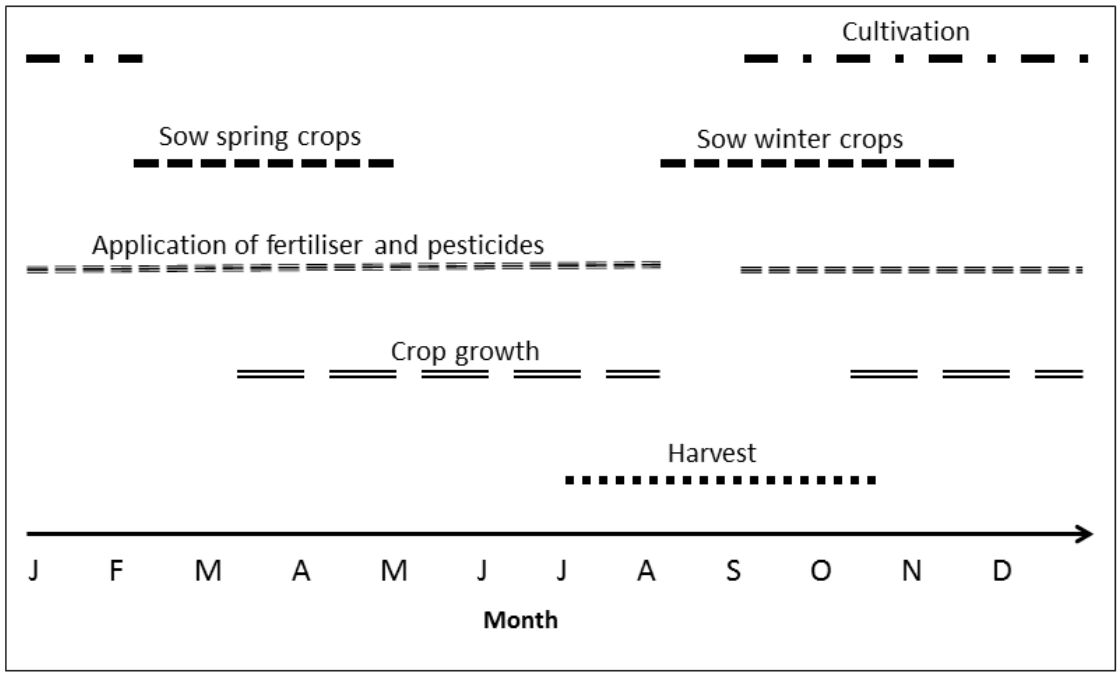


Figure 1.3 Images showing a wheat crop a) during the growing season and b) post-harvest.

a)

b)

This thesis will examine arable habitat since it is in the majority in the study region, and because the scope for spatio-temporal variation in species distribution and genetic variation is potentially greater.

1.4 Small mammals and arable agriculture

1.4.1 Small mammals as a model species

A model organism is one which can be used to gain broad insight into the general principles of a particular topic, where knowledge gained can be assumed to extend beyond the species in question (Hedges, 2002). For example, the laboratory mouse (*Mus musculus*) has been used as a model organism because it is small in size and it has a short generation time, making husbandry relatively easy (Hedges, 2002). Previously, in ecological studies of wild species, small mammals have also been referred to as model species because being midway in the food chain, their abundances can reveal something about the presumed abundances of species above and below them (Macdonald *et al.,* 2000). Additionally, they are vagile, meaning their abundance and distribution can be assumed to reflect responses to variation in habitat quality (Macdonald *et al.,* 2000). They are also relatively easy to sample, they are often common and abundant and they have short life cycles (Kikkawa, 1964; Gurnell, 1978) – other traits which make them suitable as a model organism with which to investigate ecological questions.

Despite this, care should be taken when extending the insight gained from studying one species of small mammal to other species, since their species biology and habitat requirements are never identical (Heroldová *et al.,* 2007). Perhaps, the most obvious generalisation that can

be made is that of generalist versus specialist, and within arable habitat, differences between these two categories could be large. Interspecies interactions may alter behaviour, abundance and distribution, but such interactions have proven difficult to research (Huitu *et al.,* 2004) and since species distributions can be patchy and concentrated around nesting sites (Ford, 1983), the outcome of species interactions may be locally variable. There is therefore sufficient reason to be cautious when extrapolating between species and when using the label 'model organism' in an ecological context.

1.4.2 The ecological role of small mammals in arable ecosystems

Aside from possibly providing broader indications about other wild species in agricultural habitat, small mammals may have important ecological roles themselves, and for that reason deserve landscape genetics research. There have been no known exclusion experiments in an agricultural context, which could improve confidence in the proposed ecological roles. This approach would be challenging since exclusion fencing is often costly, it must be dug into the soil to depths of greater than 25 cm, it may alter environmental conditions and exclusion zones must be checked regularly for invading individuals (Churchfield *et al.,* 1991). The ecological roles of small mammals in arable habitat have therefore been researched using alternative methods, with varying degrees of rigour.

Small mammals are well known to be a food source for raptors such as barn owls and species of harrier, and for rarer large mammals such as pine martens, wildcats and polecats (Salamolard *et al.,* 2000; Ryšavá-nováková, 2009; Rocha *et al.,* 2011; Zhou *et al.,* 2011). The species composition of the diets of these animals is less well known, but important if it is to be used as motivation for their research. For example, in one study, voles were 19 times more abundant in kestrel and buzzard pellets over winter than mice (Halle, 1988). In contrast, a study in Oxford ringed small mammals and recovered the rings in owl pellets, finding the majority of the rings recovered to be from wood mice rather than voles (Southern and Lowe, 1982). Addressing questions about the diet of birds of prey or large mammals is difficult; studies would need to be long term and obtaining a sample of independent animals requires large spatial scales to be examined.

Within arable systems, small mammals can become a pest species above certain densities because they consume crops (Brown *et al.,* 2007). The impact of common voles on mainland Europe, especially during population peaks, is well described and, in comparison to voles, the

impact of mice is thought to be insignificant (Zapletal *et al.,* 2001). However, wood mice have been recognised as a pest species and were suggested to be more of a pest during the ripening stage of grain production (Heroldová and Tkadlec, 2011). This is in line with the finding that when the diet of wood mice has been investigated previously, it has been demonstrated to include grains and plant material (Zapletal *et al.,* 2001). There is therefore scope for crop consumption. A recent study by Heroldová and Tkadlec (2011), compared the damage made by common voles (*Microtus arvalis*), wood mice and pygmy shrew (*Sorex minutus*) to culms of wheat in the laboratory, and reported that the feeding remains of the different species could be identified. This may provide more scope to compare the damage caused by the full species complement more fully.

Small mammals may also have roles within arable systems as seed or weed consumers (Westerman *et al.,* 2003). Baraibar *et al.* (2009) suggested that small mammals may play some role in the finding that 70–99% of arable weed seeds never germinate. If true, small mammals could potentially positively affect crop yields by controlling weed densities, making their ecological role more complex than simply one of crop consumption. The main method of researching seed consumption has been to fix seeds to cards and examine feeding remains or observe cards with cameras. In one seed card experiment using weed seeds, 32–70% of seed loss was due to consumption, with wood mice being the main seed consumer. Rodents were implicated in another study because of the number of droppings found around the seed cards (Marino *et al.,* 2005). In line with the finding that seeds are consumed when artificially attached to cards, wood mice were found to forage preferentially in patches of high weed density when radio-tracked (Tew *et al.,* 2000). When seeds were consumed, a preference for larger seeds was reported, suggesting that small mammals could also alter the weed species community (Westerman *et al.,* 2003).

There is evidence to suggest that, through weed grazing, small mammals could regulate weed communities in density-dependent and therefore, beneficial ways. For example, California voles (*Microtus californicus*) have been reported to selectively graze dominant grass species, which maintains open habitat and plant species richness (Batzli and Pitelka, 1971). Elkinton *et al.* (2004) also demonstrated a regulatory response to sunflower seeds (*Helianthus annuus*) by small mammals, whereby grazing increased as sunflower seed density increased. Again, regulatory roles and the shape of responses to plant availability by small mammals have received little research. However, it has also been suggested that small mammal grazing could

promote plant growth and regeneration and this could counteract the positive effects of weed grazing in arable systems (Fitzgibbon, 1997). An example is a study where grazing by voles (*Microtus oeconomus* and *Microtus middendorffi*) improved plant yield by promoting the growth of new shoots (Smirnov and Tokmakova, 1971). Similarly, rye grass (*Lolium perenne*) was found to regrow more quickly when grazed by hispid cotton rats (*Sigmodon hispidus*) in America (Howe *et al.,* 1982). The regulatory role of small mammals appears to have complexities that require additional research.

Small mammals may also act as seed dispersers by collecting seeds to cache, and by transporting seeds short distances attached to their fur. One study involved attaching seeds to the coats of yellow-necked mice (*Apodemus flavicollis)* by hand and observing the individuals until seeds were detached as far as 30 metres away (Kiviniemi and Telenius, 1998), demonstrating potential for seed to be transported. Wood mice have been reported to disperse fern spores through consumption, although more often ingested spores did not germinate (Arosa *et al.,* 2010). Through caching seeds, rodents may also act as dispersers, and there is some evidence to suggest that again this is seed size dependent (Vander Wall, 2003).

Small mammals may consume invertebrates, and therefore have scope to alter crop pest ecology or the ecology of predators of crop pests. Wood mice were found to reduce the abundance of Hymenopteran cocoons (*Diprion pini*) by up to 50% during winter (Obrtel *et al.,* 1978). A two year exclosure experiment in Berkshire grassland found that there were fewer large invertebrates outside small mammal exclusions, and this was attributed to their predation by shrew species (Churchfield *et al.,* 1991). Despite this reduction in invertebrate abundance, the species composition was not affected because shrews appeared to switch their feeding preference to the most abundant invertebrate food source (Churchfield *et al.,* 1991). Wood mice were demonstrated to include animal material in their diet during the summer months in a study in Oxford, where Lepidopteran larvae, beetle larvae, earthworms and centipedes were predated, as well as leaf-eating caterpillars during a caterpillar outbreak (Watts, 1968). In Watts (1968) study, bank voles (*Myodes glareolus*) included animal material as a smaller component of the diet than did wood mice and again primarily during summer months. It is likely that diet reflects the food types available in a habitat (Rogers and Gorman, 1995), and there is scope for small mammals to regulate invertebrate prey densities by including certain prey in their diet with a density-dependent response. In reality, few studies

have attempted to investigate the wide range of possible small mammal–invertebrate prey interactions (Elkinton *et al.,* 2004).

Small mammals may influence the concentration of nitrate in soil, affecting the cycling of nitrogen within ecosystems. In support, it has been estimated that the average kangaroo rat (*Dipodomys spectabilis*) burrow contains two kilograms of nitrogen (Greene and Reynard, 1932). In a study of shortgrass prairie, small mammals had a greater influence on nitrogen flux than any other vertebrate or ground invertebrate group (Woodmansee *et al.,* 1978). Clark *et al.* (2005) estimated the amount of faecal and urinary nitrogen produced by five small mammal species and found levels to be comparable to large mammals. They estimated a minimum of 1 kg N/ha/year and 2.7 kg N/ha/year was added from small mammal faeces and urine respectively, and to put this into context, the maximum amount of nitrogen permitted to be added in fertiliser applications in England is 150–220 kg N/ha/yr for cereal crops (DEFRA, 2013).

Small mammals have also been implicated as vectors of parasites and pathogens, such as cowpox, hantaviruses and *Toxoplasma gondii*, among others (Chantrey *et al.,* 1999; Essbauer *et al.,* 2006; Kijlstra *et al.,* 2008), providing further motivation for understanding their ecology within food systems.

Because rodent densities can be high, there is scope for any of their ecological roles to have large influences. For example, in urban scrub wood mice densities of 70–80/ha have been reported (Dickman and Doncaster, 1987) and bank vole densities reached 30/ha in urban woodland. In arable habitat, Macdonald *et al.* (2000) reported wood mouse densities of 7–26/ha in cropped habitat and during high common vole periods, densities of up to 400/ha have been reported for alfalfa grasslands (Lambin *et al.,* 2006). Population cycles have been reported for microtine rodents, and for hares (*Lepus*) and house mice (*Mus musculus*), meaning that densities can fluctuate dramatically and reach high peaks (Lambin *et al.,* 2006; Korpimäki *et al.,* 2004). The extent of the impact and roles of small mammals within agroecosystems remains to be comprehensively appraised.

1.4.3 Small mammals in arable habitat

Recent evidence has demonstrated that small mammal habitat usage could be influenced by both patch level factors and wider landscape factors (Brady *et al.,* 2011). In comparison to other groups, the influence of agriculture on small mammals has received relatively less

research compared to birds, plants and invertebrates, which have tended to be better documented (Hole *et al.,* 2005; Macdonald *et al.,* 2007).

As a field level factor within arable systems, there is evidence that wood mice prefer certain crops types over others. For example, a three year study in the Czech Republic reported that wood mice preferred alfalfa and maize at harvest rather than oilseed rape and barley (Janova *et al.,* 2011). In an earlier six year study in the Czech Republic, during which small mammals were captured at three month intervals, it was reported that the numbers of small mammal species captured varied between maize, wheat, barley, sugar beet and alfalfa plots (Heroldová *et al.,* 2007). Another study reported crop preferences, with fewer wood mice being captured in an oilseed rape plot compared to wheat and barley fields in all three years of an English study (Macdonald *et al.,* 2000). With respect to crop preferences, a consideration of a wide variety of crop types is lacking (Todd *et al.,* 2000), and studies are rarely long term despite small mammal communities fluctuating at that temporal scale (Heroldová *et al.,* 2007).

Another patch level factor to consider is the influence of non-cropped margins. Studies investigating the influence of non-cropped margins on small mammals have produced contradictory results. For example, a two year study in Yorkshire sampled in cereal, oilseed rape and bean fields, and showed no significant difference in wood mouse capture efficiency within grass margins compared to conventional field edges (Shore *et al.,* 2005) but, in contrast, bank voles and common shrews (*Sorex araneus*) preferred grassy margins over conventional field edges in Autumn months. Brown (1999) reported that small mammals were more abundant in set-aside margins of both organic and conventional fields in a six year study in Essex and in a five year study in Leicester. Further predictions of beneficial margin composition may be possible by extrapolating from work on the frequency of margin cutting, successional stages and underlying drivers of abundance patterns (Rogers and Gorman, 1995; Tew *et al.,* 2000; Tattersall *et al.,* 2001).

Another patch level factor that could be important is the influence of hedgerows within a patch. Amount and type of cover provided by the hedgerow have been tested as factors determining abundance. Voles were reported to be less abundant if there was grassy margin cover beneath the hedgerows (Smith *et al.*, 1993). In contrast, Smith *et al.* (1993) suggest hedgerow presence is important to wood mice rather than simply margin presence because they prefer cover to be at a canopy level, permitting free movement at ground level. In support, wood mice have been reported to prefer hedgerow to cut set-aside (Tattersall *et al.,*

2001). Dense cover at a canopy level was preferred in one study by both wood mice and bank voles, probably because this provides protection from aerial predators but does not impede movement and foraging (Boone and Tinklin, 1988). However in contrast, Bates and Harris (2009) compared organic and conventional hedgerows, organic hedgerows being larger in width, height and cross-sectional area, and found no significant difference in small mammal abundance or diversity. There is still considerable debate about the level of cover and type of cover that is preferred.

Several studies have confirmed the importance of wider landscape factors for small mammals. For example, a consideration of wood mice and bank voles on 38 farms, found that abundance was determined by the distance to the nearest wood, abundance of hedges nearby and the type of crops grown nearby (Fitzgibbon, 1997). Montgomery and Dowie (1993) also found the landscape matrix during winter to be an important variable affecting wood mouse abundance in field boundaries because this was determined by the distance to the nearest woodland. Furthermore, when winter and summer abundance was grouped, abundance also varied with amount of pasture in the surrounding area (Montgomery and Dowie, 1993). A recent study, reported with some uncertainty that landscape matrix factors were more important determinants of small mammal richness than patch level factors, such as habitat structure, level of patch disturbance and amount of patch resources (maximum of 12 species) (Brady *et al.,* 2011). Important matrix traits included man-made development intensity, landscape vegetation structure, presence of feral predators and resources available at landscape level (Brady *et al.,* 2011). It should be noted that the 95% confidence set of models included all four conceptual models proposed but the landscape matrix model best described the richness pattern observed (Brady *et al.,* 2011).

It is important to recognise a temporal aspect in habitat use patterns because underlying drivers, such as levels of cover and food availability, vary throughout the cropping cycle, especially in an agricultural context. For example, a study that monitored the abundance of 14 mammal species, found abundance to be greater in Spring/Summer than in Autumn/Winter within agricultural habitat (Heroldová *et al.,* 2007). Furthermore, they found that small mammal abundance was more stable in some crops across the seasons; for example, abundance within fallow and cereal was relatively stable compared to maize, sugar beet and alfalfa. Seasonal differences were also reported by Janova *et al.* (2011) who reported a significant effect of crop type on wood mice numbers captured during harvest but only a

borderline impact of crop type during spring. Studies exploring questions about wild species in arable agriculture should be specific about the time period being studied and, if possible, encompass multiple time periods.

1.5 Aims

As discussed above, landscape genetics can provide useful insight into the ecology of wild species and there is substantial scope to apply this approach more fully in studies of agricultural production systems. There have been relatively few landscape genetic studies carried out in an agricultural context and the main aim of this thesis was to examine the landscape genetics of a wild species in arable habitat, using a small mammal as the study organism. It is hoped that the work of this thesis will provide a basis from which future agricultural landscape genetic studies can be conducted.

To begin, in Chapter 2 small mammal species were surveyed at an arable study site in east Scotland and the most abundant species were determined. The broad habitat use strategies of the species were also characterised in this chapter, in order to explore whether future findings for the most abundant species might be more generally applicable to other species. The most abundant small mammal species, by a large proportion, was found to be the wood mouse and in Chapter 3, a molecular method for genetically fingerprinting this species was developed and appraised.

Recent landscape genetics reviews have emphasised the importance of carrying out sampling at appropriate spatial and temporal scales and agricultural habitat, with its frequent disturbance, is a habitat where temporal changes in genetic variation could be particularly relevant. However, temporal variation had not been previously considered by landscape geneticists working in agricultural habitat, and Chapter 4 aimed to update this. The fine scale population genetic structure of wood mice was examined at multiple time points throughout a cropping cycle and the possibility of spatial and temporal variation in genetic patterns was explored.

The role of agricultural habitat in creating barriers to gene flow for wild species has been previously researched. However, there have been inconsistencies between findings, with barriers being reported for some species but not for others. Additionally, many of the studies subjectively inferred the role of agricultural habitat in fragmenting populations but did not include control or comparison treatments. As such, in Chapter 5, a comparative approach was

taken in order to investigate the connectivity of arable compared to urban landscapes for wood mice. The aim was to address whether these habitats provided barriers to gene flow for the most abundant small mammal in the region. The work of this chapter also represented one of the first direct landscape genetic comparisons of arable and urban habitat: two of the most dramatically human-modified habitats available to wildlife.

Chapter 2

Arable habitat: a survey of small mammal species and their habitat use strategies

Manuscript submission:

This chapter contributed to an original research article, accepted for publication by the European Journal of Wildlife Research.

Chapter acknowledgements:

Small mammal data collection was assisted by Shona Jack, Rebecca Sargeant, Jennifer Rowland, Karim Hussain, Sarah King and Nick Argiropoulos. Fields and supporting crop and weed data were provided by the James Hutton Institute's Centre for Sustainable Cropping (CSC) at Balruddery Farm which is run by Cathy Hawes with the assistance from Euan Caldwell, John Bennett, Stuart Wale, Linda Nell, Paul Neave and Mark Young. Supplementary cropping data was provided by CSC staff.

Abstract

In this chapter, small mammal species of the order Rodentia were intensively surveyed at an example arable field site, with a view to choosing one species to examine using a landscape genetics approach. Arable farmland is an example of a frequently disturbed habitat, which can vary dramatically over space and time. As such, sampling was carried out during four sessions covering three key time points; early in the growing season, late in the growing season and post-harvest. The three rodent species caught were wood mice, bank voles and field voles, of which, only wood mice and bank voles were abundant. The habitat use strategies of these species were examined. Wood mice exhibited a habitat generalist strategy and made use of a variety of habitats. In contrast, bank voles made greater use of the more stable portions of the study site and exhibited greater habitat specialism. Because the study site was surveyed intensively it was possible to test whether there was continuity in the spatial distribution of individuals of both species. The distribution of wood mice varied with season but for bank voles it remained stable, possibly as a result of them specialising on stable areas within the habitat. The possibility of a species interaction was explored but there was no association, positive or negative, between the spatial distributions of the two species, suggesting that spatial competitive exclusion did not occur. Wood mice were found to be the most abundant species within the study site and were chosen as the focus of future landscape genetics chapters.

2.1 Introduction

There is variation in the small mammal species composition of arable habitat across Europe; species are present with different relative abundances and some species are absent from particular areas. For example, a study in the Czech Republic sampled 14 small mammal species, including shrews and three other rodent species that were not classified as small mammals per se (Heroldová *et al.*, 2007). The six rodent species captured most often included wood mice, bank voles, yellow-necked mice, house mice, field voles, bank voles and the pygmy field mouse (*Apodemus microps*) (Heroldová *et al.*, 2007). In contrast, a study in Western France found that wood mice and bank voles were most often captured, with field voles and common pine voles (*Microtus subterraneus*) being captured occasionally. In another alternative study in Germany , striped field mice (*Apodemus agrarius*) and common voles were the most frequently captured, but wood mice, yellow-necked mice and field voles were also

captured with some frequency (Fischer *et al.*, 2011). Substantial variation in the species composition of arable habitat across Europe is evident.

The research of this thesis will be limited to rodent species because, in Scotland, the intentional trapping of shrew species requires a Scottish Natural Heritage licence and humane traps must be checked every four hours (Scottish Natural Heritage, 2014), which would limit the number of traps that could be used in total. Within Great Britain, relative abundances of rodent species within the families Cricetidae and Muridae were estimated by Harris and Yalden (2008) but a level of uncertainty was associated with these expert estimates (Table 2.1). More relevant to this thesis, are studies carried out in agricultural landscapes but again slight variation was found in relative species abundances in British agricultural habitat. A study in southern England and Wales reported that four rodent species were well sampled, including yellow-necked mice, wood mice, bank voles and field voles (Bates and Harris, 2009). Further north in Britain, one study considering small mammal remains from owl pellets in an area of predominantly arable agriculture, reported that wood mice and bank voles were the most abundant but that field voles and harvest mice were occasionally a prey item (Askew *et al.*, 2007). In North Yorkshire, again wood mice and bank voles appeared most abundant and field voles were very rarely captured (Shore *et al.*, 2000).

Table 2.1 Rodent species in Britain. Table adapted from Harris and Yalden, 2008.

Species	Abundance estimate	Distribution
Field vole, *Microtus agrestis*	75,000,000	Widespread
Wood mouse, *Apodemus sylvaticus*	38,000,000	Widespread
Bank vole, *Myodes glareolus*	23,000,000	Widespread
Common rat, *Rattus norvegicus*	6,790,000	Widespread
House mouse, *Mus domesticus*	5,192,000	Widespread
Harvest mouse, *Micromys minutus*	1,425,000	England, south
Water vole, *Arvicola terrestris*	1,169,000	Widespread
Orkney and Guernsey vole, *Microtus arvalis*	1,000,000	Orkney and Guernsey
Yellow-necked mouse, *Apodemus flavicollis*	750,000	England, south
Ship rat, *Rattus rattus*	1,500	Scarce

The influence of arable agriculture on small mammal distributions and habitat use strategies has not been comprehensively researched. Limited small mammal research has been carried out at a regional scale in Europe, with work in France suggesting that different communities of small mammals are present under scenarios of low, medium and high intensification (Millán de

la Peña *et al.,* 2003; Michel *et al.,* 2006). When moving from the regional scale to farm scale, interpreting the literature becomes difficult, especially since exact habitat types are seldom replicated between studies and, when they are, there are sometimes conflicting results (for example, see Green 1979 and Janova *et al.,* 2011). Contributing to this uncertainty are regional climate differences, differences in habitats and differences in the complement of small mammal species between locations and the interactions between them, which have proven difficult to research (Huitu *et al.,* 2004). Further demonstrations of the range of strategies used by small mammals in agricultural habitat at a field scale are needed, especially in light of changing agricultural practices.

In the present chapter, an example arable study site was surveyed to determine the relative proportions of the small mammal species present and to identify the most abundant. The study site was typical of the land-sharing scenarios that are predicted to increase under the CAP Reform's aim to include 5% ecological focus areas within arable farms (European Commission, 2013). In this study site, frequently disturbed crop habitat was provided alongside more stable field margin habitat. The habitat use strategies of the most abundant species in the region were investigated at key time points in the cropping cycle and the study site was sampled intensively. This allowed the exact distribution of individuals throughout the cropping cycle to be considered, which provided context for future chapters. The extent to which the habitat use strategies of the most abundant species were similar was explored. Finally, to address ideas about spatial competitive exclusion, the presence of associations among the distributions of the most abundant species was tested.

2.2 Methods

2.2.1 Study area

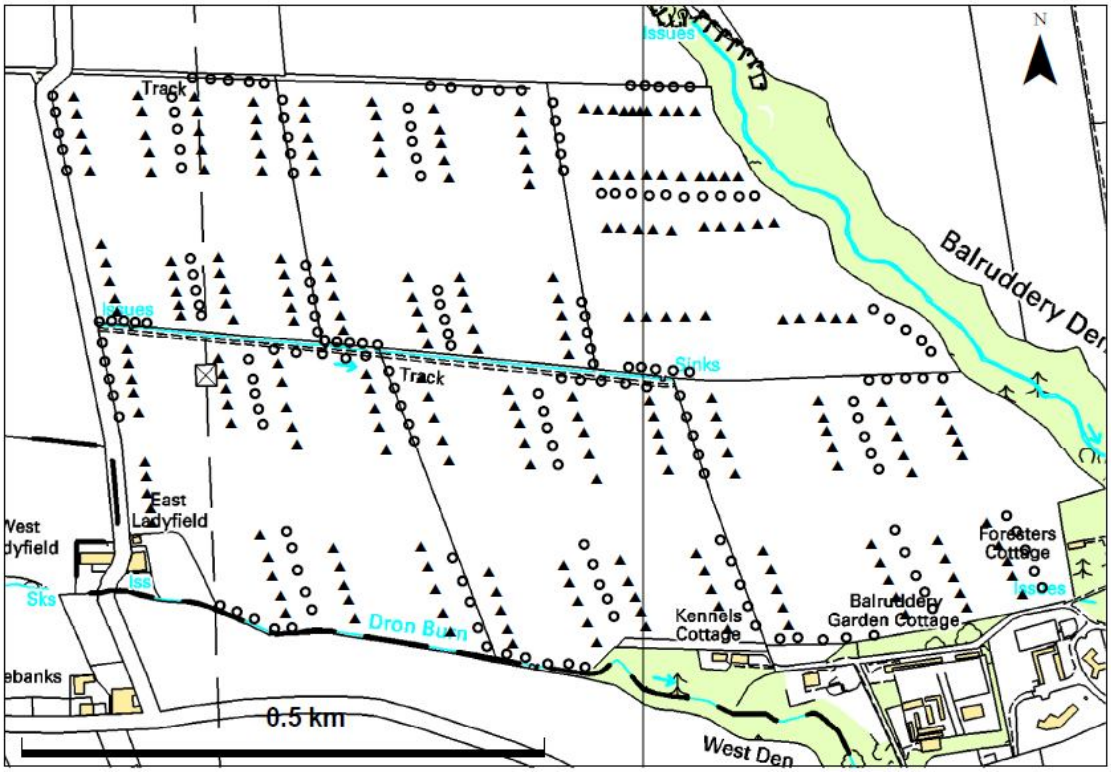
The example arable field site was at the Centre for Sustainable Cropping at Balruddery Farm (56°28'N, 3°07'W), a 170 ha arable farm in Tayside, Scotland, owned by the James Hutton Institute. The site comprised a six-field arable rotation, with naturally regenerated grass margins of equal width. Each field was divided in two, with a sustainable cropping system applied in one half and current conventional practice in the other. The two field halves were divided by a non-cropped margin, sown with a standard beetle bank grass mix. The conventional management system followed current standard management practice for the region in terms of cultivation, fertiliser, herbicide, pesticide and fungicide application. The

sustainable management system included practices such as leaving overwinter stubble, pesticide applications at threshold levels, reduced herbicide, reduced fertiliser, addition of municipal compost, precision farming and non-inversion tillage. The six crops included in the rotation were winter wheat, potato, spring barley, beans, winter barley and winter oilseed rape, and each crop is sown in five or six 18 m wide strips of different cultivars.

2.2.2 Small mammal sampling

Small mammals were sampled using Longworth traps (with shrew escape holes) filled with a mixture of peanut butter and porridge oats, along with hay for bedding. There were four trapping sessions intended to capture snapshots throughout an entire cropping cycle: early growing season 2012 (28th May – 4th July), late growing season 2012 (16th July – 2nd August), post-harvest 2012 (27th November – 20th December) and late growing season 2013 (1st July – 13th July). During each session 240 traps were laid in transects of five traps within the cropped habitat, spaced 20 metres apart, and 180 traps were laid in transects of five traps in the margin habitat spaced 20 metres apart, each for 3 consecutive nights. Eight transects were placed in each field with a regular arrangement; the first trap was placed 10 metres from the perpendicular field edge (Figure 2.1). Traps within the crop were placed 0.5 metres into the crop from tractor tramlines. Each bank or field vole captured was recorded alongside the trap position. Hair samples were taken for use in molecular analyses as described in Chapter 4 and the trap positions of each individual were recorded.

Figure 2.1 Trap layout map produced in ArcGIS 10.1. Black circles indicate Longworth trap positions in the margins and black triangles indicate traps within the crop. Traps were placed in eight transects within each field and six transects within each set of field margins, two of these within the margin dividing the field in half. © Crown Copyright/database right 2013. An Ordnance Survey/EDINA supplied service.



2.2.3 Crop sampling

In addition to the small mammal sampling, several characteristics of each crop were measured. Crop cover, weed cover and crop height were sampled during the early and late growing season in 2012 (19th-28th June, 17th July – 8th August) and during the late growing season in 2013 (22nd July – 12th August). The number of stems, stem weight and seed pod weight were sampled in the late growing season in 2012 (17th July – 8th August) and in the late growing season in 2013 (22nd July – 12th August). Twenty quadrats were placed at regular intervals in each field and percentage crop and weed cover were assessed subjectively. The total numbers of crop stems within the quadrat were counted and crop height readings were taken at three positions within the quadrat. To measure the stem and pod weight, all the crop stems within the quadrat were collected and the seed pods removed. The stems and seed pods were placed

in bags in a drying oven at 70°C for a minimum of 24 hours and the dry weight of each was recorded.

2.2.4 Handling of capture data

Further data analysis focused on wood mice and bank voles since these were the most abundant species (656 and 205 captures, respectively). Aside from 40 field vole captures, no other species were detected and there was an average capture success of 17.9%, indicating that trap saturation had not been reached.

Microsatellite marker data was available for wood mice from the work of Chapter 4, and any recaptures within a trap season could be removed from the dataset. Non-molecular marking methods for bank voles were ineffective but individuals were almost certainly recaptured. Marking methods trialled during a pilot study included marking with correction fluid (Tipp-Ex), sheep marker sprays and permanent marker pens. More sophisticated methods such PIT-tagging and fur-clipping would have required additional skilled field helpers.

Data were modelled as capture success, a proxy for levels of activity and abundance: total number of captures per total number of possible captures given the number of traps used and number of trap nights. This proxy was intended to be a measure of activity levels within the field site rather than a measure of population density, which was not a focus. For wood mice, capture success and unique individual capture success were both examined to obtain some insight into the effect of including recaptured individuals in the analysis.

2.2.5 Data analysis

To test whether season, habitat (crop or margin) and an interaction between season and habitat could explain capture success, logistic regression was used. Each season's data were analysed separately because an interaction between season and habitat was detected. The significance of habitat as a factor was tested with a likelihood ratio test (LRT).

To consider whether the arrangement of individuals within the study site remained stable throughout the study, the capture success per transect in one season was modelled against previous season's capture success per transect using logistic regression. The significance was tested with likelihood ratio tests (LRT). Transects were separated into crop and margin transects, to account for the possibility that crop and margin might differ in stability.

To investigate the possibility that competitive exclusion occurred, the extent to which wood mice and bank voles overlapped in their distribution was tested using logistic regression by including the capture success per transect of the other species in the model. The significance was tested with likelihood ratio tests (LRT). Again, crop and margin transects were separated. All seasons were tested separately, since differences in habitat use with season were observed.

The use of the cropped area by wood mice was considered further, and a preference for certain crop types was tested using wood mouse capture success per half field, modelled against crop type using logistic regression. This was repeated for all four trap seasons. The use of winter planted crops was compared to the use of spring planted crops using an exact binomial test for each season. Sustainably and conventionally managed field halves were compared using a paired t-test to control for differences in field types, after capture success per half field was transformed using an arcsin square root transformation, enabling the use of a paired t-test, for which there is no binomial equivalent.

Finally, because wood mice were most abundant, and because their habitat use strategy involved a greater use of the cropped habitat than bank voles, the relationships between several crop traits and wood mouse activity were considered: crop cover, weed cover, crop height, number of stems, stem weight and seed pod weight. Using linear regression, each crop trait was modelled with crop type as a factor. The mean and standard deviation of each trait were calculated for each crop – six field half replicates of each. As might be expected, these traits co-vary and for that reason they have not been modelled against wood mouse counts.

All analyses were performed using R 2.7.2 (R Development Core Team, 2014) and assumptions of deviance (goodness of fit), dispersion and random residuals were tested and judged to be satisfactory before continuing with each analysis. When appropriate, multiple testing was accommodated using a Bonferroni correction (Holm, 1979).

2.3 Results

2.3.1 Preparatory analysis

Here, 89% of samples were successfully genotyped and the $P_{(ID)sib}$ value for the nine microsatellite loci was adequately low ($3x10^{-5}$), suggesting individual wood mice could be confidently identified (Waits *et al.,* 2001). There was an error rate of 0.128 errors per allele,

mostly due to allele drop-out but by setting an appropriate level of mismatch in allelematch these errors can be accommodated (Galpern *et al.*, 2012; R Development Core Team, 2014). Removing recaptures resulted in 87 unique captures in the early growing season 2012, 84 in the late growing season, 104 post-harvest and 74 in the late growing season 2013 (32, 36, 30 and 29 recaptures removed respectively).

2.3.1 Habitat use

Wood mice and bank vole capture success varied with both season and habitat type (Figure 2.2 and Figure 2.3). Because of an interaction between habitat and season for each species (wood mice: LR χ^2= 108 (3.s.f.), df = 1679, p-value<0.0001; bank voles: LR χ^2= 46.2, df = 1679, p-value<0.0001), the data for each season was analysed separately. Wood mice demonstrated significantly greater use of the crop than the margins in the early and late growing season 2012 but used the crop significantly less than the margins during the post-harvest period (Table 2.2). In contrast, bank voles consistently demonstrated greater use of the margins than the crop areas in the early and post-harvest seasons 2012, when no bank vole individuals were captured within the crop (Table 2.2). The data for the late growing seasons showed the same pattern although the difference in capture success between the habitats was not significant (Table 2.2). It was noted that fewer small mammals were captured in 2013, possibly due to a long winter. The significance of the patterns above was verified as also being true for count data grouped by transect. The effect of including wood mouse recaptures was that the significance of the trends was slightly more pronounced – given the overwhelming magnitude of the effect for bank voles, it is likely that the overall conclusion would not be affected by accounting for recaptures.

Table 2.2 Habitat preference (crop versus margin) for both wood mice and bank voles (df = 419).

Season	**Wood mice**		**Bank voles**	
	LR χ^2	**p-value**	LR χ^2	**p-value**
Early growing season, 2012	14.4	<0.0001***	18.9	<0.0001***
Late growing season, 2012	17.6	<0.0001***	5.90	0.015
Post-harvest, 2012	51.8	<0.0001***	72.6	<0.0001***
Late growing season, 2013	2.60	0.104	2.30	0.128

Asterisks indicate significance level: p < 0.006 *, p < 0.001 **, p < 0.0001 ***

Figure 2.2 Total numbers of unique wood mice and bank voles captured in crop and margins during the four seasons. Crop data are displayed with black bars and margin data with white bars. Asterisks indicate significance level: $p < 0.006$ *, $p < 0.001$ **, $p < 0.0001$ ***

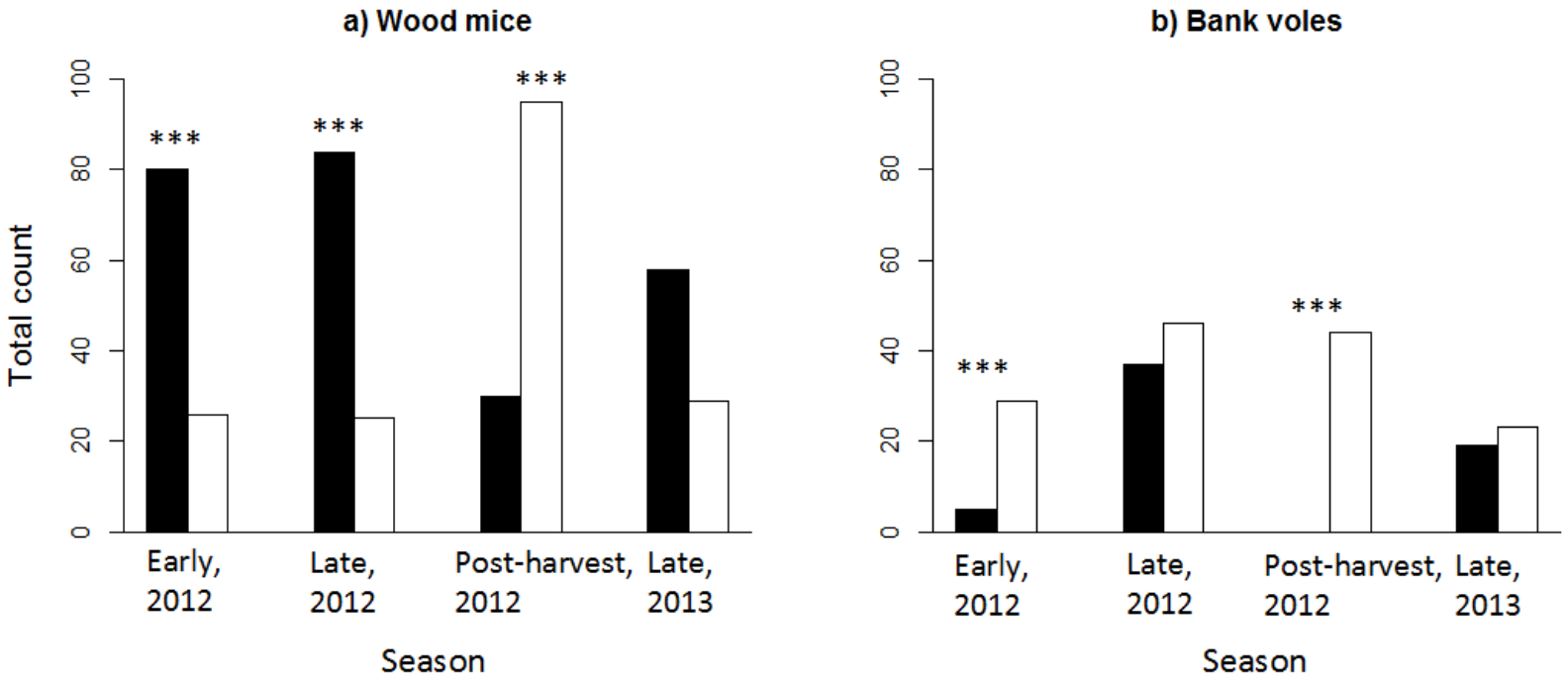
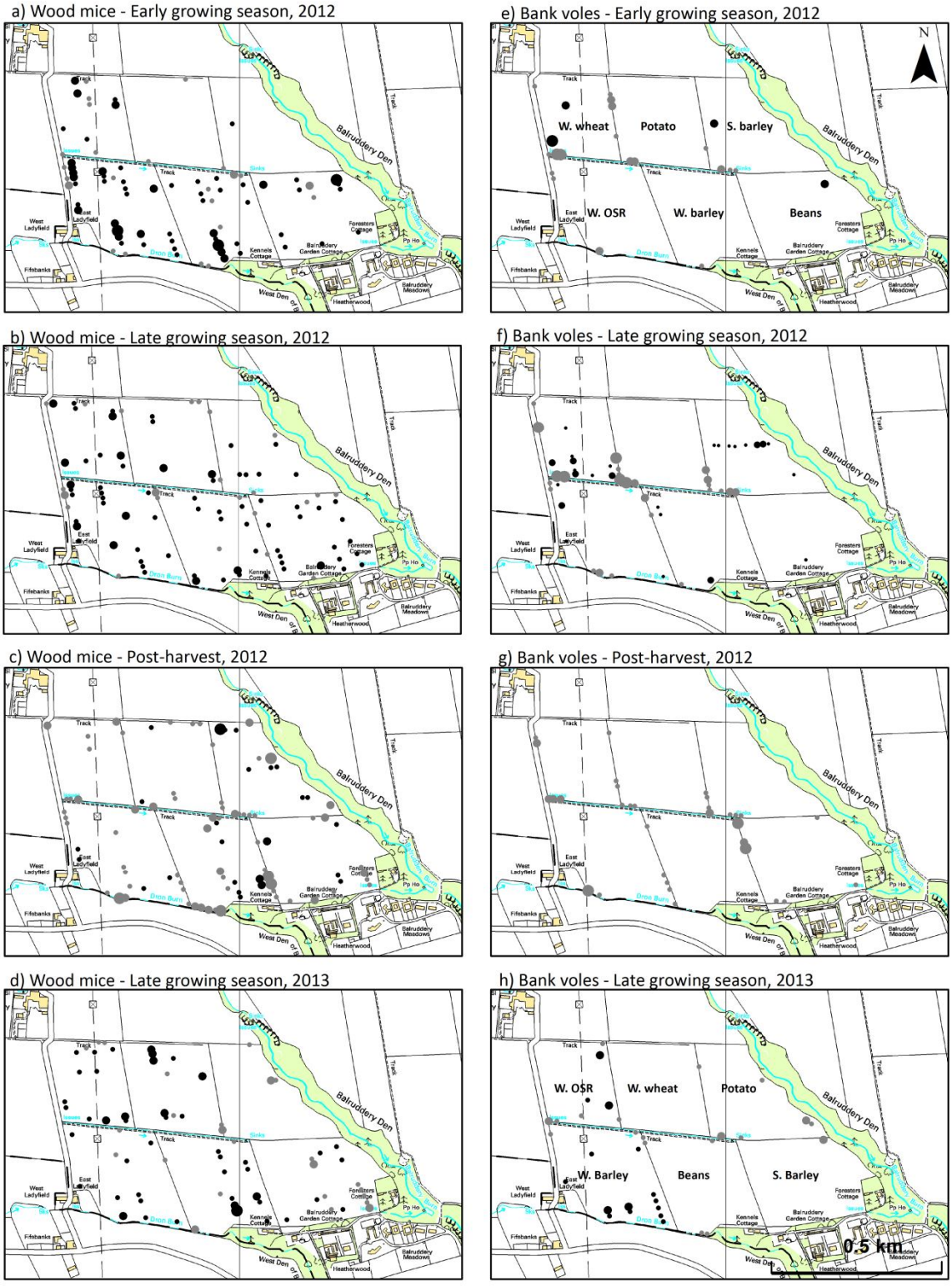


Figure 2.3 Distribution of the species maps were produced in ArcGIS 10.1. Circles represent individuals caught and are scaled to indicate one, two or three individuals. Grey circles indicate captures within margins and black circles indicate captures within cropped habitat. © Crown Copyright/database right 2013. An Ordnance Survey/EDINA supplied service.



2.3.3 Spatial continuity in distribution

To investigate the continuity of wood mice and bank vole spatial distribution, capture success of each season was modelled against capture success of the previous season. Although early growing season wood mouse capture success could predict late growing season wood mouse capture success in the crop (Table 2.3), the late growing season did not predict post-harvest capture success (Table 2.3), supporting the hypothesis that harvesting would influence this species. For bank voles, early growing season capture success predicted late growing season capture success in the margin but not the crop (Table 2.3) and in contrast to mice, late growing season could predict post-harvest growing season capture success in the margins (none in crop) (Table 2.3), suggesting that this species was less affected by the harvesting of crops, probably because it mostly inhabited grassy field margins. 2013 wood mouse and bank vole capture success could not be predicted by post-harvest 2012 capture success (Table 2.3).

Table 2.3 Tests for whether capture success of a current season could be predicted by the previous season ($df_{margins} = 35$, $df_{crops} = 47$). Asterisks indicate significance level ($P_{crit} = 0.005$).

		Wood mice		Bank voles	
Seasons being tested	**Habitat**	**LR χ^2**	**p-value**	**LR χ^2**	**p-value**
Late growing season, 2012 ~ early growing season, 2012	Crop	17.6	<0.0001***	0.00	0.95
	Margins	5.60	0.018	29.4	<0.0001***
Post-harvest, 2012 ~ late growing season, 2012	Crop	0.00	0.944	NA[a]	NA[a]
	Margins	0.00	0.914	23.6	<0.0001***
Late growing season, 2013 ~ post-harvest, 2012	Crop	5.80	0.016	NA[a]	NA[a]
	Margins	0.60	0.452	0.20	0.625

Asterisks indicate significance level: p < 0.005 *, p < 0.001 **, p < 0.0001 ***

[a] Bank voles only caught in margins during post-harvest, 2012 session.

2.3.4 Interaction between the species

There was no evidence for a relationship (positive or negative) between wood mice and bank voles. Separate models were generated for crop and margin transects. In no season did bank

vole capture success significantly predict wood mice capture success (Table 2.4). This suggests the species were randomly distributed with respect to each other.

Table 2.4 Tests for a relationship between wood mice and bank voles ($df_{margins}$ = 35, df_{crop} = 47) (P_{crit} = 0.006).

Season	Habitat	LR χ^2	p-value
Early growing season, 2012	Margins	0.528	0.467
	Crop	0.932	0.334
Late growing season, 2012	Margins	0.020	0.887
	Crop	0.005	0.941
Post-harvest, 2012	Margins	0.264	0.607
	Crop	0.005	0.941
Late growing season, 2013	Margins	2.433	0.119
	Crop	1.687	0.194

No significant p-values.

2.3.5 Wood mice within the crop

Wood mouse capture success varied significantly between the six crops available, when capture success was modelled at the half field level against crop type (Table 2.5). More specifically, significantly greater numbers of wood mice were captured within winter crops compared to spring crops during the early growing season 2012 (exact binomial test, number of trials = 80, p-value <0.0001) and this trend was still evident during the late growing season of 2012 (exact binomial test, number of trials = 85, p-value = 0.05). After harvest, the preference for fields under winter cropping was no longer evident (exact binomial test, number of trials = 42, p-value = 0.877) with numbers caught under the two types being almost identical (winter = 22 mice, spring = 20 mice). For late growing season 2013, the preference for winter crops was not significant (exact binomial test, number of trials = 58, p-value = 0.512). There was no significant difference in wood mouse capture success in sustainable compared to conventional field halves (paired t-test: t-statistic = -1.60, df = 23, p-value = 0.126).

Table 2.5 Tests for crop preference (df = 11). Asterisks indicate significance level (P_{crit} = 0.017).

Season	LR χ^2	p-value
Early growing season, 2012	71.8	<0.0001***
Late growing season, 2012	11.3	0.045
Late growing season, 2013	24.5	<0.001**

Asterisks indicate significance level: p < 0.017 *, p < 0.003 **, p < 0.0003 ***

Several characteristics of the field plots were considered in an attempt to reveal possible drivers of the pattern. Characteristics which were found to differ significantly between the crops were: crop cover, crop height, number of stems, stem weight and seed pod weight (Table 2.6). Weed cover did not differ significantly between the crop types (Table 2.6). Average values of the characteristics that differed significantly are given in Table 2.7. Attention should be drawn to the differences between winter and spring crops.

Table 2.6 Tests for a significant difference in crop traits for six crops types sampled during three sessions.

	Av. crop cover (%)	Av. weed cover (%)	Av. crop height (m)	Av. num of stems	Av. stem weight (g)	Av. pod weight (g)
[a]Early growing season, 2012	< 0.0001*** (df = 11)	0.632 (df = 11)	< 0.0001*** (df = 11)	NA[b]	NA[b]	NA[b]
[a]Late growing season, 2012	0.007 (df = 11)	0.765 (df = 11)	< 0.0001*** (df = 11)	<0.0001*** (df = 9)	< 0.0001*** (df = 9)	0.0003*** (df = 9)
[a]Late growing season, 2013	0.0002** (df = 9)	0.167 (df = 9)	< 0.0001*** (df = 9)	0.054 (df = 9)	<0.0001*** (df = 9)	<0.0001*** (df = 9)

Asterisks indicate significance level: p < 0.017 *, p < 0.003 **, p < 0.0003 ***
[a]Due to variation in the magnitude of values in 2012 and 2013, the seasons have been tested separately.
[b] Not sampled.

Table 2.7 Averages for crop characteristics with corresponding standard deviations in brackets after each value.

Crop type	[a]Av. crop cover (%)	[a]Av. crop height (m)	[b]Av. num of stems	[b]Av. stem weight (g)	[b]Av. pod weight (g)
Winter OSR	70.3 (42.6)	93.9 (71.9)	25.0 (3.6)	277.8 (75.8)	238.0 (81.1)
Beans	53.1 (38.0)	69.3 (59.5)	14.8 (2.2)	192.9 (54.9)	108.5 (66.9)
Winter barley	72.6 (7.1)	90.6 (3.7)	241.4 (98.1)	231.5 (82.2)	212.8 (50.6)
Winter wheat	44.7 (21.1)	44.8 (34.5)	148.4 (42.8)	260.8 (109.3)	211.3 (55.8)
Spring barley	54.1 (30.1)	47.4 (41.3)	297.7 (99.6)	319.7 (185.1)	214.9 (73.0)
Potato	27.1 (26.4)	13.5 (11.7)	[c]NA	[c]NA	[c]NA

[a]Crop cover, weed cover and crop height were averaged across two sample sessions in 2012 and one in 2013. [b]Number of stems, average stem weight and average pod weight were averaged during one sample session in 2012. [c]Data for potato crop unavailable.

Table 2.8 Comparison of key characteristics and life history traits of the three species. Information derived from Harris and Yalden (2008).

Trait	Wood mice	Bank voles	Field voles
Size	Females = 17.8 grams; males = 19.1 grams	Females = 21.9 grams; males = 26.1 grams	Females = 30.9 grams; males = 39.7 grams
Lifespan	< 1 year	Up to 21 months but high mortality	Usually <1 year, few survive until 2 years
Patterns of activity	Mainly nocturnal	Crepuscular in summer, mostly diurnal in winter	Crepuscular or nocturnal in summer, mostly diurnal in winter
Diet	Opportunistic, omnivorous	Mostly herbivorous	Mostly herbivorous
Social system	Winter: both sexes may nest communally. In summer males may have dominance hierarchy, females defend territories either singularly or in groups	Females defend territory, males have overlapping territories	Extensive overlap of territories in winter but less overlap in summer, male dominance hierarchy proposed, males occupy distinct territories in summer
Mating system	Promiscuity	Promiscuity	Promiscuity
Breeding season	Mainly March to October	Mainly March to October	Mainly March to October
Litters/year	Up to six	Up to six	Up to six
Young/litter	Usually 4-7 individuals	Usually 3-5 individuals	Usually 1-8 individuals
Gestation	19-32 days	19-25 days	20.8 days
Development of young	Weaning from 18 days, breed from 7-8 weeks.	Weaned at 18 days, breed from 6-8 weeks.	Weaned at 14-21 days, breed from weaning.
Chromosomes	2n = 48	2n = 56	2n = 50

2.4 Discussion

For the arable study site surveyed, wood mice, bank voles and field voles were the only species captured, with wood mice being the most abundant species. A summary of the key characteristics and life history traits of the three species captured are included in Table 2.8 for reference.

This chapter demonstrated differences in the habitat use strategies of two most abundant small mammal species: wood mice and bank voles. In response to harvest, wood mice switched their habitat preference, whereas bank voles consistently made greater use of the more stable margin habitat. One of the main differences between the species was that the spatial arrangement of individuals remained consistent through harvest for bank voles but was altered for wood mice. This suggested that the habitat use strategy displayed had implications on the stability and persistence of individuals and populations. Wood mice appeared to act as habitat generalists, whereas bank voles were more habitat specialist.

Wood mice were clearly demonstrated to use both kinds of habitat, with their capture numbers, even in the margins, almost equalling that of bank voles, and with active burrows being observed within crop stubble overwinter. However, wood mice were demonstrated to show preferences likely to reflect differences in the underlying quality of the available habitats. In general, there was a preference for cropped habitats over margin habitat during the growing season but possibly to the detriment of individuals because the distribution of individuals changed between late growing season and post-harvest, despite having been maintained during the growing season. Wood mice have previously been found to exhibit greater use of cropped areas during the growing season than margin habitat (Pollard and Relton, 1970; Ylönen *et al.,* 1991; Macdonald *et al.,* 2000; Ouin *et al.,* 2000; Butet *et al.,* 2006) but the stability of areas of high activity has received less attention. The change in habitat use was likely to be due to the impacts of crop harvesting. As an alternative explanation, increased competition during peak densities has been suggested. Although there is no evidence for multi-annual population cycles in either species in Britain, seasonal fluctuations in density have been reported for both wood mice and bank voles (Harris and Yalden, 2008). Nevertheless, competition is less likely as a driver because, at their peak post-harvest (Kikkawa, 1964; Gurnell, 1978; Montgomery, 1989), capture success within the crop declined, instead of remaining stable alongside increases in the margins, as would be predicted under competition scenarios. It is also possible that seasonal changes in habitat preference were sex-biased but

this remains to be tested. Given that males are known to have a larger home range (Wolton *et al.,* 1985; Macdonald *et al.,* 2000), an interaction between sex and habitat use cannot be ruled out. Despite a possibly higher mortality rate at harvest as a result of this strategy (Tew and Macdonald, 1993), the generalist strategy appears to be successful, with capture numbers being at their highest following harvest. This suggests that, during the growing season, cropped habitats supported high reproductive output in wood mice.

Again there was evidence that, although wood mice generally used all crop types, they demonstrated preferences for some crops over others. Although the number of crop replicates was small, the preference appeared to be for winter sown crops over spring crops; the establishment of winter crops perhaps coincided with the choosing of burrow sites on moving into the cropped habitat, when adequate cover became available (Macdonald *et al.,* 2000; Butet *et al*., 2006). In support of this idea, is the fact that winter crop preferences did not persist to the post-harvest period when burrows may have been destroyed during harvest. Winter crop traits which may contribute to this preference over spring crops include a greater crop cover, greater crop height and fewer stems. The crop preferences exhibited in the present study do differ slightly from preferences reported in the literature but a preference for winter wheat during summer has been suggested (Green, 1979) and wheat and barley have been found to be preferred over oilseed rape (Macdonald *et al.,* 2000). The differences in crop preference between studies may be due to a restricted home range size or other constraints, resulting in the inability to exhibit a preference based on the full complement of habitats seemingly available but differences between winter and spring sown crops is something that deserves further consideration.

The evidence for bank voles being more habitat specialist than wood mice is convincing, with the specialism being most strictly exhibited post-harvest, when no bank voles were captured within the previously cropped area. Probably as a result of the strategy of being restricted to permanent margins, their distribution remained stable through the growing season to post-harvest, unlike for wood mice whose spatial distribution was different after harvest. Previous work has alluded to a specialisation on more permanent habitats of various types including hedgerows, single woodlots, forests and permanent grassland (Kikkawa, 1964; Pollard and Relton, 1970; Ylönen *et al.,* 1991; Hansson 1987; Shore *et al.,* 2005; Butet *et al.,* 2006). However, it seems that bank voles in the present study made relatively greater use of the

cropped habitat during the late growing season than in these previous studies, where on the whole they rarely ventured from the permanent habitat.

As possible factors driving the distributions of small mammal species, vegetative cover and food availability have been suggested to be important (Macdonald *et al.,* 2000; Macdonald et al., 2007; Janova *et al.,* 2011). One study reported that nine variables related to cover provision, explained 64% of the variation in small mammal abundance, with wood mice and bank voles being included within the three most abundant species (Panzacchi *et al.,* 2010). Greater levels of cover are believed to provide protection from aerial predators (Tew and Macdonald, 1993) and small mammals are thought to perceive covered habitats as less risky (Jacob and Brown, 2000). In addition to greater levels of cover, preferred areas may provide greater food availability in the form of weed plants and from the crop itself. Previous research reported that radio-tracked wood mice foraged selectively in areas with a naturally high abundance of weedy plant species (Tew *et al.,* 2000) but they are also known to consume crops (Heroldová and Tkadlec, 2011). Given the dramatic changes in both vegetative cover and food availability that accompany the harvesting of crops, it seems likely that the habitat use shift observed for wood mice was at least partly driven by these changes. The alternative habitat use strategies exhibited by wood mice and bank voles could be partly due to previously observed differences in their diets (Watts, 1968).

In the present study, the capture success of wood mice was not significantly associated or disassociated with that of bank voles in any season or habitat, even post-harvest when both species were forced to share the margins. It is an important time to be considering how these two species manage to coexist within available habitat because invasive bank voles acting alongside the greater white-toothed shrew have been implicated in a decline of the native wood mice within the affected parts of the Republic of Ireland (Montgomery *et al.,* 2012). The occurrence of interactions, most intuitively competition, between the two species has not previously been well investigated and to do so would be challenging (Huitu *et al.,* 2004). In the present study, the differing strategies for making use of unstable arable habitat might facilitate coexistence between wood mice and bank voles or alternatively, be the result of competitive displacement. Indeed, coexistence has been predicted to be favoured when two competing species display differing strategies, with one being a selective specialist and the other an opportunistic generalist (Rosenzweig, 1987). Alternatively, the coexistence of the two species in the studied region may be the result of other factors, such as dietary differences (Watts,

1968), different daily activity patterns (Harris and Yalden, 2008) or other unmeasured factors. If one was aiming to detect competition in arable habitat, post-harvest would be the period to focus on since this is when both species appear to be confined to the same small subset of habitat. In the present study, if the species did compete, it did not appear to be played out in terms of spatial competitive exclusion.

In summary, a survey of an example arable field site in this region, suggested that three species were present at appreciable frequencies: wood mice, bank voles and field voles. For the chapters that follow, wood mice will be used as the species of investigation since they were the most abundant. Wood mice appeared to display a habitat generalist strategy and therefore, future findings cannot be extrapolated to species with more specialist requirements, for example, the bank vole.

Chapter 3

Wood mice: Development of a microsatellite multiplex reaction for genotyping individuals from hair samples

Chapter acknowledgements:

Prior to the beginning of my PhD, David Soutar, an honours student working with Gaynor Malloch and Brian Fenton, had extracted DNA from wood mice using the phenol/chloroform method outlined. David used three microsatellite markers, AS11, AS12, AS20, which gave products on polyacrylamide gels and he had trialled the use of Illustra puReTaq ready-to-go PCR beads with fluorescently labelled AS11, AS12 and AS20 with some success. It was at this point that I began.

Abstract

The use of molecular genetics has improved the range of questions that can be addressed within the field of ecological sciences. Here a microsatellite marker multiplex reaction was developed so that wood mice could be genotyped and used as the study organism to address arable landscape genetics questions. Two methods of DNA extraction were trialled in order to obtain DNA from hair samples. Microsatellite markers that had been reported in the literature were tested and different combinations of fluorescently labelled markers used in multiplex were trialled. Stutter peaks were given consideration, as were other potential problems associated with the use of microsatellite markers and an error rate was calculated. Two multiplex mixes, using nine microsatellite markers in total, were chosen for use in future chapters.

3.1 Introduction

Within the field of ecology, genetic methods are becoming increasingly popular and they have made it possible to address a new suite of ecological questions. Improvements in computer power for analysing genetic data and a reduction in the cost of laboratory reagents and equipment have resulted in genetic methods being more accessible (Selkoe and Toonen, 2006). A popular choice within the range of genetic methods available for use in ecological and landscape genetics research is that of microsatellite marker techniques (Balloux and Lugon-Moulin, 2002; Selkoe and Toonen, 2006; Manel *et al.*, 2010; Wang, 2010) because of their relatively high evolution rate, which provides a good degree of resolving power when considering closely related individuals.

Microsatellite markers are sequences of DNA made up of tandem repeats of one to six bases in length found within the nuclear genomes of individuals (Selkoe and Toonen, 2006; Fletcher and Hickey, 2012). The regions of repeats vary in length up to several hundred base pairs (Selkoe and Toonen, 2006). The repeat units are created by slippage of DNA polymerase enzymes during DNA replication but crucially, the flanking sequences of the tandem repeats are usually conserved within species and often within genera and families (Selkoe and Toonen, 2006). Microsatellite markers are believed to be selectively neutral and display Mendelian inheritance, meaning that they accurately confer information about relatedness between individuals (Selkoe and Toonen, 2006).

Microsatellite markers isolated for wood mice have previously been reported in the literature. Harr *et al.* (2000) reported six microsatellite markers, that were found to be highly polymorphic (7–15 alleles) within a sample population of 30 individuals. These six markers had high levels of heterozygosity (0.73–0.97) and did not display linkage disequilibrium when tested on a sample population. Makova *et al.* (1998) reported nine microsatellite markers that were polymorphic (3–14 alleles) and varied in levels of heterozygosity (0.35–0.92). Gockel *et al.* (1997) also reported two microsatellite markers found within wood mice observed to have a heterozygosity of 0.9 and to be polymorphic (8 and 16 alleles). Known pedigrees were considered and no evidence for null alleles was found within the sample population (Gockel *et al.,* 1997). There was no evidence for linkage disequilibrium (Gockel *et al.,* 1997).

Several studies have been published which made successful use of some of these reported microsatellite markers. For example, Booth *et al.* (2009) considered both small and large scale genetic structure of wood mice populations using AS7, AS20, AS34 (Harr *et al.*, 2000), GCATD7S, TNF (Makova *et al.*, 1998) and MSAF8 (Gockel *et al.*, 1997). Bartmann and Gerlach (2001) used MSAF3, MSAF8 (Gockel *et al.*, 1997), AS7, AS20, AS27 and AS34 (Harr *et al.*, 2000) to assign parentage to wood mice in an experimental set-up. Berckmoes *et al.* (2005) considered the genetic diversity and structure of wood mouse individuals on a gradient of heavy metal pollutants using MSAF3, MSAF8 (Gockel *et al.*, 1997), CAA2A, GACAD1A, GCATD7S, TNF (Makova *et al.*, 1998), AS11, AS20, AS34 and AS7 (Harr *et al.*, 2000). And finally, Booth *et al.* (2007) used GACA3BA, GCATD7S (Makova *et al.*, 1998), AS7, AS11, AS12 and AS34 (Harr *et al.*, 2000) to investigate polyandry in the species. Only Berckmoes *et al.* (2005) provided evidence that linkage disequilibrium had been tested and they concluded that loci were independent.

Despite the popularity of microsatellite markers in ecological studies, a wide variety of issues have been raised that need to be considered prior to their use.

There are several fundamental assumptions that underpin the use of microsatellite markers. Most fundamentally, the flanking sequence of any markers needs to be conserved within the species of concern (Hoffman and Amos, 2005; Selkoe and Toonen, 2006). For the most part, the pattern of mutation and the rate of mutation for microsatellite markers are still unclear, although a stepwise mutation model has received attention, whereby repeats are added or deleted one pair at a time (Selkoe and Toonen, 2006). Selective neutrality of markers is often assumed but there has been some suggestion that this should be tested (Selkoe and Toonen,

2006) since some microsatellite markers have been shown to have roles in the organisation of chromatin and in the regulation of gene activity (Goldstein and Schlotterer, 1999; Li *et al.*, 2002). Microsatellite markers could also in theory hitchhike on genes under selection that are nearby on the chromosome (Selkoe and Toonen, 2006). Selkoe and Toonen (2006) suggested that tests for selective neutrality lack power and selection is only revealed when it is very strong but making use of multiple microsatellite markers should help reduce the effect of weak selection at one or two of the loci. The assumption that microsatellite markers display Mendelian inheritance should always be tested but there are few reports to the contrary (for examples see Dobrowolski *et al.*, 2002; Recce *et al.*, 2004); however, the crossing of known individuals is often not feasible (Selkoe and Toonen, 2006). Finally, microsatellite loci are initially assumed to be independent of each other but this can be assessed by testing for the independent assortment of loci (Selkoe and Toonen, 2006).

Perhaps of less concern, are problems associated with the correct scoring of genotypes since these often create more obvious errors and can usually be accounted for when analysing genetic data. Errors of this kind include null alleles, allele drop-out and problems due to stutter peaks. Null alleles occur when there is a mutation in the flanking sequence resulting in no PCR product for a proportion of individuals. Null alleles are also thought to display Mendelian inheritance. Allelic drop-out occurs when primers do not anneal in the first round of PCR and it is believed to be more common in larger alleles (Hoffman and Amos, 2005). Scoring problems due to stutter peaks are thought to be one of the most common error-generators within microsatellite studies (Hoffman and Amos, 2005), although they appear to be rarely acknowledged in published work (but see example trace in Arif *et al.*, 2010). Stutter peaks make it difficult to distinguish heterozygotes with alleles separated by one or two base pairs from true homozygotes and they are thought to be generated by the slippage of the *Taq* polymerase during PCR (Dewoody *et al.*, 2006). The magnitude and shape of stutter peaks varies between loci (Dewoody *et al.*, 2006) so being familiar with the shape of single allele peaks can help with identification of the problem (Hoffman and Amos, 2005). Stutter peak scoring error gives rise to an excess of homozygotes and a deficit of heterozygotes with alleles separated by two base pairs making it possible to detect the problem (Van Oosterhout *et al.*, 2004). Genotyping errors can be especially common when poor quality DNA is used, as is often the case for non-invasive sampling methods (Taberlet *et al.*, 1999; Hoffman and Amos, 2005).

Several suggestions have been made for ensuring rigorous use of microsatellite markers. Firstly, an awareness of the possible problems is crucial so that difficult genotypes can be given close attention (Hoffman and Amos, 2005). Regenotyping of individuals allows calculation of an error rate which will help identify error due to some of these problems (Hoffman and Amos, 2005), although null alleles would be consistently scored, as could stutter peaks. Micro-Checker software (Van Oosterhout *et al.*, 2004) can perform a check of microsatellite data by using the allele frequencies calculated from heterozygotes to predict observed total allele frequencies and the frequency of allele combinations if Hardy-Weinberg Equilibrium (HWE) is assumed (Van Oosterhout *et al.*, 2004). Error due to stutter peaks, large allele drop out and null alleles can therefore be detected using Micro-Checker since these problems have a particular signature revealed when all allele frequencies are considered (Van Oosterhout *et al.*, 2004). In reality, published studies rarely report having rigorously tested their microsatellite marker method, making comparisons between studies difficult (Hoffman and Amos, 2005). There is also currently no well-defined limit on what might be an acceptable error rate, although Smith and Wang (2014) estimated the number of samples that would be necessary to detect genetic differentiation at certain error rates.

In this chapter, two methods of DNA extraction were tested using hair samples from wood mouse individuals. Microsatellite markers reported in the literature were tested on samples of individuals from populations closer to future study sites, and, to save time and money, an attempt was made to develop a multiplex mix of microsatellite markers. The problem of stutter peaks was investigated more closely. Finally, to determine whether null alleles, allele drop-out and stutter peak scoring error occurred, Micro-Checker was used and repeat genotyping was used to calculate an error rate for the methods used. The method developed will be used to explore the landscape genetics of wood mice in arable habitat.

3.2 Methods

3.2.1 Collecting material for trials

Longworth small mammal traps were placed at Balruddery Farm to obtain hair samples from wood mouse individuals. Traps were set in transects of five traps evenly spaced within margin and crop habitat. Hair was sampled from each individual by isolating the individual in a clear plastic bag and gripping it securely with one hand whilst plucking hairs from between the

shoulder blades. The date and location of collection was recorded on the corresponding microfuge tube and the hair samples were stored in a -20°C freezer within 24 hours.

3.2.2 DNA Extraction

Two extraction methods were trialled. In each case, the success of a 1:10 dilution and a 1:100 dilution in a PCR reaction was compared.

A phenol/chloroform method was derived from Chia *et al.* (1985) as follows. 200 µl grinding buffer (100 mM Tris HCl pH 7.5, 10 mM EDTA, 350 mM NaCl, 2% SDS, 7 M urea ultrapure) was added to the hair sample and it was crushed in its microfuge tube using a pestle whilst immersed in liquid nitrogen. The hair sample was then placed on a heating block at 60 °C for 5 min. In a fume hood 200 µl phenol/chloroform/isoamylalcohol (25:24:1) was added to each sample and the samples were inverted 100 times. These were centrifuged in a microfuge at 14,000 g for 5 min and the top layer was removed to a fresh microfuge tube. This phenol/chloroform/isoamylalcohol step was repeated once and then 200 µl chloroform/isoamyl alcohol (24:1) was added to the top layer and the sample centrifuged for 2 min. To the top layer 500 µl of ice cold 100% ethanol was added and the samples were left at -20 °C overnight. The samples were then centrifuged at 14,000 rpm for 10 min and the ethanol was poured off leaving a DNA pellet. 100 µl of 70% ethanol was added to each sample to wash the pellet, they were re-centrifuged and this was poured off so that the pellet could be dried on a heating block at 60 °C until all the ethanol had evaporated. The pellet was resuspended in 10 µl TE buffer (10 mM Tris /HCl pH 7.6, 1 mM EDTA) using a vortex mixer. The quantity of DNA obtained when using this extraction method was determined using a NanoDrop meter for 16 extractions. However, it should be noted that the accuracy of this machine is questionable, and it may not provide accurate readings if DNA is not of uniform concentration in the sample (G. Malloch, pers. comm.).

The second extraction method used sodium hydroxide as described in Stanton *et al.* (1998). 80 µl of 0.25 M NaOH was added to each sample ensuring this amount covered the hair and the samples were left on the bench overnight. The samples were then incubated at 99 °C for 3 min and centrifuged to remove condensation from the Eppendorf lid. 40 µl of 0.25 M HCl, 20µl 0.5 M Tris HCl and 20 µl 2% Triton X-100 were added to each sample and the samples were incubated at 99 °C for 3 min. After they had cooled, the samples were stored at -20 °C.

3.2.3 Choosing microsatellite markers

Potentially suitable microsatellite markers for wood mice were chosen from the literature (Harr *et al.*, 2000; Makova *et al.*; 1998, Gockel *et al.*; 1997). Markers reported to have the most alleles were selected in order to provide the best chance of distinguishing between closely related individuals at a fine spatial scale. The following markers were considered: AS7, AS11, AS12, AS20, AS34, AS27 (Harr *et al.*, 2000), GACAB3A, GCATD7S, GTTD9A, GACAE12A, TNF, CAA2A (Makova *et al.*, 1998), MSAF-8, MSAF-3, (Gockel *et al.*, 1997).

Initially the forward and reverse primers for each marker were ordered in non-labelled form (100 pmol/µl) to confirm that PCR products could be obtained and these were tested using eight individuals. The primers were diluted 1:10 to give a working stock of 10pmol/µl. The working stock of the forward and reverse primer for each marker was combined 1:1. Each PCR reaction contained 6.25 µl of Qiagen Type-It Microsatellite PCR Kit, 4 µl sterile distilled water, 1.25 µl of the working stock primer mix and 1.3 µl of DNA diluted 1:10. As suggested in the Qiagen Type-It Microsatellite PCR Kit manual, the PCR included an initial activation step of 95 °C for 5 min, and 33 cycles of 95 °C for 30 sec, 57 °C for 90 sec, 72 °C for 30 sec before a final extension step of 60 °C for 30 min. Once the PCR was complete, 3µl of gel loading dye was added to each PCR tube and the products were run on a polyacrylamide gel, stained with ethidium bromide, destained and viewed with a UV trans-illuminator and photographed. Prior to trialling each non-labelled microsatellite marker, a multiplex reaction of AS11, AS12, and AS20 markers was trialled using the Qiagen Kit, since this had proved previously to be successful when used with illustra puReTaq ready-to-go PCR beads. This mix gave products on a gel when used in PCR with the Qiagen Kit and it was therefore used in each batch of PCR reactions above as a positive control.

Once it was confirmed that products of the anticipated size had been synthesised, fluorescently labelled versions of reverse primers were ordered. The literature from which the microsatellite primer sequences were taken listed anticipated allele sizes meaning bands of fluorescence corresponding to different microsatellite loci could be easily distinguished. Markers were labelled with one of three fluorescent labels (VIC™, NED™, FAM™), chosen strategically on the basis of anticipated allele sizes reported in the literature so as to later permit multiplex reactions without producing fluorescent peaks that overlapped. Markers that appeared to show some allele size variation were ordered first, since these would give greatest resolution for distinguishing between unique individuals. The same protocol as above was

followed to test the fluorescently labelled primers, except rather than analysing products on a gel, they were analysed using an ABI 3730 Genotyper. The ABI 3730 Genotyper accepts 48 or 96 samples per plate, and injects a small volume of the fluorescently labelled DNA from each well into polymer filled capillaries by an electrokinetic injection (brief pulse of electrophoresis when a voltage is applied) (Applied Biosystems, 2002). When electrophoresis is applied to each capillary, fragments move through the polymer with small fragments moving more quickly so that fragments are separated by size (Applied Biosystems, 2002). The machine has a detection window that emits a laser beam which excites dye molecules causing them to emit fluorescence and this is picked up by a charge-coupled device camera (Applied Biosystems, 2002). The fluorescence information is read by data collection software and displayed as an electropherogram – a trace showing the amount of fluorescence against size in base pairs (Applied Biosystems, 2002). Allele sizes can be scored using GeneMapper software which displays such traces. Allele size calibration was made possible by adding a ladder solution containing bands of known size to each well. The GeneMapper software performed the calibration but in some cases it had to be adjusted manually. Once bands of anticipated allele size had been scored, loci were categorised as appearing heterozygous, homozygous or showing no products. The frequency of heterozygotes was used as an indicator of the success, since heterozygotes should be in the majority, and it is possible that errors such as allele drop out or stutter peak interference could be present in homozygotes. The number of unique alleles observed was also counted to give some indication as to the extent of polymorphism in a small sample.

Next, different combinations of the markers were used in multiplex reactions using the same PCR reagents and conditions as above. Product sizes found in multiplex reactions were compared to results from single reactions for the same individual to confirm that they were consistent.

3.2.4 Scoring allele sizes accurately and consistently

Using a subset of the markers in multiplex (AS11, AS12, AS20) the accuracy of scoring was subjected to various tests.

To investigate the occurrence of stutter peaks, artificial peaks with known two base pair separation were created by mixing samples of known allele sizes for the AS11 and AS12 markers. It was hoped that this would provide further insight into the appearance of a true

heterozygote, with alleles separated by two base pairs compared to a homozygote. DNA from the individuals was mixed in a 1:1 ratio and a 2:1 ratio in case the starting amount of DNA had an effect.

Micro-Checker software was used to detect scoring problems such as null alleles, incorrect detection of alleles separated by two base pairs and allele drop out. For this test, 40 of the trial individuals were scored at AS11, AS12 and AS20.

An overall error rate was calculated when all nine markers were used in multiplex for a small number of samples: 12 individuals with DNA extracted using the phenol/chloroform method were regenotyped and six individuals with DNA extracted using the sodium hydroxide method were regenotyped.

3.3 Results

3.3.1 DNA Extraction

Both phenol/chloroform and sodium hydroxide DNA extractions methods gave products when the extraction was diluted 1:10. When the extraction was diluted 1:100, PCR products were less frequently obtained. The NanoDrop meter suggested the average concentration of DNA obtained from the phenol/chloroform extraction was 20.9 ng/µl but with a standard deviation of 40.3 ng/µl (N = 20 individuals). This suggested that the amount of DNA obtained could have been highly variable but there have also been concerns about using this machine when DNA is not uniformly mixed throughout a sample.

3.3.2 Choosing microsatellite markers

All microsatellite markers tested proved successful when used in non-labelled form, giving different sized products when analysed on a gel (Table 3.1). An example gel is shown in Figure 3.1.

When fluorescently labelled microsatellite markers were run in a single reaction, there was variation in the levels of success. Only those assigned to heterozygotes could be confidently judged to have been successful since assignment to homozygotes could be the result of allele drop out or scoring error (Table 3.2).

Table 3.1 Success of each microsatellite marker in terms of generating a PCR product of the appropriate size and showing signs of polymorphism. The number of alleles reported in the literature is given since this also guided the decision. Eight reactions were carried out for each marker.

Microsatellite marker name	Number of successful runs	Products show polymorphism	Number of alleles in original publication
MSAF8	8	Yes	16
GCATD7S	8	Yes	9
AS20	8	Yes	11
AS12	8	Yes	14
AS11	8	Yes	15
AS7	7	Yes	10
MSAF3	7	Yes	8
GACAB3A	7	Yes	14
GACAE12A	6	Yes	6
GTTD9A	5	Yes	5
AS27	5	Yes	7
AS34	5	Yes	12
TNF	4	Yes	7
CAA2A	4	Yes	6

Figure 3.1 Example gel showing the result of trialling non-labelled markers GACA3BA, AS7 and MSAF3. A ladder and a six control samples were also run. Products show a range of sizes suggesting polymorphism.

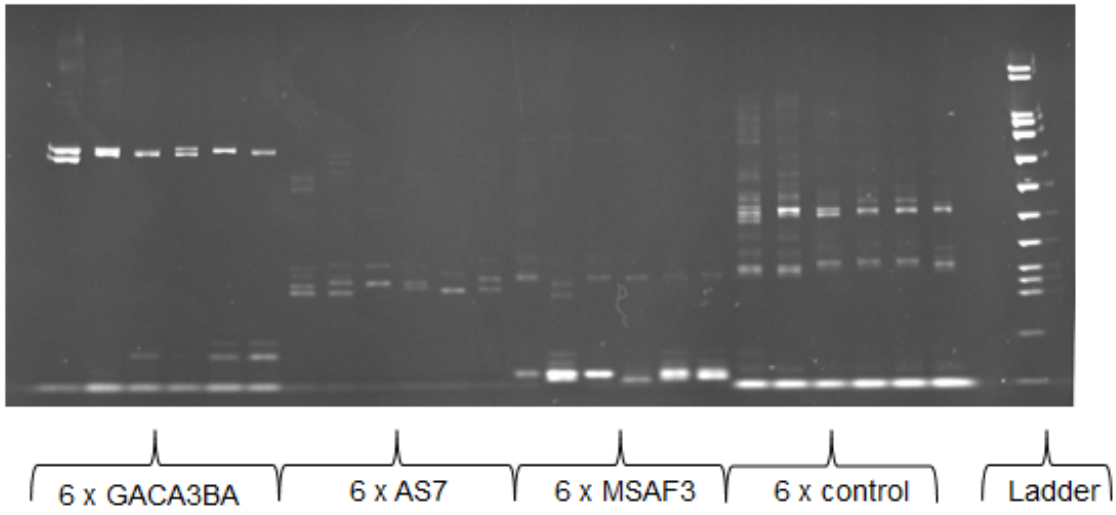


Table 3.2 Success of each marker in a single reaction with the final column showing the proportion of individuals assigned to heterozygotes. Number of unique alleles observed is given. The number of trials varied because 48 or 96 samples are needed to fill a genotyper plate and samples were fitted into gaps on plates.

Marker	No. tested	No. unique alleles	No. heterozygotes	No. homozygotes	No. giving no product	Proportion heterozygotes
AS7	6	5	5	0	1	0.8
AS12	20	15	15	5	0	0.8
AS11	20	10	14	6	0	0.7
GACAB3A	6	4	4	2	0	0.7
AS20	20	8	12	8	0	0.6
MSAF8	15	9	9	1	5	0.6
CAA2A	11	3	6	2	3	0.5
GACAE12A	11	3	6	2	3	0.5
TNF	4	3	2	1	1	0.5
MSAF3	20	6	9	0	11	0.5
GCATD7S	14	6	6	2	6	0.4
GTTD9A	11	4	4	3	4	0.4
AS34	16	3	4	2	10	0.3
AS27	14	1	0	1	13	0.0

When attempting to create a multiplex, firstly a simple multiplex containing AS11, AS12, AS20 proved successful in that it gave products and there was consistency between their sizes in multiplex and in single reaction. A series of combinations of multiplex reaction was then carried out with varying levels of success and using a degree of trial-and-improvement (see Figure 3.2 for an example of a GeneMapper trace). For each multiplex run, allele sizes were scored and compared to the allele sizes inferred from single reactions. To the initial multiplex of AS11, AS12, AS20, various markers were added in combination and the success considered until there was confidence that a successful combination had been achieved (see Table 3.3). This meant ruling out AS27 because it did not give products in single reaction or multiplex and ruling out AS34 because there were suspicions about whether the products observed were truly microsatellites since the peak was a different shape. Finally MSAF8, which gave low amounts of product in multiplex, was combined successfully in multiplex with CAA2A. Therefore, the two multiplex combinations to be used in later work are:

Multiplex 1) AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, TNF.

Multiplex 2) MSAF8, CAA2A.

Figure 3.2 Example GeneMapper trace showing fluorescence on the y-axis in an arbitrary unit and size in base pairs along the x-axis. The AS20 marker was one of the markers included in this multiplex reaction and its heterozygous allele pair is marked by way of example.

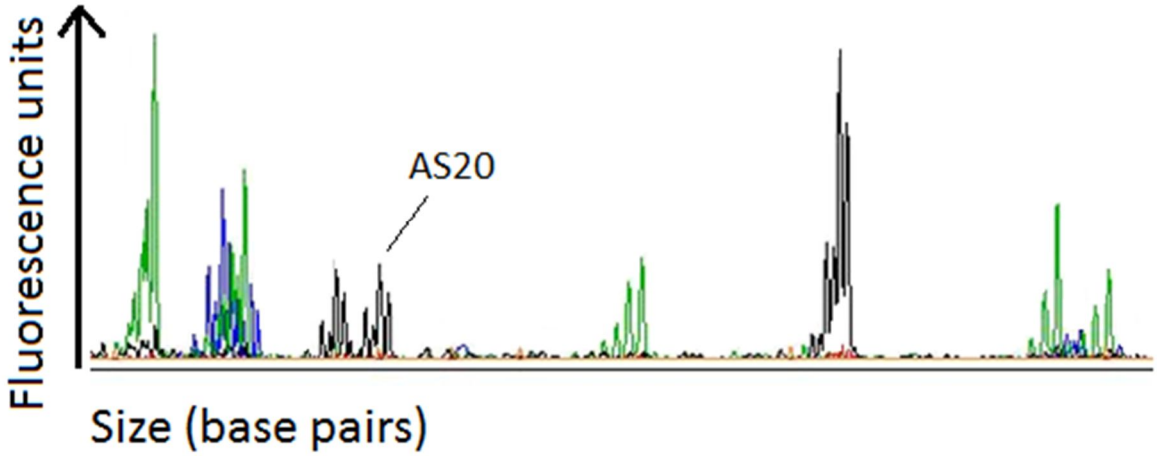


Table 3.3 Table detailing the success of mixing different combinations of microsatellite markers. The allele sizes determined from each multiplex reaction were compared to allele sizes obtained when each marker was run in single reaction. Comments provide rational for each subset of combinations.

Multiplex	No. of reactions	Total no. of allele pairs (A)	No. allele pairs that match single reaction (B)	B / A	Comments
AS11, AS12, AS20	20	60	54	0.90	Adequately successful.
AS11, AS12, AS20, GCATD7S	6	24	18	0.75	Adding any of the following is adequately successful: GCATD7S, MSAF8, AS7, AS34, GACA3BA.
AS11, AS12, AS20, MSAF8	6	24	19	0.79	
AS11, AS12, AS20, AS7	6	24	18	0.75	
AS11, AS12, AS20, AS34	6	24	22	0.92	
AS11, AS12, AS20, GACAB3A	6	24	23	0.96	
AS11, AS12, AS20, AS27, GCATD7S, MSAF8	6	36	19	0.53	Poor. Disregard AS27: no product as with single reaction.
AS11, AS12, AS20, AS7, AS34, GACAB3A	10	60	47	0.78	All mixes are adequately successful. MSAF8 has very low peaks. AS34 shows little variation & trace peaks are a different shape compared to the other microsatellites.
AS11, AS12, AS20, AS7, AS34, GACAB3A, MSAF8	11	77	67	0.87	
AS11, AS12, AS20, AS7, AS34, GACAB3A, GCATD7S	11	77	66	0.86	
AS11, AS12, AS20, AS7, AS34, GACAB3A, GCATD7S, MSAF8	11	88	62	0.70	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, MSAF8	6	42	38	0.90	

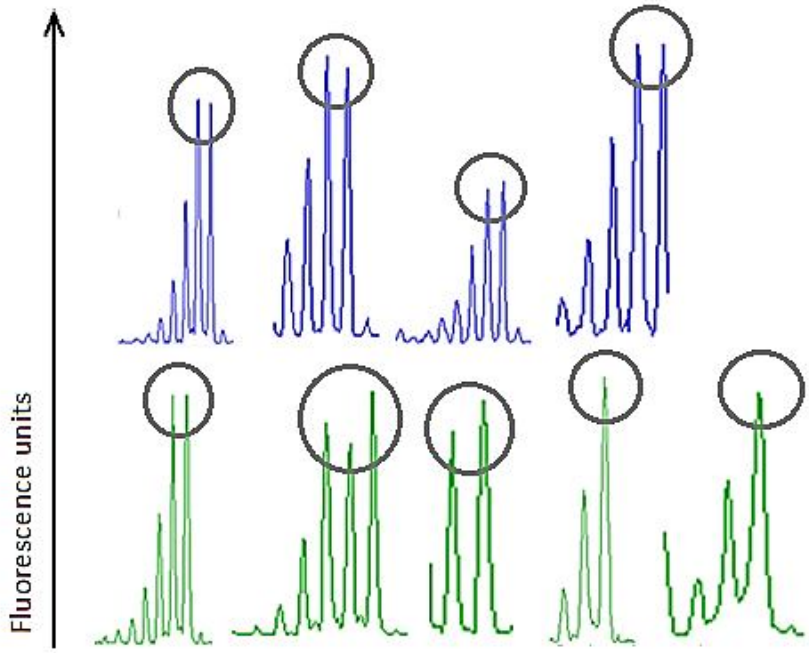
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S	6	36	34	0.94	Very successful.
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, double conc of MSAF8	6	42	17	0.40	Doubling the concentration of MSAF8 and AS34 reduced success further.
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, MSAF8, double conc of AS34	5	40	27	0.68	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, GTTD9A	3	21	6	0.29	Adding more markers to the successful combination above gave poor success for all combinations except when TNF was added.
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, TNF	3	21	18	0.86	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, CAA2A	3	21	6	0.29	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, GACAE12A	3	21	7	0.33	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, TNF, GTT9DA	7	56	8	0.14	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, TNF, GACAE12A	7	56	8	0.14	
AS11, AS12, AS20, AS7, GACAB3A, GCATD7S, TNF, CAA2A	7	56	20	0.36	
MSAF8, AS34	6	12	5	0.42	Poor success except for MSAF8 and CAA2A in combination.
MSAF8, GTT9DA	7	14	7	0.50	
MSAF8, CAA2A	7	14	12	0.86	
MSAF8, GACAE3A	7	14	2	0.14	
GTTD9A, GACAE12A	6	12	3	0.25	

3.3.3 Scoring allele sizes accurately and consistently

Firstly, using the AS11, AS12 and AS20 subset of markers in multiplex, the accuracy and consistency of scoring was considered.

To investigate stutter peaks, DNA from individuals of known allele sizes were strategically combined in order to generate example peaks for the situation where a heterozygote had alleles separated by two base pairs. In most cases this was not successful and the PCR either failed or it appeared that DNA from only one individual was amplified because only two alleles were shown (unlike the expected three or four for two diploids). This problem was not eliminated by altering the concentration of the two individuals' DNA in 2:1 ratios. Cases where this method did appear successful are shown in Figure 3.3. There was no predictable pattern in the relative heights of peaks representing two alleles with two base pair separation; sometimes the final two peaks were the same size and sometimes one was larger than the other.

Figure 3.3 A collection of GeneMapper traces showing artificially generated heterozygote peaks for AS11 or AS12 with two alleles separated by two base pairs each time. The peaks of interest are circled in each example. As demonstrated, there is no obvious pattern in the relative heights of these final peaks.



Micro-Checker software was used to detect whether there was any evidence for common scoring problems: null alleles, allele drop-out or error due to incorrect scoring of stutter peaks. When 40 individuals were scored at the three loci (AS11, AS12 and AS20), Micro-Checker reported no evidence for null alleles, allele drop-out or scoring error due to stutter peaks across all three loci.

Finally, the overall error rate when using the chosen two multiplex reactions was determined by re-genotyping individuals (Table 3.4).

Table 3.4 Errors occurring when alleles were scored. The rate of occurrence calculated as the number of times an error occurred divided by the total number of alleles, and the % each type of error contributes to the total error (bp = base pair, sample size for phenol/chloroform extraction was 12 individuals and for sodium hydroxide extraction was 6 individuals).

Extraction method		Incorrect by 1bp	Incorrect by 2bp	Other error	Allele drop-out from one repeat
Phenol/ chloroform extraction	Number of alleles with error (out of 216)	4	4	3	17
	% of total error	14.3	14.3	10.7	60.7
	Rate of occurrence (errors per allele)	0.019	0.019	0.014	0.079
NaOH extraction	Number of alleles with error (out of 108)	4	0	1	4
	% of total error	44.4	0.0	11.1	44.4
	Rate of occurrence (errors per allele)	0.037	0.000	0.009	0.037

3.4 Discussion

DNA was successfully extracted from hair samples taken from wood mice using two different DNA extraction methods. Diluting the extraction product 1:10 proved more successful than when it was diluted 1:100, possibly because a lower DNA concentration at 1:100 decreased the likelihood of microsatellite markers annealing to the DNA in the first few rounds of PCR, resulting in greater allele drop-out. The extraction method that made use of sodium hydroxide

had fewer and less time intensive steps than the phenol-chloroform method. However, it is believed that the phenol-chloroform method provides better quality DNA that can be stored for longer (G. Malloch, pers. comm.).

A selection of microsatellite markers from the literature was trialled in order to generate a suitable multiplex reaction for fingerprinting wood mouse individuals that would be time and cost saving. All markers chosen from the literature gave variable products when non-labelled and run on a polyacrylamide gel. When markers were fluorescently labelled, the success in terms of amplification varied for unknown reasons. By a trial-and-error process, two multiplex mixes were designed:

1) CAA2A, MSAF8.
2) AS7, AS11, TNF, AS12, GACA3BA, AS20, GCATD7S.

It is necessary to consider some fundamental assumptions when using microsatellite marker methods. Here, it can be assumed that the flanking sequences of the markers are highly conserved since, for the most part, the markers chosen gave products indicating successful annealing of primers. Violations of the assumption of selective neutrality are difficult to detect unless markers are subject to strong selection (Selkoe and Toonen, 2006). Using a wide range of markers should reduce the effect of any selection on results (Selkoe and Toonen, 2006) and if strong selection was acting on a locus, it is likely that polymorphic loci would have tended over time to monomorphic loci, whereas all loci chosen here were polymorphic. Violations of the assumption of Mendelian Inheritance have been rare (Selkoe and Toonen, 2006) and breeding individuals of known genotype wood mice was not possible during the project. The patterns of mutation and the mutation rate were not considered here. Makova *et al.* (2000), investigated mutation patterns in the TNF microsatellite sequence found in the genus *Apodemus* by generating a phylogeny using microsatellite flanking sequences and mapping microsatellite allele sizes onto this phylogeny. When considering the *Apodemus* loci as a whole, they found support for the stepwise mutation model of microsatellite evolution, but when only wood mouse alleles were considered, the one step mutation model was rejected – here a single base mutation and recombination generated variation between alleles rather than stepwise mutation. It is therefore necessary to remain open-minded as to the types of mutation occurring in microsatellite sequences. Departure from HWE and linkage disequilibrium will be tested for in future chapters, when the sample size is large and once any

population substructure has been determined – since this could give rise to certain allele combination occurring more frequently than expected, as with linkage, but not being caused by it.

The occurrence of stutter peaks when scoring microsatellite traces on GeneMapper was investigated further. Attempts to create two base pair separation of peaks only were not always successful. It was initially thought that this might be due to the sample with the highest concentration of DNA being preferentially amplified in the first few rounds of PCR. However, doubling the concentration of one individual of the pair in turn made no improvement. On the occasions where peaks separated by two base pairs were created, no predictable pattern of stutter peaks was observed, suggesting that it would not be possible to find a rule for scoring such peaks. It is possible to imagine the scenario illustrated by the schematic in Figure 3.4, where on joining the traces of two adjacent microsatellite peaks, a single compound peak of given shape is produced. However, in reality, often the two microsatellite peaks of a heterozygote are different heights (different amounts present) and this may explain why it is more difficult to predict the resulting shape. The schematic presented in Figure 3.5 is therefore closer to reality. Being familiar with this issue and the occurrence of heterozygotes with alleles separated by two base pairs, increases the chance of appreciating the occasions when it could potentially cause a problem (Hoffman and Amos, 2005). From the complement of nine markers used here, CAA2A, GACAB3A and GCATD7S rarely displayed stutter peaks but they were evident in the other markers.

Figure 3.4 Schematic showing the predicted peak shape when alleles are separated by two base pairs assuming no stutter peaks occur. Ultimately the trace obtained appears more like the compound peak shown by a solid red line (in reality, stutter peaks would add further complication). This is made up of the two underlying peaks, assuming both alleles of the heterozygote were present in the starting sample in equal amount. Adding in the effect of stutter peaks makes the ultimate compound peak difficult to distinguish from peaks of homozygotes.

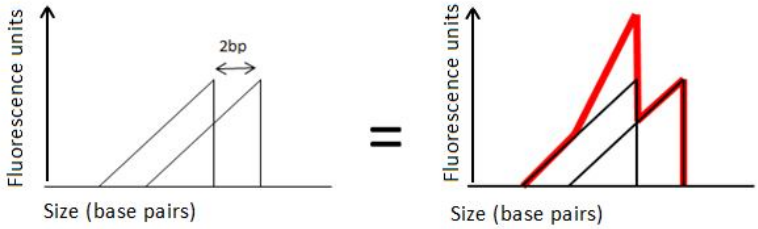
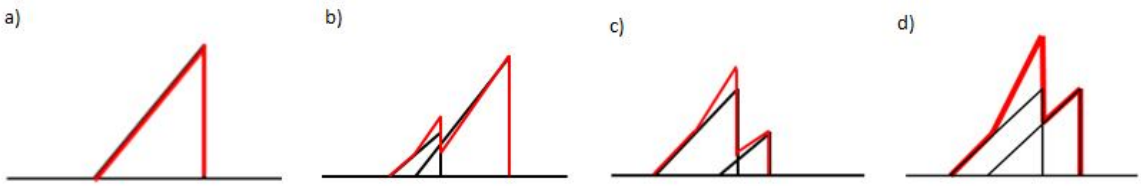


Figure 3.5 More realistic schematic accounting for the fact that the two allele peaks are often different heights. In reality, all that is seen on the trace is the red line, but usually with visible stutter increasing the difficulties in distinguishing between the four scenarios below. a) Shows true homozygote with two alleles of same allele size but in reality this would show up as a series of stutter peaks. b) Shows heterozygote with alleles separated by 2bp but with more of the larger allele being present. c) Shows heterozygote with alleles separated by 2bp but with more of the smaller allele being present. d) Shows heterozygote with alleles separated by 2bp with both alleles being present in equal amount. When the effect of stutter peaks is added to this simplistic representation, all of the patterns below appear more similar, introducing scoring difficulties.



A trial of Micro-Checker for a sample of 40 individuals scored at AS11, AS12 and AS20 suggested that null alleles, allele drop-out and error due to mis-scoring of stutter peaks did not influence the reliability of data collection for the chosen microsatellite markers.

Error rates were calculated by regenotyping a subset of the samples for both the phenol/chloroform extraction method and the sodium hydroxide extraction method and noting any discrepancies between the microsatellite genotypes. The types of error were separated into categories because they will affect analysis in different ways. Error due to missing alleles when samples were repeated was the most common error (phenol/chloroform method: 0.079 errors per allele; NaOH method: 0.037 errors per allele), probably caused by allele drop out, which is more likely when low concentrations of DNA are used (Hoffman and Amos, 2005; and indeed here, when 1:100 dilutions of DNA were used in the PCR rather than 1:10, alleles did appear to drop out more often). The concentration of DNA in each extraction measured with the NanoDrop meter was highly variable but, on the whole, not unreasonably low when compared with literature values (Gagneux *et al.*, 1997; Goossens *et al.*, 1998). The possibility of machine error having occurred here suggests that DNA concentration values should be considered with caution. Errors due to mis-scoring by one base pair also occurred occasionally (phenol/chloroform method: 0.019 errors per allele; NaOH method: 0.037 errors per allele). These could occur due to problems with the calibration ladder but is perhaps more likely to be due to the difficulty associated with scoring stutter peaks. Error due to mis-scoring by two base pairs occurred occasionally (phenol/chloroform method: 0.019 errors per allele; NaOH method: zero errors per allele), again likely to be a problem due to stutter peaks. Problems other than these occurred infrequently also (phenol/chloroform method: 0.014 errors per allele; NaOH method: 0.009 errors per allele) and could be due to contamination or bleed through from one trace colour to another.

In comparison to the literature on error rates, the values observed in the present chapter are close to the higher end of what is reported. Hoffman and Amos (2005) reviewed the literature and thought that 0.001-0.02 errors per allele was the approximate range; although they provided no citations for this. Selkoe and Toonen (2006) suggested 1% of alleles (i.e. 0.01 errors per allele) being misidentified would be an 'uncommonly good' number, which if true, might suggest the error rate here is adequate. Unfortunately, studies seem to rarely report error rates in their published work and when they do, the units of measurement vary, making comparisons with this and other studies unsatisfactory (Hoffman and Amos, 2005; Selkoe and

Toonen, 2006). A study using faecal DNA from Eurasian badgers reported that allele drop-out occurred in 27% of the reactions involving heterozygotes, and erroneous allele scoring in 8% of the reactions (Frantz *et al.*, 2003). Data to make comparisons to the error rate calculated by Frantz *et al.* (2003) was not available. In contrast, Goossens *et al.* (1998) obtained an error rate of 0.0029 per reaction, solely due to allele drop-out, when DNA was extracted from 10 hairs from alpine marmots but a much greater error rate when only one hair was used (0.14 per reaction).

Several factors may influence error rates of studies. For example, it is likely that highly polymorphic loci will be more difficult to score, generating more scope for error. Furthermore, the type of system used to view and score microsatellite alleles almost certainly affects the accuracy and consistency – one might hypothesise that using fluorescently-labelled markers with systems that allow very precise readings would be more accurate that the traditional method of using non-labelled markers run on gels, which can be difficult to calibrate and score precisely. However, because of the level of precision, errors might also be more obvious when using fluorescently labelled primers improving the accuracy of the calculated error rate. These three factors should be given consideration when interpreting error rate. In reality, the error rate value itself is of little use, rather the types of error and how they affect future data analysis are of key importance.

In conclusion, two multiplexes have been produced that can be used in the work of future chapters to fingerprint wood mouse individuals. Potential problems associated with the scoring of microsatellite traces have been considered and the insights gained will be applied when addressing landscape genetics questions.

Chapter 4

Landscape genetics in space and time: insights from sampling wood mice at multiple time points in arable habitat

Manuscript submission:

This chapter contributed to a manuscript submitted for possible publication.

Chapter acknowledgements:

Small mammal data collection was assisted by Shona Jack, Rebecca Sargeant, Jennifer Rowland, Karim Hussain, Sarah King and Nick Argiropoulos. Fields were provided by the James Hutton Institute's Centre for Sustainable Cropping at Balruddery Farm which is run by Cathy Hawes with assistance from Euan Caldwell, John Bennett and Stuart Wale. Three anonymous reviewers provided feedback on an earlier version of a manuscript derived from this chapter and submitted for consideration to Molecular Ecology.

Abstract

No landscape genetics studies have considered temporal variation in arable landscapes, despite recent reviews calling for genetic sampling to take place over appropriate spatial and temporal scales. The aim of this chapter was to test the hypothesis that spatial genetic structure could vary over short time periods in arable habitat, by sampling at multiple time points. Hair samples were collected, DNA was extracted, and individuals were genotyped at nine highly polymorphic microsatellite markers. Using the Bayesian-clustering methods of Structure and Geneland and distance-based methods it was shown that the fine-scale spatial genetic structure was not fixed. Three genetic clusters were detected but two of these showed significant variation in the assignment strength of individuals over time. A comparison of the recapture rates between time points, suggested that the turnover of individuals between all seasons was high, and that the harvest process may have been accompanied by mortality of individuals. These results highlight that there are situations when the influence of short scale temporally acting processes should be accommodated in landscape genetic studies by sampling at multiple time points, an approach which has not been previously recognised for arable habitat.

4.1 Introduction

Several reviews in landscape genetics have highlighted the importance of conducting studies at relevant temporal, as well as spatial scales (Storfer *et al.*, 2007; Balkenhol *et al.*, 2009; Bolliger *et al.*, 2014). A large proportion of the ecological genetics literature has comprised work that, at a single point in time, samples genetic measures, often genetic structure, diversity, or differentiation between populations (Heath *et al.*, 2002; Nussey *et al.*, 2005; Nichols *et al.*, 2012). Furthermore, when addressing questions using molecular methods, samples are commonly grouped across multiple years and assumed to be representatives of a fixed genetic picture. The fact that genetic measures can vary over time is frequently acknowledged in work considering large evolutionary timescales, given that bottlenecks, founder effects and drift are well known to influence population subdivision, genetic variation and genes under selection. However, at much shorter timescales there are situations where it might be equally prudent to accommodate temporal processes that could influence genetic measures.

Research explicitly considering processes which act continuously or intermittently on a short timescale is very limited. The effect of direct animal exploitation (i.e. hunting or harvesting) on

genetic measures is perhaps the best example of a short-term temporal process with known effects on population subdivision, genetic variation and genes under selection (Allendorf *et al.*, 2008). For example, altering the culling regime of red deer on an island in Scotland reduced the genetic differentiation between female red deer populations over a 24 year study period (Nussey *et al.*, 2005). Other examples addressing the effect of short-term processes of influence, by sampling at multiple time points, have focused on the effect of life history traits, such as dispersal and social structure or the effect of population cycles (Berthier *et al.*, 2006; Schweizer *et al.*, 2007; Piertney *et al.*, 2008; Ehrich *et al.*, 2009; Pilot *et al.*, 2010; Nichols *et al.*, 2012; Rikalainen *et al.*, 2012). Depending on the question being addressed, processes which act on short time-scales may necessitate genetic sampling at equally short time-scales, if a true genetic picture is to be obtained.

A regularly occurring process, known to affect the ecology of a large proportion of wildlife, is that of agricultural disturbance (Tilman *et al.*, 2001), and arable farmland is an example of a habitat that varies dramatically in space and time. In arable habitat, potential for disturbance is high throughout the cropping cycle, with crop sowing, crop maintenance, harvest and ploughing all requiring the use of heavy machinery and affecting habitat quality to various degrees, depending on the species being considered (Hole *et al.*, 2005). Gauffre *et al.* (2008) recently hypothesized that agricultural disturbance forced the dispersal and hence gene flow of common voles in intensively farmed agroecosystems at a large scale. No known studies have considered whether fine scale genetic structure of populations varies throughout the cropping cycle.

In this chapter, the fine-scale genetic structure of wood mice was examined, by sampling at four time points throughout the cropping cycle, in an effort to detect any changes that occurred over time. In one of the few studies considering the genetics of wood mouse populations, Booth *et al.* (2009) reported that wood mice had a microgeographic genetic structure (<3 km), with four subpopulations identified within a small area, at a fixed time point. This species is common and widespread throughout Europe and, although a generalist species, it utilises and can nest within cropped habitat during the summer months but is present in non-farmed marginal habitat in higher numbers post-harvest (Ylönen *et al.*, 1991; Macdonald *et al.*, 2000; Ouin *et al.*, 2000; Butet *et al.*, 2006). The life history traits, social structure and mating system of wood mice are still being researched but it is likely that promiscuity occurs, with each litter having multiple fathers (Booth *et al.*, 2007). Few

individuals survive longer than one year, and population abundance is at its lowest in spring and highest in autumn, after the June to September breeding season (Kikkawa, 1964; Gurnell, 1978; Montgomery, 1989). Females have four to seven litters during the breeding season, consisting of four to six offspring per litter (Macdonald and Tattersall, 2001; Booth *et al.*, 2007). Cooperative breeding has been observed in laboratory studies, but in a field study, females were found to nest alone during summer (Wolton, 1985; Gerlach and Bartmann, 2002). Home range size varies with sex and habitat quality but, in general, nightly movements are within 1 km from the nest site (Wolton *et al.*, 1985; Macdonald *et al.*, 2000).

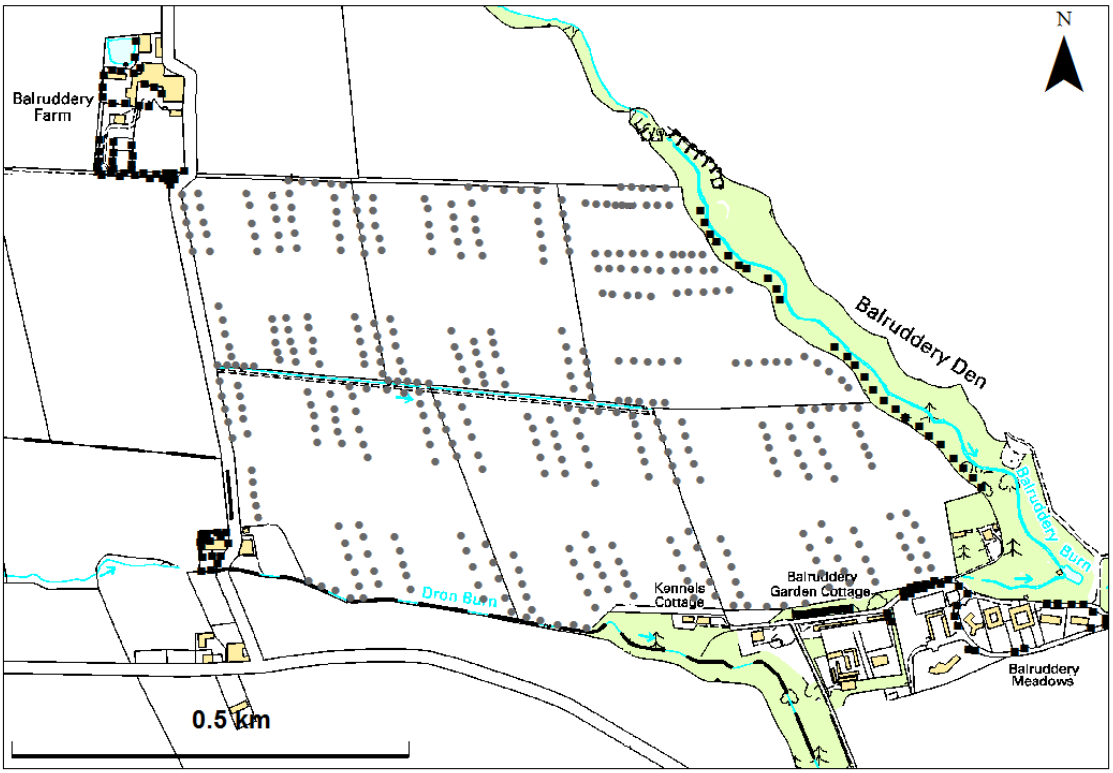
In this chapter, the arable study site was sampled intensively at four time points throughout a cropping cycle. Given that genetic structure was previously evident at a small scale for wood mice (Booth *et al.*, 2009), and because the spatial arrangement of wood mouse individuals was shown to alter over time in Chapter 2, the main aim of the study was to test the hypothesis that changes in fine scale genetic structure might be observed on short time scales.

4.2 Materials and methods

4.2.1 Field sampling

The hair samples used in this chapter came from individuals sampled for work in Chapter 2. From each wood mouse, a hair sample was taken before release and the position of the trap was recorded. There were four trapping sessions intended to obtain genetic structure snapshots at critical time points during an entire cropping cycle: early growing season 2012, late growing season 2012, post-harvest 2012 and the growing season 2013. Crops were harvested during September–October and fields were ploughed from October–March. Sampling was intensive with 120 traps laid in transects in the crop for 3 consecutive nights and 90 traps laid in margins for 3 consecutive nights and this protocol was repeated twice during each trapping session with the transect layout rotated 180° (i.e. each session has 12 trap nights). Transects had a regular arrangement within each field (Figure 4.1). During the post-harvest period, additional trapping of 139 individuals was conducted in adjacent non-farmed habitats, to investigate whether these habitats may have acted as over-wintering refugia (Figure 4.1).

Figure 4.1 Map showing the fields and surrounding features at the study site. Grey circles indicate Longworth trap positions. Black squares indicate trap positions for additional sampling in possible over-wintering refugia at the Balruddery Farm buildings (NW), Balruddery Den (E), Balruddery Meadows (SE), Balruddery Garden Cottage (SE) and domestic gardens to the SW of the site. © Crown Copyright/database right 2013. An Ordnance Survey/EDINA supplied service.



4.2.2 Genetic data

As described in Chapter 3, DNA was extracted from each hair sample using the phenol/chloroform method derived from Chia *et al.* (1985) and each sample was genotyped at nine microsatellite markers: AS7, AS11, AS12, AS20, GACAB3A, GCATD7S, TNF, CAA2A and MSAF-8 (Gockel *et al.*, 1997; Makova *et al.*, 1998; Harr *et al.*, 2000). An error rate was calculated from inconsistencies between 30 recaptured individuals and their original genotype.

4.2.3 Preparatory analysis

To test for null alleles, error due to stutter peaks and allele-dropout, Micro-Checker was used (Van Oosterhout *et al.*, 2004).

To remove recaptured individuals (duplicated genotypes), the first sample of an individual within a trapping session and its capture coordinate were selected to be used in population genetic analysis. The first capture was chosen because many individuals were only caught once, and recapture instances were not sufficient for the centroid of the recapture positions to provide a much improved estimate of the individuals' locations. To ensure that the probability of catching two individuals with identical genotypes by chance was adequately small, the sibling probability of identity ($P_{(ID)sib}$) was calculated for the nine microsatellite marker combination using GenAlEx 6.5 (Waits *et al.*, 2001; Peakall and Smouse, 2006; 2012). Recaptured individuals were determined using the allelematch package implemented in R statistical software, with a permitted allele mismatch of five as calculated to be appropriate for this sample by the package (Galpern *et al.*, 2012; R Development Core Team, 2014). A threshold level of mismatch is chosen so as to reduce the number of multiple matches in a dataset; that is, the number of times an individual is assigned to more than one unique genotype (Galpern *et al.*, 2014). The level of mismatch recommended can be affected when the scoring of a small number of individuals is less complete or has more errors than average (Galpern *et al.*, 2014). Matching genotypes are aligned and displayed, so the user can appraise the allocations, which the software encourages by providing a sibling probability of identity for each grouping so that the chance of obtaining genotypes with that level of mismatch by chance is appreciated (Galpern *et al.*, 2014).

For each determined genetic cluster, departure from HWE was tested using exact Hardy Weinberg tests, and linkage disequilibrium was examined using Genepop with 10,000 permutations (Raymond and Rousset, 1995; Rousset, 2008).

Multiple testing was accommodated using a Bonferroni correction to reduce p_{crit} (Holm, 1979).

4.2.4 Analysis of population structure

To investigate possible genetic clustering, the Bayesian clustering methods of Structure were used (Pritchard *et al.*, 2000). This software uses Bayesian methods to assign individuals to clusters in an attempt to minimise overall departure from HWE and linkage disequilibrium (Pritchard *et al.*, 2000). To determine optimal cluster number, the posterior probabilities of runs at different numbers of clusters (K) were examined and the considerations of Evanno *et al.* (2005) were implemented in Structure Harvester (Earl and vonHoldt, 2012). Structure was used with 10 independent runs for K = 1 to K = 10 assuming admixture, correlated allele

frequencies between clusters, with a burn-in of 500,000 followed by 500,000 Markov Chain Monte Carlo (MCMC) samples. For the most appropriate cluster number (K), the assignment coefficients for each individual across ten independent runs were averaged using CLUMPP (Jakobsson and Rosenberg, 2007). The individual assignment coefficients suggested by Structure were compared to those of Geneland, which has a similar Bayesian-clustering algorithm but is independent from it (Guillot *et al.*, 2005). As recommended (The Geneland Development Group, 2012), to find the optimal number of clusters, uncorrelated allele frequencies between clusters were initially assumed and 10 independent chains were run with K varying from 1 to 8, with 500,000 MCMC iterations and with a thinning of 100.

To investigate spatial variation in the genetic clustering, the CLUMPP assignment values were mapped separately for each trapping session using ArcGIS 10.1. To test whether the distribution of individuals within a cluster at each time point was spatially random, SADIE software was used (Perry *et al.*, 1999). SADIE implements an algorithm that estimates the effort that would be required to move all individuals to a regular arrangement in sampled space (Perry *et al.*, 1999).

To investigate temporal variation in the genetic clustering, the average assignments and the number of individuals assigned to each cluster with greater than 50% assignment were compared between time points using a Kruskal-Wallis test (a rank based test) and χ^2 test respectively. These were implemented in R statistical software (R Development Core Team, 2014).

As an alternative method to investigating possible genetic clustering, a distance-method was used alongside the Bayesian-clustering methods of Structure and Geneland (Pritchard *et al.*, 2000; Guillot *et al.*, 2005). Allele-sharing distances between individuals (DAS) captured at each time point were displayed using Splitstree with a neighbor-joining algorithm (NJ) (Saitou and Nei, 1987; Huson and Bryant, 2006). The allele-sharing distance was the chosen genetic distance measure because mutation and drift were unlikely to play large roles in generating variation between individuals at the small spatial and temporal scale, as demonstrated empirically by Paetkau *et al.* (1997).

Additionally, the extent of genetic differentiation between pairs of identified clusters was estimated using the Weir and Cockerham (1984) fixation index θ and the significance of θ was tested using 10,000 permutations performed in GenAlEx 6.5 (Peakall and Smouse, 2006; 2012).

Given the recent discussion regarding the effect of isolation-by-distance on Bayesian-clustering methods, evidence for a continuous isolation-by-distance pattern was examined (Frantz *et al.*, 2009; Meirmans, 2012). Two types of genetic structuring are commonly observed across a landscape, concurrently or otherwise: a portion of the genetic variation may be discontinuous in space leading to distinct clusters (which can generate stepwise isolation-by-distance) and a portion may be continuous in space leading to a gradient (which can generate continuous isolation-by-distance). To examine the shape of any isolation-by-distance relationship, geographic distance matrices were calculated using Geographic Distance Matrix Generator (Ersts, 2014) and Euclidean distance was plotted against genetic distance (DAS). A Mantel test with 50,000 permutations was performed for each trapping session using the vegan package available in R statistical software (Oksanen *et al.*, 2012; R Development Core Team, 2014). Additionally, a multivariate spatial autocorrelation analysis was conducted to examine the shape of any isolation-by-distance pattern (Smouse and Peakall, 1999). Correlograms were plotted for distance classes of 50 and 100 metres with a 95% confidence envelope calculated using 1,000 permutations in GenAlEx 6.5, to allow significance testing (Peakall and Smouse 2006; 2012). Spatial autocorrelation was judged to be significant for distance classes where the spatial autocorrelation statistic fell outside the 95% confidence envelope (Peakall and Smouse 2006; 2012).

4.2.5 Factors influencing population structure

Several processes could potentially influence the spatial and temporal patterns in genetic structure. To identify periods of high mortality or dispersal, the recapture rates within and between seasons were calculated. To determine whether there was a period of dispersal to adjacent non-farmed locations over-winter, the relationship between individuals caught in possible refugia (locations listed in Figure 4.1) and previously captured individuals was examined. To further investigate whether there were periods of greater immigration and emigration, the allele richness and private allele richness of each cluster were estimated using the rarefaction methods of HP-Rare (for unequal sample sizes) (Kalinowski, 2005) and compared between time-points using χ^2 tests.

4.3 Results

4.3.1 Preparatory analysis

In total, 583 samples were successfully genotyped with 0.13 errors per allele, an error rate consistent with other studies (Goossens *et al.*, 1998; Frantz *et al.*, 2003; Selkoe and Toonen, 2006). Missing alleles accounted for 0.074 errors per allele, and these were accommodated in analysis software. The probability of identity between siblings ($P_{(ID)sib}$) was 0.0001 for each season, which is within the acceptable range and recaptured individuals were removed from the dataset (Waits *et al.*, 2001). In total, 87 individuals were identified in early growing season, 84 in late growing season, 104 post-harvest and 74 in 2013.

No loci deviated from HWE in any identified cluster ($N_{loci} = 24$, $p_{crit} = 0.002$) and only two pairs of loci showed significant linkage disequilibrium ($N_{pairs} = 84$, $p_{crit} = 0.0006$).

The MSAF8 locus was removed from population analyses because of a possible excess of homozygotes, identified using Micro-Checker (Van Oosterhout *et al.*, 2004).

4.3.2 Analysis of population structure

Three genetically distinct clusters were identified using the Bayesian clustering methods of Structure (Pritchard *et al.*, 2000). Adopting the rationale of Evanno *et al*. (2005), a ΔK peak at $K = 3$ and a smaller peak at $K = 5$ were identified, suggesting that three was the uppermost level of hierarchical structure but that further substructure may have existed (Figure 4.2) (Evanno *et al.*, 2005). Examining the spatial and temporal distribution of assignment probabilities clearly demonstrated both spatial and temporal patterns, which were not evident when assignments were pooled over time (Figure 4.3 e,j,o). Structure histograms for each time point are given in Figure 4.4 for reference.

Figure 4.2 Structure Harvester (Earl and vonHoldt, 2012) Evannoplots showing a) the posterior probabilities against number of K and b) ΔK against number if K. The largest ΔK peak was at K = 3, with a secondary peak at K = 5.

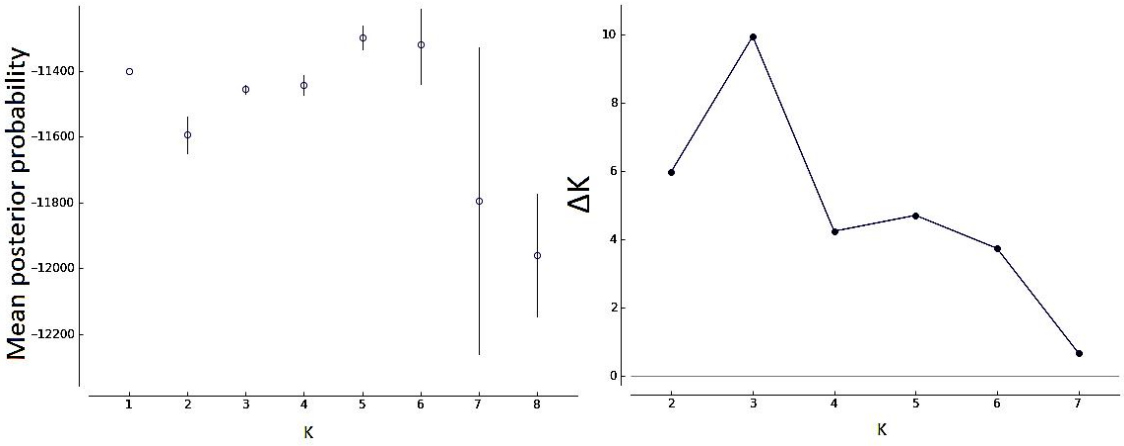


Figure 4.3 (next page) Maps showing the spatial distribution of the clusters defined by Structure. Circles represent capture positions for unique individuals and are scaled by the assignment probabilities for that cluster, with a smaller circle indicating a lower assignment to that group. Three assignment probability classes are shown 0.4–0.6, 0.6–0.8, 0.8–1.0. Cluster A is shown in a-d by blue circles, cluster B in f-I by red circles, cluster C in k-n by green circles, with trapping sessions given in the order: early growing season 2012, late growing season 2012, post-harvest 2012, growing season 2013. The product of these, showing individuals grouped across all time points is given in e, j and o.

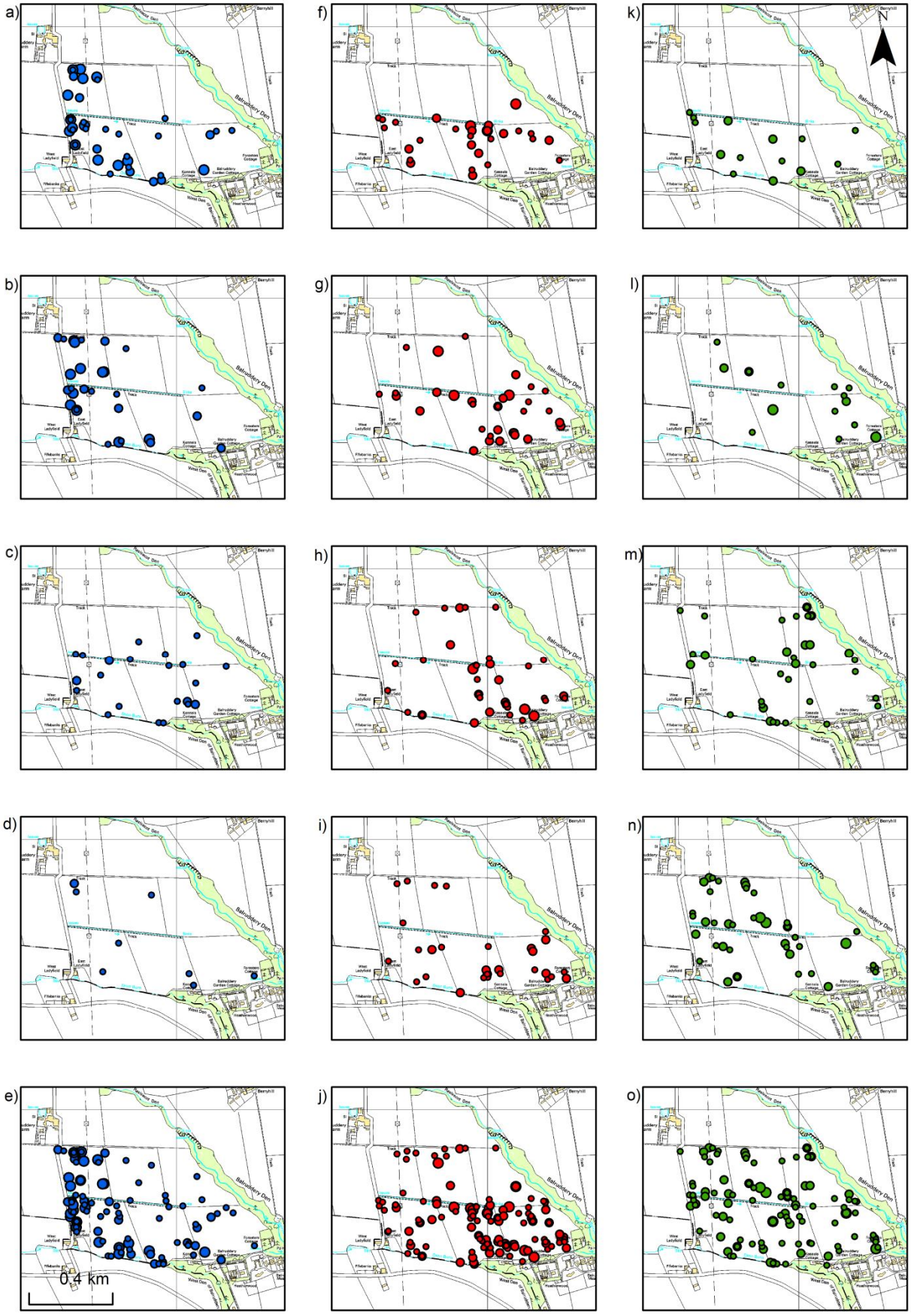
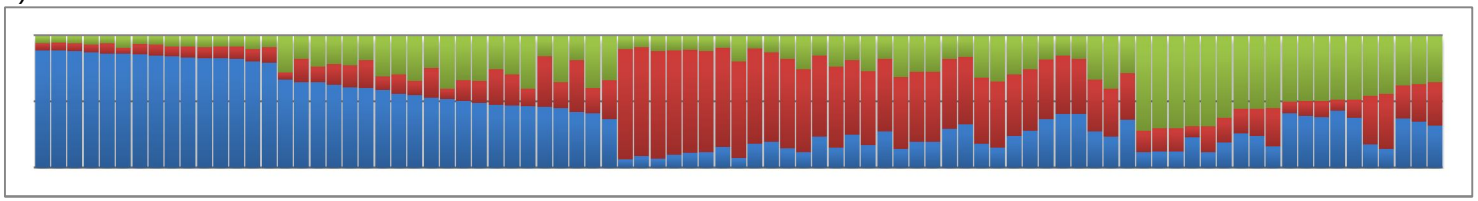
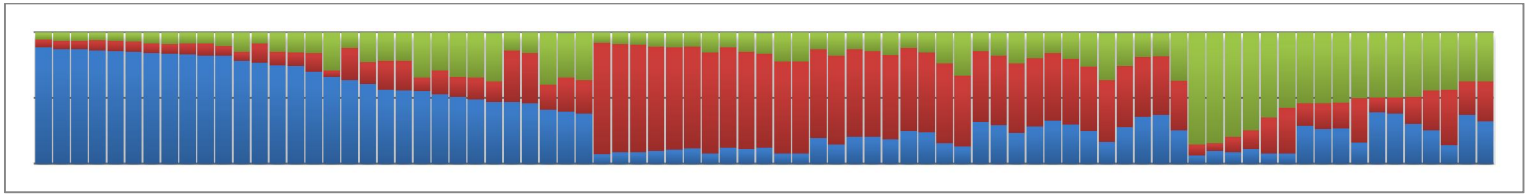

a)
f)
k)
b)
g)
l)
c)
h)
m)
d)
i)
n)
e)
j)
o)
N
0.4 km

Figure 4.4 Structure histograms showing assignments to each of the three clusters for all individuals within a) early growing season 2012, b) late growing season 2012, c) post-harvest 2012 and d) growing season 2013. Colours correspond to Figure 4.3. Individuals are ordered by majority assignment to cluster A, cluster B and cluster C.

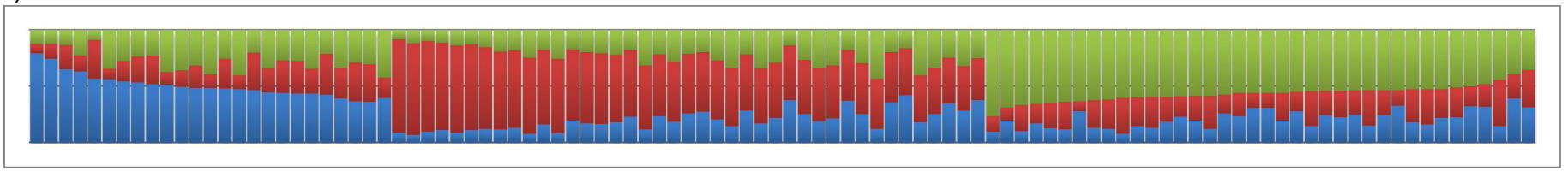
a)



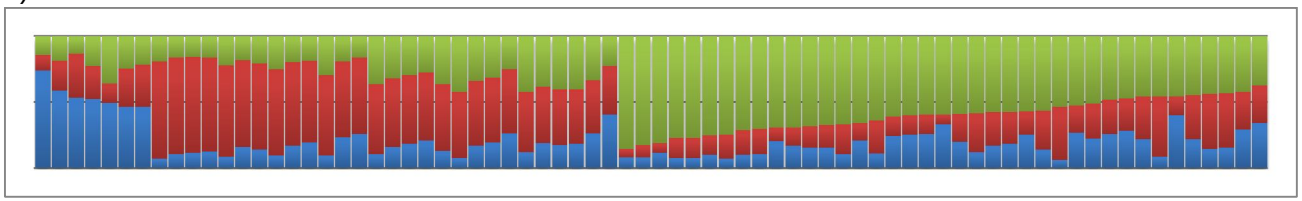
b)



c)



d)



The distribution of all clusters was found to be spatially non-random, except for post-harvest C, 2013 C and 2013 A, which had few individuals (Table 4.1). Spatial overlap between clusters was evident (Figure 4.3).

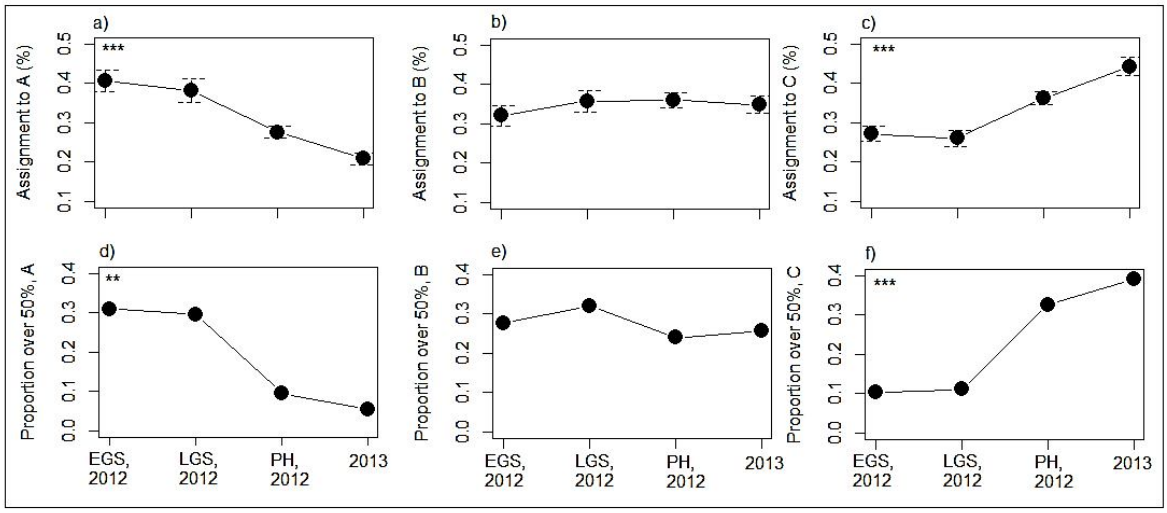
Table 4.1 Result of the SADIE analysis (I_a and P_a values) testing for a spatially random distribution for each cluster.

Season and cluster	I_a	P_a
Early, A	2.42	<0.001***
Early, B	1.57	0.011*
Early, C	1.43	0.033*
Late, A	1.98	<0.001***
Late, B	1.72	0.004**
Late, C	1.55	0.015*
Posthar, A	1.39	0.036*
Posthar, B	1.98	<0.001***
Posthar, C	1.32	0.061
2013, A	1.11	0.227
2013, B	2.06	<0.001***
2013, C	1.32	0.061

Asterisks indicate significance level: $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***

As well as spatial variation, temporal variation in genetic structure was also highlighted. For cluster A and C, there was a significant difference in assignment strength to the cluster over time but assignment strength to cluster B was constant (Figure 4.5) (average assignment, Kruskal-Wallis test, df = 3: A, $\chi^2 = 34.4$, p-value < 0.0001; B, $\chi^2 = 4.49$, p-value = 0.213; C, $\chi^2 = 49.0$, p-value <0.0001; Number of individuals greater than 50% assigned, χ^2 test, df = 3: A, $\chi^2 = 13.1$, p-value = 0.005; B, $\chi^2 = 0.654$, p-value = 0.884; C, $\chi^2 = 33.5$, p-value <0.0001). Cluster A was more abundant during the growing season of 2012, whereas cluster C became more abundant post-harvest and in 2013. The importance of considering this temporal variation was highlighted when assignments were plotted as if for a single time-point, which disguised the temporal variation (Figure 4.3 e, j, o).

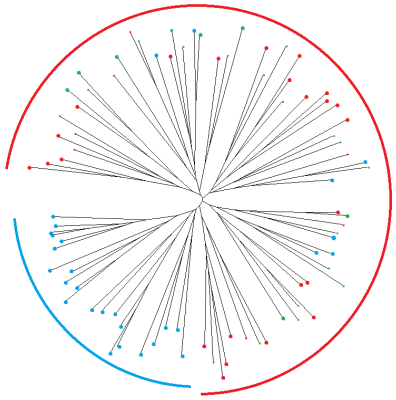
Figure 4.5 Plots showing the average assignment (+/- standard error) to each cluster (a-c) and the proportion of individuals sampled at each time point that were at least 50% assigned to each cluster (d-f).



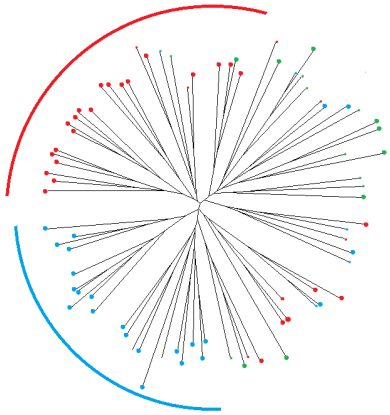
When the dataset was analysed using Geneland (Guillot *et al.*, 2005), a single genetic cluster was identified but when K was set to K = 3, 80% of the individual assignments matched those of Structure (Pritchard *et al.*, 2000), suggesting that an additional level of hierarchical structure could also be identified by Geneland (Guillot *et al.*, 2005). Slight discrepancies between results from the two Bayesian-clustering have also been reported previously (Baker and Hoelzel, 2013; Olsen *et al.*, 2014). The existence of more than one genetic group was validated by plotting genetic distances between pairs of individuals for each trapping season on a neighbor-joining (NJ) tree (Figure 4.6) (Saitou and Nei, 1987), which suggested groupings consistent with the clusters identified by Structure (Pritchard *et al.*, 2000). Smaller groupings were also present on the NJ tree, possibly due to family relationships.

Figure 4.6 Dendrograms showing the genetic distances between pairs sampled during a) early growing season 2012, b) late growing season 2012, c) post-harvest 2012 and d) growing season 2013. Circles added to the end of edges denote the Structure assignment for each individual (A = blue, B = red, C = green). Large circles indicate individuals were more than 50% assigned to that cluster and small circle indicate a majority assignment but weaker than 50%.

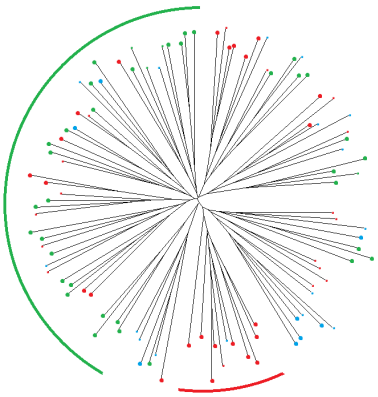
a)



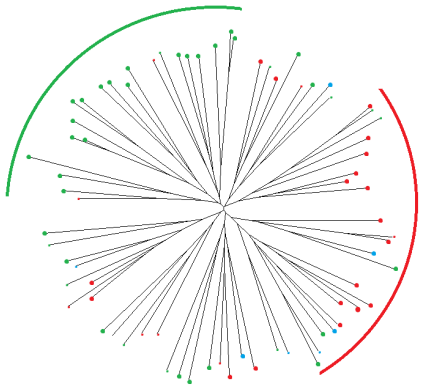
b)



c)



d)



The Weir and Cockerham pairwise θ values provide further evidence in favour of three genetic clusters because the three identified clusters were found to be significantly differentiated from each other (Table 4.2a). For each cluster, the early and late growing season groups were not significantly differentiated (Table 4.2b). The presence of temporal variation was further supported because, for cluster B and C, there was significant differentiation between the growing season 2012 and the post-harvest samples and between the post-harvest samples and 2013 (Table 4.2c).

Table 4.2 Pairwise Weir and Cockerhams θ values for a) the three genetically distinct clusters, b) the three clusters separated into early and late growing season samples, c) the three clusters separated into growing season (GS), post-harvest (PH) and 2013 samples. Pairwise θ values are given below the diagonal and p-values above the diagonal.

a)

	A	**B**	**C**
A	-	***	***
B	0.031	-	***
C	0.016	0.024	-

Asterisks indicate significance level: p < 0.017 *, p < 0.003 **, p < 0.0003 ***

b)

	A, early	**A, late**	**B, early**	**B, late**	**C, early**	**C, late**
A, early	-	0.457	***	***	***	***
A, late	0.000	-	***	***	***	***
B, early	0.032	0.041	-	0.448	**	***
B, late	0.032	0.040	0.000	-	**	***
C, early	0.019	0.020	0.017	0.020	-	0.467
C, late	0.024	0.027	0.031	0.033	0.000	-

Asterisks indicate significance level: p < 0.003 *, p < 0.0007 **, p < 0.00007 ***

c)

	A, 2013	**A, GS**	**A, PH**	**B, 2013**	**B, GS**	**B, PH**	**C, 2013**	**C, GS**	**C, PH**
A, 2013	-	0.005	***	0.082	***	***	***	***	***
A, GS	0.01	-	***	0.005	***	***	***	***	***
A, PH	0.023	0.017	-	0.012	***	***	***	***	***
B, 2013	0.013	0.027	0.022	-	0.423	0.463	0.034	0.048	0.121
B, GS	0.043	0.039	0.052	0.001	-	0.006	***	***	***
B, PH	0.031	0.027	0.02	0.00	0.01	-	***	***	***
C, 2013	0.018	0.024	0.031	0.015	0.034	0.031	-	*	***
C, GS	0.019	0.023	0.029	0.014	0.027	0.024	0.009	-	***
C, PH	0.031	0.032	0.018	0.009	0.037	0.024	0.016	0.015	-

Asterisks indicate significance level: p < 0.001 *, p < 0.0003 **, p < 0.00003 ***

Mantel tests for a linear relationship between genetic distance (DAS) and geographic distance provided evidence for isolation by distance with small significant correlation coefficients (r), except for the early growing season session (Table 4.3). The spatial autocorrelation coefficient mostly fell within 95% confidence envelope for all distance classes except for the 0–200m distance classes in the early and late growing season 2012 and post-harvest, indicating significant spatial autocorrelation at this scale (Figure 4.7). This suggests a stepwise isolation-by-distance pattern, with the 'step' being at 200 m, rather than simply continuous genetic variation.

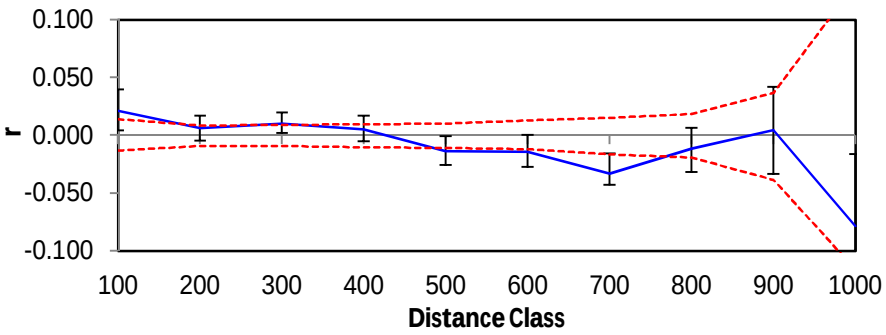
Table 4.3 Mantel test results for a correlation between geographic and genetic distance.

Season	Correlation coefficient	p-value
Early growing season	0.021	0.201
Late growing season	0.134	<0.001***
Post-harvest, 2012	0.061	0.009**
2013	0.046	0.034*

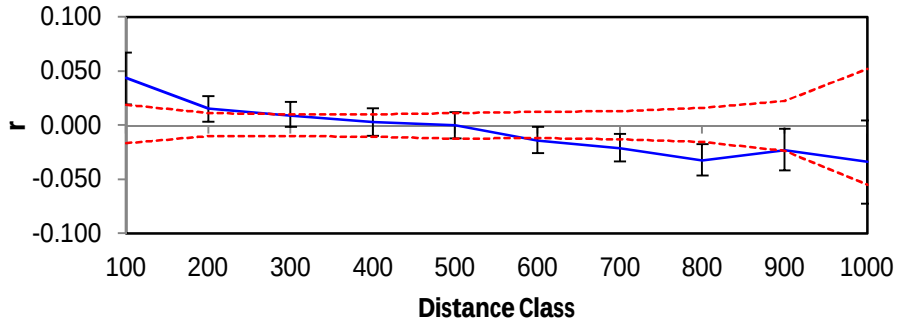
Asterisks indicate significance level: $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***

Figure 4.7 Example correlograms for 100 m distance classes with the 95% confidence envelopes marked by red dashed lines for a) early growing season 2012, b) late growing season, 2012, c) post-harvest 2012 and d) growing season 2013. A correlation coefficient (r) similar to Moran's I is plotted against distance classes. Distance classes with autocorrelation coefficients falling outside the confidence envelope, illustrated using red dashed lines, show significant autocorrelation.

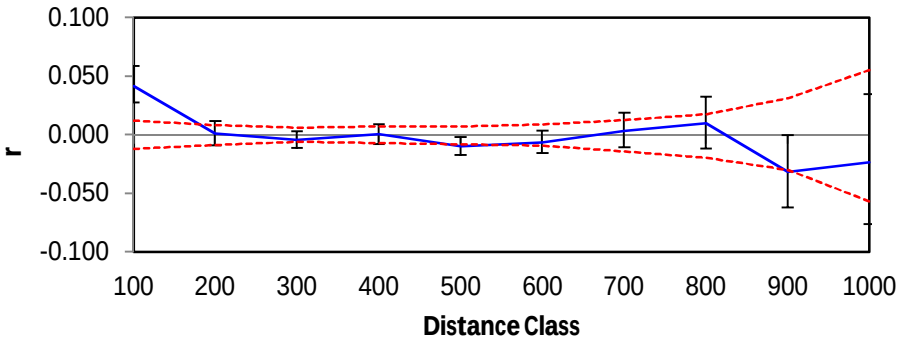
a)



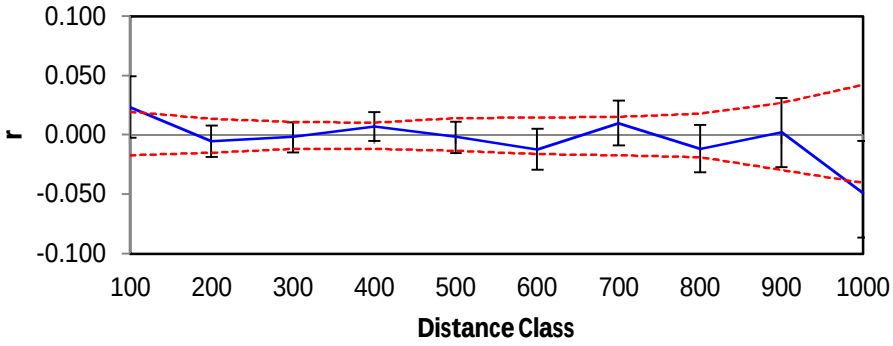
b)



c)



d)



4.3.3 Factors influencing population structure

To identify periods of high mortality or dispersal, the recapture rate of individuals was considered (Table 4.4). Within a season (36.9% on average) and between the early and late growing season (35.7%), a substantial number of the individuals were recaptured. In contrast, few individuals captured post-harvest and in 2013 had been recaptured in a previous season,

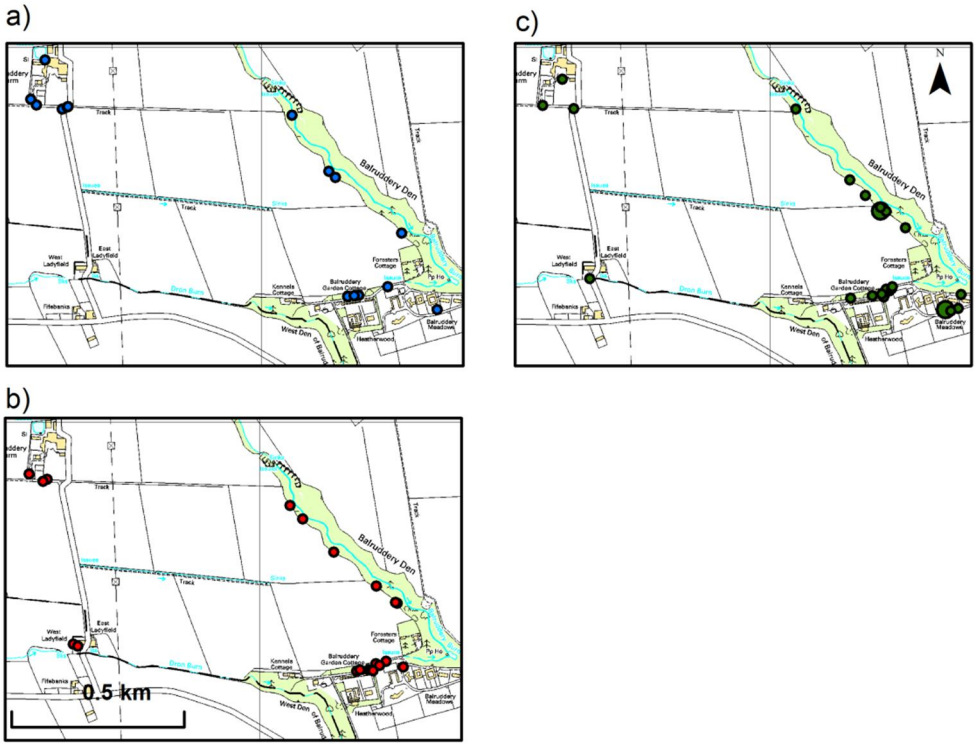
suggesting death or dispersal from the field site between these trapping sessions (0.5%; 0.93% on average respectively).

Table 4.4 Number of recaptures within and between seasons as a total number and as a percentage of the total captures within a season. Sample sizes for each season are given.

	Season (n= num of unique individuals)	**Recaptures**	**% recaptures**
Within a season	Early (n = 87)	32	36.8
	Late (n = 84)	36	42.9
	Post-harvest (n = 104)	30	28.8
	2013 (n = 74)	29	39.2
Between seasons	Late, 2012 from early, 2012	30	35.7
	Post-harvest, 2012 from late, 2012	1	1.0
	2013 from post-harvest, 2012	1	1.4
	Post-harvest, 2012 from early, 2012	0	0.0
	2013 from early, 2012	0	0.0
	2013 from late, 2012	1	1.4

Previously sampled individuals did not appear to disperse to adjacent non-farmed locations post-harvest, since only one individual captured during the growing season was recaptured in non-farmed locations (139 unique individuals captured). Only 51% of the individuals sampled in possible refugia could be assigned to the three original clusters with greater than 50% confidence, compared to previously when 70% of individuals were well assigned. This suggests that individuals belonging to the identified genetic clusters, perhaps offspring of sampled individuals, shared refugia with alternative non-sampled genetic clusters. Of the 68 individuals captured in possible refugia that were well assigned to the identified clusters, there was no clear spatial pattern (Figure 4.8), which is consistent with fact that sampling took place during a period of disruption.

Figure 4.8 Maps showing the positions of individuals well assigned to a) cluster A, b) cluster B and c) cluster C. Circles are scaled according to the assignment probability for each individual (0.5-0.75, 0.75-1) and coloured according to Figure 4.3.



There was no convincing evidence that immigration or emigration rates varied over time. However, allele richness and private allele richness were higher post-harvest and in 2013, but not significantly, perhaps because there were only four data points resulting in a low test power (allele richness, $\chi^2 = 0.299$, p-value = 0.960; private allele richness, $\chi^2 = 3.24$, p-value = 0.356) (Table 4.5).

Table 4.5 Allele richness and private allele richness estimated using rarefaction ($N_{rarefacted} = 74$).

	Allele richness	**Private allele richness**
Early growing season, 2012	108	8
Late growing season, 2012	111	9
Post-harvest	117	15
2013	118	18

4.4 Discussion

The results of this chapter show that wood mouse populations can exhibit genetic structure on a fine spatial scale, with three spatially defined clusters evident within an area of 42 hectares. As hypothesised, the fine scale population structure was dynamic, and the representation of two of these clusters within the area varied significantly over a short time period. This finding represents an original contribution to the literature about the landscape genetics of arable systems because temporal variation in spatial genetic patterns for this habitat type had not been previously demonstrated.

In previous studies, genetic structure at the fine scale has been attributed to social organisation (Schweizer *et al.*, 2007; Gauffre *et al.*, 2008; Booth *et al.*, 2009). In this chapter, genetic structure existed at a fine scale with evident spatial overlap between the clusters. There were no obvious geographical barriers to movement but the spatial autocorrelation analysis indicated a patch size of approximately 200 m, which also corresponded with the area of high density of each identified genetic cluster. Recently, population structure at a fine scale was also demonstrated in wood mice living in more stable habitats (woodland and pastoral farmland), with 850 m between sampled populations, and in common voles, with sampled populations separated by 330–2560 m (Schweizer *et al.*, 2007; Booth *et al.*, 2009). In the absence of spatial separation between clusters, population structure may be driven by social processes, for example, with individuals having a behavioural affinity for other individuals (Schweizer *et al.*, 2007; Gauffre *et al.*, 2008; Booth *et al.*, 2009). In common with the fine scale vole clusters identified by Schweizer *et al.* (2007), family groupings were unlikely to be the major explanation for the genetic clustering identified in this chapter. As for Schweizer *et al.* (2007), there was no departure from HWE in any cluster and each was comprised of at least 98 individuals. Additionally, there was no evidence for substantial linkage disequilibrium and, within a group of close relatives, certain allele combinations would be expected to occur together more frequently than otherwise predicted (Stewart *et al.*, 1999). There has been some suggestion that the inclusion of closely related individuals when using Bayesian-clustering methods could produce artefactual clusters, that is, artificial groups with no biological basis (Anderson and Dunham, 2008; Rodríguez-Ramilo and Wang, 2012). However, it seems unlikely that the genetic clusters identified in the present chapter are artefacts for several reasons. In addition to the groups being in HWE and without linkage disequilibrium, the groups identified are spatially coherent. In contrast, artefactual groups would be expected

to show a more random spatial distribution. Furthermore, the distance based neighbor-joining tree was consistent with the results obtained from Bayesian-clustering methods. Additionally, Anderson and Dunham (2008) suggested that artefacts would be more likely in studies of monogamous species but wood mice have been shown to have a promiscuous mating system (Booth et al., 2009) so any family structure would be expected to be relatively weak. Nevertheless, since juveniles remain dependent for 28 days after birth (Gerlach and Bartmann, 2002), at least some family structure is likely and, future method development for detecting and accommodating family relationships in Bayesian-clustering methods may provide additional insight.

The spatial population processes of wood mice and in particular, the patterns of habitat use discussed in Chapter 2, provide the context within which their population genetics should be viewed. It has been proposed that the dispersal of wood mouse populations between arable habitats provides an example of balanced dispersal dynamics, where dispersal rates between various habitats, such as crop and margins, are equal (Tattersall *et al.*, 2004). Alternatively, it has been proposed that the relative quality of cropped habitat and semi-natural margin habitat drives source/sink dynamics (Ylönen *et al.*, 1991; Macdonald *et al.*, 2000; Ouin *et al.*, 2000; Butet *et al.*, 2006). Estimates of population growth rates necessary to confirm or refute this suggestion are difficult to obtain. However, at the very least, the majority of studies have suggested that wood mice move from stable woodlots and hedgerows in spring into cropped habitat, and recolonise the non-cropped, more stable habitats after harvest (Ylönen *et al.*, 1991; Macdonald *et al.*, 2000; Ouin *et al.*, 2000; Butet *et al.*, 2006). In the present study, the genetic information has demonstrated that few individuals captured pre-harvest were recaptured post-harvest in or around the field site, consistent with either a loss of individuals through high mortality or dilution by population expansion. High mortality around harvest has been reported by Tew and Macdonald (1993) in a radio-telemetry study. However, in the present study equal numbers of wood mice were caught pre- and post-harvest, suggesting a period of population expansion may have taken place balancing out subsequent mortality. In Chapter 2, habitat preferences were shown to shift from cropped habitat during the growing season to margin habitat outside of the growing season. This, in addition to the loss of individuals captured during the growing season, may suggest that cropped habitat supports sink populations that are maintained by the seasonal immigration of individuals from surrounding source populations present in the crop margins. However, confirmatory studies measuring demographic parameters for a range of study sites may provide further insight.

The results clearly identified three genetically distinct and spatially non-random subpopulations at the fine scale, and two of these varied significantly over time. Cluster B had a consistent distribution and representation throughout the study period. In contrast, the representation of cluster A and C was dynamic; the number of individuals in cluster A decreased between the growing season 2012 and the post-harvest sampling, whereas the number in cluster C increased at this time point. The explanation for this change was unclear, but perhaps the simplest explanation is that agricultural disturbance, more specifically the harvesting of crops, reduced the number of individuals in cluster A. At the same time cluster C was able to expand, possibly but not necessarily as a consequence of the decline of cluster A. It is notable that grouping samples across time points and plotting the three groups, appeared to mask these spatial and temporal patterns, which highlights the importance of allowing for temporal variation in future studies.

The possible role of agricultural disturbance on the genetic structuring of populations has received little consideration previously (Gauffre *et al.*, 2008). In one study, Gauffre *et al.* (2008) hypothesised that the disturbance effect of agriculture may have been the explanation for finding only a single genetic cluster in common voles at a scale of 500 km^2. Several other studies reported that in agricultural habitat gene flow was also promoted, again perhaps due to disturbance; for example, for the wild rodent *Calomys musculinus* and for Columbia spotted frogs (Goldberg and Waits, 2010; Chiappero *et al.*, 2011). If agricultural disturbance did influence genetic variation temporally, the persistence of the clusters may have been influenced by factors such as crop type, crop management, and adjacent margin quality. The influence of these factors may also vary between generalist and specialist species and temporal variation may be less apparent for specialist species that are more restricted to non-cropped stable portions of arable landscapes. This may be the case for the bank vole which was found to be more habitat specialist in Chapter 2.

An alternative possible explanation for the temporal variation is that juvenile dispersal drives changes in genetic structure. Nevertheless, the role of agricultural disturbance acting on temporal variation is perhaps a more convincing explanation because the pattern of temporal variation was not uniform across all clusters. Uniformity might be expected if changes were due to routine juvenile dispersal. In the existing literature, the pattern discussed above of seasonal dispersal by all age classes between cropped habitat and non-cropped margin habitat has over-shadowed the juvenile dispersal system. To address questions about juvenile

dispersal would be challenging, given that large numbers of genetic markers and a high proportion of the population would need to be sampled to build a family tree. This will be particularly true for this promiscuous species where birth siblings can have different fathers (Van De Casteele *et al.,* 2001; Blouin, 2003). From the current results it is not possible to distinguish between the relative magnitude of disturbance-driven change and routine juvenile dispersal and ideally future studies would incorporate control sites comprised of arable habitat with delayed harvesting. Manipulations that influenced the financial return from cropped habitat were not feasible in the present project.

The approach of examining small mammal populations at multiple time points has demonstrated that considering genetic structure as a fixed spatial pattern may not enable a complete understanding of population structuring. As hypothesised, the landscape genetics of arable habitat for this species appeared to be best considered across space and time. Because of the possible role of agricultural disturbance in shaping genetic structure, samples gathered in arable habitats, particularly on a small scale, should not be grouped across unreasonably long time periods without reason. Had a single time point been sampled or samples grouped across time points, the conclusion about the genetic structure of wood mice in this arable system would likely be different.

Chapter 5

Landscape genetics and connectivity: a comparison of local genetic diversity and gene flow between wood mouse populations in arable and urban habitat

Manuscript submission:

This chapter contributed to a manuscript submitted for possible publication.

Chapter acknowledgements:

Small mammal sampling was assisted by Jennifer Rowland, Karim Hussain, Hazel Brown, Vicky Stavropoulou, Bethany Westwood, Sarah King and Nick Argiropoulos. Land access was provided by Dundee City Council, the Scottish Wildlife Trust, University of Dundee Botanic Garden, Inverarity Farms Ltd, Balruddery Farm staff and Rossie Priory Estate.

Abstract

As discussed in Chapter 1, studies considering the connectivity provided by arable habitat for wild species have had mixed results. In this chapter, landscape genetic techniques were applied in order to investigate the local genetic diversity and gene flow between wood mouse populations in arable habitat, by comparing this habitat type to urban habitat. Using the nine microsatellite marker multiplex developed in Chapter 3, individuals were genotyped from six arable and seven urban sample sites. Inter-population genetic differentiation was significantly greater in urban than arable habitat, while allele richness, private allele richness and heterozygosity were higher for arable sample sites, with varying degrees of significance. Both suggest that urban habitat was sufficiently fragmented to limit gene flow, whereas arable habitat was more connected. To test the effect of landscape features on gene flow, several cost–distance measures were generated. In arable habitat, overland distance and Euclidean distance best correlated with inter-population genetic differentiation, whereas, in urban habitat, distances that accommodated differences in habitat quality better explained differentiation. No strong evidence was found to support the hypothesis that margins adjacent to roads, rivers or railways facilitated gene flow. The results presented, are discussed in the context of the other landscape genetics studies, detailed in Chapter 1, that investigated whether agricultural habitat acted as a barrier to gene flow.

5.1 Introduction

Landscape genetics aims to better understand how landscape configuration shapes population genetic structure (Manel *et al.*, 2003; Storfer *et al.*, 2007). Urbanisation and agriculture represent two of the most dramatic human modifications of landscape configuration, with 40% of global land surface being dedicated to agriculture (Foley *et al.*, 2005) and 0.2–2.4% being urbanised (Seto *et al.*, 2011). Given that the human population size is predicted to increase, further conversions worldwide are predicted (Seto *et al.*, 2011). The importance of landscape connectivity has been recognised for some time in metapopulation research; dispersal and recolonisation are recognised to be essential for replenishing individuals lost through random extinctions (Levins, 1969). It is now known that due to genetic implications small poorly connected subpopulations may be more vulnerable to extinction (Saccheri *et al.*, 1998; Frankham, 2005). Recently the role of connectivity in enabling species' range shifts in response to climate change has been acknowledged (Heller and Zavaleta, 2009).

When exploring connectivity for a particular habitat type, a comparative landscape genetics approach can be insightful because it allows genetic differentiation due to fine scale social organisation to be accounted for. This is particularly important for landscape genetic studies of wood mice, given that genetic differentiation between their populations has been shown at the fine scale (Booth *et al.*, 2009) and, in Chapter 4, for spatially overlapping groups. In the only other study comparing urban and agricultural habitat for a small mammal, sampling of drylands vesper mouse (*Calomys musculinus*) was undertaken alongside a continuous secondary road verge in arable habitat and this was compared to populations within urban habitat (Chiappero *et al.*, 2011). Without such a comparison the genetic differentiation in urban habitat could be solely attributed to social organisation.

As discussed in Chapter 1, the impact of agriculture on the genetic structure of wild populations has received little consideration (Gauffre *et al.*, 2008). The majority of the literature has focused on appraising whether agricultural habitat can act as a barrier to gene flow but there have been mixed results. A possible barrier effect was suggested for a selection of species in several studies (for example, Cegelski *et al.,* 2003; Coulon *et al.,* 2004; Lindsay *et al.,* 2008), but no barrier effect was evident for other species (for example, Johansson *et al.,* 2005; Purrenhage *et al.,* 2009; Schmidt *et al.,* 2009). It has even been suggested that arable agriculture, with its particularly frequent disturbance, may promote dispersal and widespread gene flow (Gauffre *et al.*, 2008). This was a possible explanation given for the finding that common voles across a 500 km^2 intensive agricultural area belonged to a single genetic unit (Gauffre *et al.*, 2008). In the case of amphibians, agricultural habitat has also been found to offer less resistance to gene flow for Columbia spotted frogs than grassland, forest or developed land (Goldberg and Waits, 2010), though this species was distinguished from the long toed salamander whose gene flow was greatest through forest habitat (Goldberg and Waits, 2010). There has been some movement to promote connectivity within agricultural landscapes, through the strategic implementation of agri-environment schemes (Donald and Evans, 2006) but as demonstrated by Goldberg and Waits (2010), species' responses to landscapes differ, limiting the potential to achieve connectivity for all species.

Urban ecology is a relatively recent research field and many of the consequences of urbanisation are poorly understood (Mcdonald *et al.*, 2008; Munshi-South and Kharchenko, 2010). Within urban landscapes remnants of natural habitat or artificially created green space exist but these are usually small and poorly connected. As a result of this fragmentation, even

species adapted to urban habitat have been shown to form genetically distinct groups within the urban landscape (Munshi-South and Kharchenko, 2010; Chiappero *et al.*, 2011; Gardner-Santana *et al.*, 2009; Gortat *et al.*, 2013). Urban populations have been reported to have low genetic diversity and heterozygosity and these have been linked to lower fitness in some cases (Wandeler *et al.*, 2003; Hitchings and Beebee, 1998, Lens *et al.*, 2000; Chapman *et al.*, 2009). Although urban greenspace is most often implemented for human enjoyment, fitness and health benefits (Maas *et al.*, 2006; Swenson and Franklin, 2000), there has been some effort to evaluate the levels of connectivity it provides for wildlife (Rudd *et al.*, 2002).

In this chapter, the connectivity of arable agricultural habitat and urban habitat for wood mice was directly compared to determine their barrier to movement and gene flow. Wood mice were found to be abundant in both arable and urban habitat. As demonstrated in Chapter 2 and, as has been previously suggested, wood mice are considered a generalist species (Tew and Macdonald, 1993) but they exhibit habitat preferences in both land use types. In urban habitat, they have been shown to prefer undisturbed habitat, for example, orchards and allotments, over disturbed habitat with a high density of buildings (Dickman and Doncaster, 1989) but they have been found to make use of woodland, scrub, orchard, grassland, allotment and domestic gardens (Dickman and Doncaster, 1987; Baker *et al.*, 2003). In arable habitat, they have been shown to move from stable woodlots and set-aside margins or hedgerows into cropped habitat during the growing season (Yl*önen et al.,* 1991; Macdonald *et al.*, 2000; Ouin *et al.*, 2000; Butet *et al.*, 2006) and this pattern of habitat use was also demonstrated in Chapter 2. Wood mice have also been shown to make use of grass verges alongside roads (Bellamy *et al.*, 2000), features which may promote connectivity in both land use types but could also act as barriers to gene flow if they are seldom crossed (Richardson *et al.*, 1997).

In this study, wood mouse individuals from six arable and seven urban sites were genotyped at nine highly variable microsatellite markers. Genetic diversity was compared and variation in the levels of inter-population genetic differentiation under the two land use types was investigated. Given that habitat preferences have previously been demonstrated in urban habitat, some level of inter-population genetic differentiation in urban habitat was hypothesised (Dickman and Doncaster, 1987; Baker *et al.*, 2003). In contrast, because frequent disturbance in arable farmland may promote dispersal (Gauffre *et al.*, 2008), populations were hypothesised to be more similar. As a result of differences in genetic

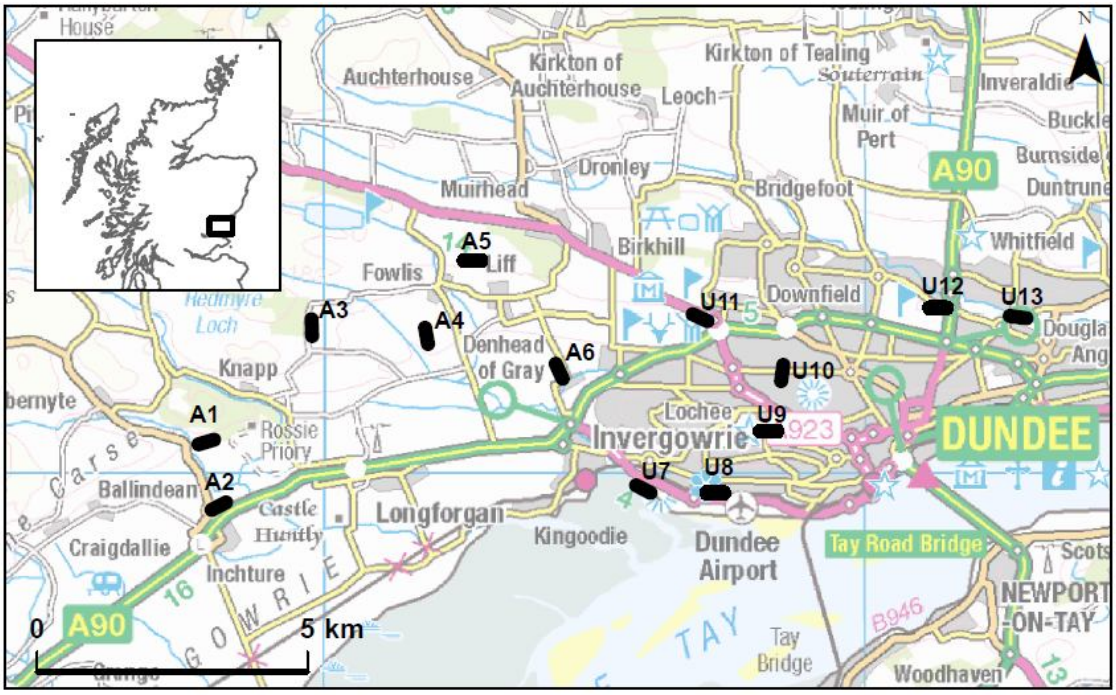
differentiation under agricultural and urban land use, lower genetic diversity was predicted for urban habitat. Finally, various distance metrics were constructed, and a correlation between these and genetic differentiation was examined, to begin to explore pathways for gene flow through the landscape.

5.2 Materials and methods

5.2.1 Field sampling

The work of this chapter was carried out in Dundee, the fourth largest city in Scotland and in an adjacent area of predominantly arable habitat. Sampling was carried out on transects at seven urban and six arable locations, independent from the study site used in Chapters 2 and 4 (arable: mean separation distance = 4109 m, SD = 1883 m, max = 7172 m, min = 1434 m; urban: mean separation distance = 4233 m, SD = 2174 m, max = 7504 m, min = 928 m) (Figure 5.1). Arable transects were placed within crops (barley, oilseed rape or wheat), one metre from tractor tramlines which started 10 meters into the crop. Urban transects were placed in city greenspace with each trap being concealed by existing vegetation. Each 300 m transect consisted of 25 evenly spaced Longworth traps and trapping was repeated for at least three trap nights at each site during both May/June and July/August 2013. The aim was to sample 20–30 unique individuals from each site, whilst accommodating the possibility that genetic variation could vary throughout the breeding season. Given that temporal variation in genetic structure was identified in Chapter 4, field sites in this chapter were sampled during a single arable growing season, the period when spatial genetic variation was previously found to be most stable in Chapter 4. From each wood mouse (90% of all captures) a hair sample was taken before release, and marks made with correction fluid (Tipp-Ex) were used to identify recaptures in the field.

Figure 5.1 Map showing the study area. Transect sites are shown as black lines with arable sites being labelled A1-6 and urban sites labelled U7-13. In the underlying Ordnance Survey base map, grey shading denotes urban habitat, the River Tay is marked (SE) and the remainder of the habitat is predominantly arable farmland with scattered patches of woodland and grassland. Roads and rivers are abundant and marked. *© Crown Copyright/database right 2014. An Ordnance Survey/EDINA supplied service.*



5.2.2 Genetic data

DNA was extracted from each hair sample using the method derived from Stanton *et al.* (1998) detailed in Chapter 3. The two microsatellite multiplexes developed in Chapter 3 were used to genotype the samples using the protocol described and allele sizes were scored. An error rate for this protocol was calculated from observed inconsistencies between the genotypes of 40 regenotyped samples and their original genotype.

5.2.3 Population information

To detect null alleles, allele drop-out and scoring error due to stutter peaks, Micro-Checker was used (Van Oosterhout *et al.*, 2004).

To identify recaptures allelematch was implemented in R statistical software (Galpern *et al.*, 2012; R Development Core Team, 2014) and these were removed from further population analyses. To ensure recaptured individuals could be confidently detected, the chance that two siblings, sampled randomly from the population, had the same genotype ($P_{(ID)sib}$) was calculated using GenAlEx 6.5 (Peakall and Smouse, 2006; 2012).

To test whether actual population sizes differed between arable and urban habitat, the Peterson method was used to estimate population size for each site (White and Searle, 2008). For each day at each sample site, the number of individuals marked on day one (M), the number caught on day two (C) and the number of those that were marked (R) were recorded. An estimate of population size each day (P) was given using P = (M x C) /R and this was averaged across all sample days to gain an estimate of population size. When zero marked animals were recaptured (R = 0), the estimate of population size was undefined and, the population size (P) was estimated by the number of unique animals captured on that day, as has been done in other studies (White and Searle, 2008). The Peterson index correlated with the values obtained when the number of unique individuals per site was divided by the sampling effort in days (Linear regression, $R^2 = 0.742$, df = 11, p-value < 0.0001), providing some confidence in this index. Effective population sizes were also estimated using the molecular co-ancestry method of NE Estimator (Do *et al.*, 2014). To compare population size estimates for urban and arable sites, a Wilcoxon rank-sum test was implemented in R statistical software (R Development Core Team, 2014).

Departure from HWE by loci was tested for at each site using exact Hardy Weinberg tests, implemented in Genepop with 1,000 permutations (Raymond and Rousset, 1995; Rousset, 2008). Linkage disequilibrium was also tested for using 1,000 permutations in Genepop (Raymond and Rousset, 1995; Rousset, 2008).

When appropriate, multiple testing was accommodated with a Bonferroni correction (Holm, 1979).

5.2.4 Habitat differences

Differences between arable and urban landscapes are perhaps obvious, but an objective comparison was made by summarising the habitat composition around each transect, and by estimating habitat connectivity using Fragstats (McGarigal *et al.*, 2002). Habitats displayed on the UK Land Cover Map 2007 were recategorised as arable, urban, suburban, good quality

semi-natural, poor quality semi-natural and water using ArcGIS 10.1 (Morton *et al.*, 2011). The motivation for these categories is detailed in Table 5.1. In order to compare equal areas for urban and arable, the amount of habitat of each category within a 1km buffer around each 300 m transect was calculated using ArcGIS 10.1.

Habitats were then reclassified subjectively as suitable or less suitable for wood mice, by considering whether vegetative cover or food resources would be provided, since these are among the most important resources for small mammals (Macdonald *et al.*, 2007). Those in the less suitable category included habitats such as water, poor quality habitat around the Tay Estuary and urban habitat. A binary raster showing the habitat categories for 1 km buffers around each transect was imported to Fragstats (McGarigal *et al.*, 2002) and three connectivity metrics were calculated: the largest patch index (LPI), the contiguity index and the percentage of like adjacencies (PLADJ). In a more connected habitat, the largest patch would be expected to comprise a greater proportion of the landscape and the contiguity and percentage of like adjacencies would be higher (McGarigal *et al.*, 2002). To test whether the LPI, the CI and the PLADJ differed significantly between urban and arable habitat types, they were compared using Wilcoxon rank-sum tests in R statistical software, suitable for non-parametric data (R Development Core Team, 2014).

5.2.5 Genetic diversity

The genetic diversity was assessed using allelic richness, private allele richness and heterozygosity and a comparison of urban and arable sites was made using Wilcoxon rank-sum tests implemented in R statistical software (R Development Core Team, 2014). To account for slight differences in sample sizes between transects, the rarefaction methods of HP-rare (Kalinowski, 2005) were used. Observed and expected heterozygosity was obtained using GenAlEx 6.5 (Peakall and Smouse 2006; 2012).

Related to genetic diversity, inbreeding coefficients (F_{IS}) were calculated for each site using FSTAT (Goudet, 1995) and the significance of their deviation from zero tested by comparing them to values obtained after 10,000 permutations of the data. A significantly positive value indicates inbreeding, a significantly negative value indicates avoidance of inbreeding and a value of zero is expected under panmixis (Goudet, 1995).

5.2.6 Genetic differentiation

Weir and Cockerham (1984) pairwise θ values were calculated for all pairwise site comparisons and these were categorised as either urban pairs, arable pairs or urban/arable pairs. The significance of each pairwise θ value's deviation from zero was tested in GenAIEx 6.5 using 10,000 permutations of the dataset (Peakall and Smouse 2006; 2012). Differences between the pairwise θ values for the three categories were tested using a Kruskal-Wallis test implemented in R statistical software (R Development Core Team, 2014).

To illustrate the pattern of genetic differentiation between sites, the pairwise θ values were plotted on a neighbor-joining tree (Saitou and Nei, 1987) in Splitstree (Huson and Bryant, 2006). Two further inter-population distance metrics were also plotted; the simple and transparent shared allele distance was chosen, as well as the Cavalli-Sforza chord distance measure (Cavalli-Sforza and Edwards, 1967) which accommodates the effect of genetic drift in its calculation but does not assume that stepwise mutation influences genetic distances. This assumption is unlikely to be appropriate at this scale and for at least one of the markers chosen, as discussed in more detail in Chapter 3 (Paetkau *et al.*, 1997; Makova *et al.*, 2000).

The Bayesian-clustering of Structure was also implemented to determine whether individuals would be assigned to populations corresponding to unique sample sites (Pritchard *et al.*, 2000). To determine optimal cluster number for each trapping session, the posterior probabilities of runs at different number of clusters (K) were considered and the suggestions of Evanno *et al.* (2005) were implemented in Structure Harvester (Earl and vonHoldt, 2012). For the samples pooled across both habitat types and then for each habitat type separately, Structure was run 10 times for K = 1 to K = 20 assuming admixture, correlated allele frequencies between clusters, with a burn-in of 500,000 followed by 500,000 Markov Chain Monte Carlo (MCMC) samples. A similar analysis was repeated with Geneland, which also uses Bayesian-clustering methods (Guillot *et al.*, 2005). As recommended (The Geneland Development Group, 2012), to determine the most likely number of genetic clusters, Geneland was run without spatial information, assuming uncorrelated allele frequencies, for K = 1–20, with 100,000 iterations followed by a thinning of 100 and this was repeated 10 times.

5.2.7 Factors affecting gene flow

To explore the possible effects of landscape features on gene flow, ArcGIS 10.1 was used to calculate several distance measures: Euclidean distance, overland distance, distance along

roads, distance along roads, rivers and railways and land quality cost distance measures (Figure 5.2). The distance calculations were based on the midpoints of each transect.

A Euclidean distance matrix was obtained using the Geographic Matrix Distance Generator (Ersts, 2014). An overland distance matrix, which accounts for any additional distance between sites due to topography, was produced using the Ordnance Survey Panorama Digital Terrain Model (EDINA Digimap Ordnance Survey Service) and the ArcGIS 10.1 3D Analyst toolbox.

Digitised maps of roads, rivers and railways were obtained from the Ordnance Survey Strategi maps (EDINA Digimap Ordnance Survey Service). Pairwise distances between sites were calculated using the ArcGIS 10.1 Network Analyst toolbox, assuming gene flow occurred preferentially along the margins of roads, rivers or railways.

Producing land quality cost surfaces is more challenging since the user must define the cost of moving through each habitat type, which requires additional field data or confidence in a subjective expert judgement (Spear *et al.*, 2010). For this reason, a simple binary method of classifying habitats was initially chosen, which involved recategorising the UK Land Cover Map 2007 (Morton *et al.,* 2011) into either suitable or less suitable habitat as before. Using the Landscape Genetics toolbox (Etherington, 2011) distances between sites were calculated assuming gene flow is higher through suitable habitat. For the urban habitat, which was perhaps less uniformly suitable for wood mice than arable, further categories were defined as before: arable, urban, suburban, good quality semi-natural, poor quality semi-natural and water (See Table 5.1, rationale). Costs were assigned to each category and the analysis repeated with five different sets of potential cost values, since relative costs can have a large influence on the distance values obtained (Rayfield *et al.*, 2010) (Table 5.1).

Figure 5.2 Surface layers used to calculate distance matrices. a) The digital terrain surface for calculating overland distance. b) Road network layer. c) Roads, rivers, and railway network layer. d) Habitat shaded according to the assigned land quality categories. Transects are shown as short black lines. Figure a-c: *© Crown Copyright/database right 2014. An Ordnance Survey/EDINA supplied service.* Figure d: *Based upon LCM2007 © NERC (CEH) 2011. © Crown copyright 2007. Ordnance Survey Licence number 100017572. © third party licensors.*

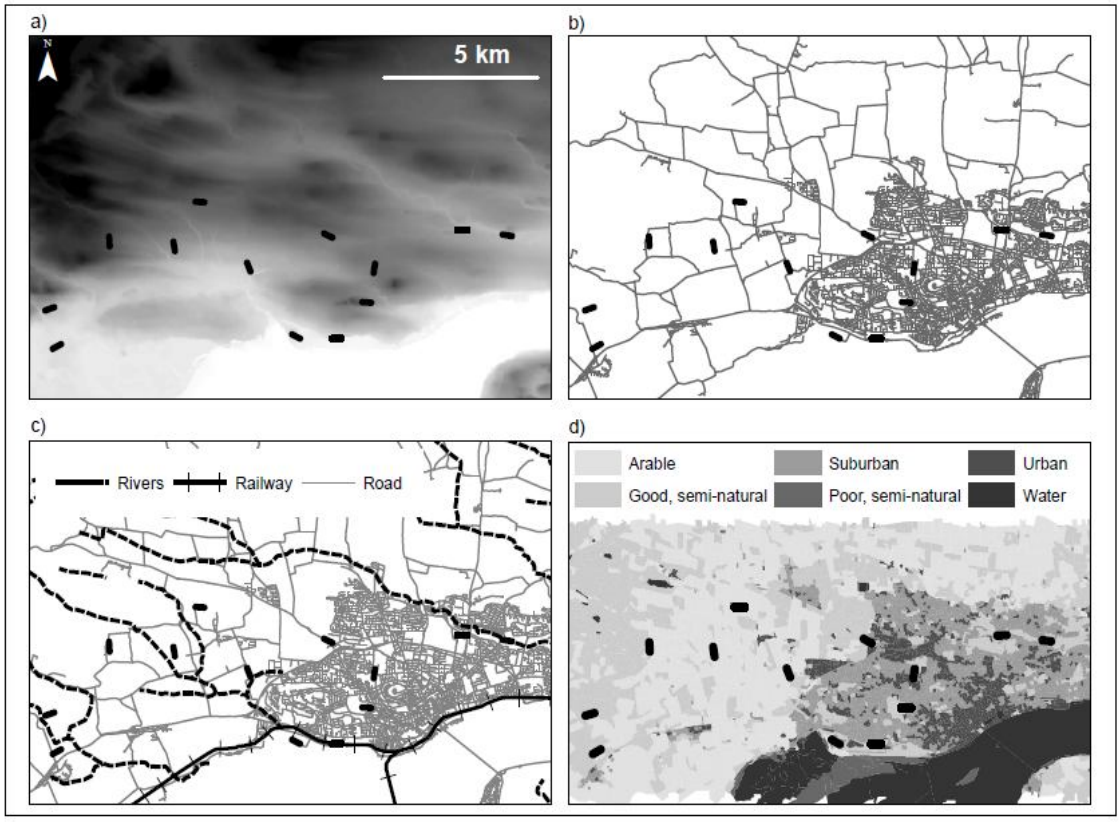


Table 5.1 Habitat categories and parameters sets used for the cost–distance land quality surfaces. Five surfaces were produced using these categories, each with a different cost parameter set as given. A rationale for the relative costs is included.

Category (LCM categories)	Cost parameters					Rationale based on studies cited
	Set 1	Set 2	Set 3	Set 4	Set 5	
Water (Estuary, sea, lakes)	1000	1000	1000	1000	1000	Where wood mice occupy islands, morphological traits suggest gene flow is limited (Berry, 1986). Reports of wood mice making use of water habitats are lacking.
Urban (Urban, urban industrial)	25	40	40	20	20	Fewer individuals were captured in Longworth traps placed outside of vegetative cover in urban habitat (Baker *et al.*, 2003) and fewer numbers were captured in disturbed urban habitat with a greater housing density, than in undisturbed urban habitat with a lower housing density (Dickman and Doncaster, 1989). Vegetative cover low.
Poor, semi-natural (Littoral mud, sand, shingle)	15	15	40	20	20	A small area of habitat around the tidal Tay Estuary, frequently underwater. Little vegetative cover or burrowing opportunities provided.
Suburban (Suburban)	5	5	15	2	1	Wood mice have utilised suburban habitats such as private gardens, cemeteries, churchyards, woodland and scrub (Baker *et al.*, 2003; Dickman and Doncaster, 1987).
Good, semi-natural (Grassland, woodland)	2	2	2	1	1	Provides vegetative cover and food resources which are important (Macdonald *et al.*, 2007). Wood mice have utilised woodland (Mallorie and Flowerdew, 1994) and grassland (Churchfield *et al.*, 1997).
Arable (Arable)	1	1	1	1	1	Wood mice have utilised arable habitat and gene flow is predicted to be high, given the balanced dispersal system believed to operate in this habitat (Tattersall *et al.*, 2004; Macdonald *et al.*, 2000; Ouin *et al.*, 2000; Ylönen *et al.*, 1991).

The correlation of these distance measures with the pairwise θ distance matrix was tested using a combination of simple and partial Mantel tests, with 50,000 permutations performed in the vegan package in R (Oksanen *et al.*, 2012; R Development Core Team, 2014). There has been controversy associated with statistical methods for examining relationships between genetic distances and geographic or landscape distance metrics (Balkenhol *et al.*, 2009; Cushman *et al.*, 2013; Guillot and Rousset, 2013; Castillo *et al.*, 2014; Smith *et al.*, 2014). The reliability of the various methods available is still being debated and no consensus has been reached (Guillot and Rousset, 2013; Castillo *et al.*, 2014; Smith *et al.*, 2014). The debate has centered around the levels of type I and type II error associated with simple and partial Mantel tests (Cushman *et al.*, 2013; Guillot and Rousset, 2013; Castillo *et al.*, 2014; Smith *et al.*, 2014), with several simulation studies reporting an elevated type I error associated with both types of test, resulting in spurious relationships between factors between identified (Cushman and Landguth, 2010a; Cushman *et al.*, 2013). There has been some suggestion that comparing the relative magnitude of the correlation coefficients between competing variables to determine the most influential variables may be adequate (Cushman *et al.*, 2013). No alternative rigorously examined approaches have been suggested (Castillo *et al.*, 2014). Regression analysis and Approximate Bayesian Computation Methods have been discussed (Jaquiéry *et al.*, 2011) but remain largely unexplored and were not pursued in this chapter since the number of data points was limited and over-fitting was considered likely. Recent studies have used combinations of simple and partial Mantel tests to partial out confounding factors – referred to as a causal modelling framework (Cushman *et al.*, 2013; Castillo *et al.*, 2014; Smith *et al.*, 2014). In general, Bonferroni corrections appear not to have been applied to accommodate the effects of multiple testing (Castillo *et al.*, 2014) but there has been suggestion that the risk of type I error could be reduced by using more stringent critical p-values (Cushman *et al.*, 2013). Here, p-values have not been adjusted using a Bonferroni correction because this would result in extremely conservative values but the results can be interpreted whilst appreciating these points.

It is notable that although correlating landscape factors with genetic differentiation is standard in landscape genetics (Manel *et al.*, 2003), inferring causation from this method requires some caution, particularly when there is scope for multiple landscape variables to be inter-correlated (Cushman and Landguth, 2010a).

5.3 Results

5.3.1 Population information

In total, 268 unique individuals were successfully genotyped from seven urban (N = 15, 16, 21, 23, 23, 24, 25) and six arable sample sites (N = 13, 14, 21, 23, 24, 26). By regenotyping 40 samples, the error rate was calculated to be 0.075 errors per allele, within the range calculated by other studies (Goossens *et al.*, 1998; Frantz *et al.*, 2003; Selkoe and Toonen, 2006). The $P_{(ID)sib}$ for each site was less than 0.0001, meaning any recaptures could be confidently identified and removed (Waits *et al.,* 2001). Micro-Checker (Van Oosterhout *et al.*, 2004) confirmed that there was no evidence for null alleles, allele-dropout or error due to stutter peaks.

Information about recaptured individuals was used to estimate a population size for each site (N_{recap} = 392). Using the Peterson method, there was no significant difference between the estimated population size of urban sites compared to arable sites (Wilcoxon rank-sum test: N = 13, W = 10, p-value = 0.136), suggesting that any genetic differences observed could not be attributed to differences in population size. Additionally, there was no significant difference in the effective population sizes of urban and arable sites (Wilcoxon rank-sum test: N = 13, W = 19.5, p-value = 0.886) (Table 5.3).

No loci were in linkage disequilibrium (P_{crit} = 0.0004) and there was a low incidence of loci deviating from HWE (MSAF8 significantly deviated from HWE at three sites), suggesting no overall departure from HWE.

5.3.2 Habitat differences

The habitat composition of a 1 km buffer around each 300m transect was calculated (Table 5.2). UK Land Cover Map 2007 habitats (Morton *et al.*, 2011) were then reclassified as suitable or less suitable and the proportion of suitable habitat was found to be significantly greater in arable than urban sites (Wilcoxon rank-sum test, N = 13, W = 42, p-value = 0.003). Three connectivity metrics, the largest patch index, the contiguity index and the percentage of like adjacencies, were found to differ significantly for urban and arable habitat (LPI: Wilcoxon rank-sum test, N = 13, W = 42, p-value = 0.003; CI: N = 13, W = 41, p-value = 0.005; PLADJ: N = 13, W = 42, p-value = 0.001). In arable habitat, the largest patch comprised a greater proportion of the total area than in urban habitat (mean arable LPI = 98.5, SD = 2.94, mean urban LPI = 72.1,

SD = 14.3), the habitat was more contiguous (mean arable CI = 0.947, SD = 0.057, mean urban CI = 0.785, SD = 0.073) and there was a higher percentage of like adjacencies (mean arable PLADJ = 99.4, SD = 0.204, mean urban PLADJ = 98.0, SD = 0.465). This suggests that connectivity was higher in the sampled arable habitat than the urban habitat, when habitat types were classified in this way.

Table 5.2 Habitat composition (in km^2) of the 1 km buffer around each 300m transect. *Based upon LCM2007 © NERC (CEH) 2011. © Crown copyright 2007. Ordnance Survey Licence number 100017572. © third party licensors.*

	Site	Arable	Urban	Suburban	Semi-natural, good quality	Semi-natural, poor quality	Water
Arable	A1	1.16	0.00	0.00	2.38	0.21	0.00
	A2	1.89	0.04	0.23	1.55	0.02	0.00
	A3	3.39	0.00	0.00	0.35	0.00	0.00
	A4	1.93	0.00	0.00	1.81	0.00	0.00
	A5	3.07	0.02	0.01	0.63	0.00	0.00
	A6	2.43	0.28	0.24	0.80	0.00	0.00
	U7	0.24	0.28	1.04	0.72	0.60	0.86
	U8	0.03	0.25	1.52	0.90	0.90	0.14
	U9	0.00	0.84	2.26	0.62	0.01	0.00
Urban	U10	0.00	1.44	1.82	0.49	0.00	0.00
	U11	0.24	0.90	0.94	1.65	0.00	0.01
	U12	0.03	0.28	1.99	1.43	0.00	0.01
	U13	0.00	0.56	2.23	0.95	0.00	0.00

5.3.3 Genetic diversity

The genetic diversity of urban and arable sites was compared using allele richness, private allele richness and heterozygosity as diversity measures (Table 5.3). Allele richness, calculated using rarefaction, was found to be significantly lower in urban sites than arable (Wilcoxon rank-sum test: N = 13, W = 42, p-value = 0.001). Private allele richness and heterozygosity were greater for arable sites but not significantly (Wilcoxon rank-sum test: private allele richness, N = 13, W = 25.5, p-value = 0.565; heterozygosity, N = 13, W = 24.5, p-value = 0.667).

F_{IS} did not differ significantly from zero at any site (Table 5.3) (Pcrit = 0.004).

Table 5.3 Sample size, rarefacted allele richness (N = 13), rarefacted private allele richness (N = 13), observed and expected heterozygosity and F_{IS} at each sample site.

	Site	N	N_E	N_A	N_P	H_O	H_E	F_{is}
Arable	A1	24	22.8 (8.4)	7.5	0.48	0.88	0.86	-0.02
	A2	23	23.8 (0.1)	7.47	0.26	0.78	0.84	0.07
	A3	13	12.9 (∞)	7.51	0.35	0.86	0.85	-0.02
	A4	21	20.8 (0.1)	7.67	0.26	0.87	0.86	-0.02
	A5	14	13.5 (∞)	7.85	0.52	0.81	0.86	0.06
	A6	26	26.0 (0.1)	7.25	0.65	0.83	0.85	0.02
Urban	U7	23	22.8 (5.5)	6.47	0.26	0.83	0.82	-0.02
	U8	25	24.8 (∞)	6.78	0.28	0.81	0.82	0.01
	U9	24	22.8 (7.6)	6.66	0.38	0.79	0.83	0.04
	U10	23	23.0 (3.0)	6.82	0.32	0.85	0.83	-0.02
	U11	16	14.9 (0.1)	6.76	0.73	0.84	0.82	-0.03
	U12	15	15.6 (5.7)	5.9	0.24	0.84	0.79	-0.06
	U13	21	20.8 (∞)	6.7	0.35	0.86	0.83	-0.03

N, sample size; N_E, effective population size (95% confidence intervals), N_A, rarefacted allele richness; N_P, rarefacted private allele richness; H_O, observed heterozygosity; H_E, expected heterozygosity.

5.3.4 Genetic differentiation

Urban sites showed a greater degree of genetic differentiation from each other than urban/arable pairs, which in turn were significantly more differentiated than pairs of arable sites (Table 5.4) (Kruskal-Wallis test, df = 2, $\chi^2 = 39.9$, p-value<0.0001; urban average pairwise $\theta = 0.035$, SD = 0.002; urban/arable average pairwise $\theta = 0.026$, SD = 0.002, arable average pairwise $\theta = 0.010$, SD = 0.001). Using 10,000 permutations, the significance of each pairwise θ value was tested (Pcrit = 0.0006). For the most part, the arable sites were not significantly differentiated from each other (1 of 14 pairs significantly differentiated), whereas the majority of urban sites were significantly differentiated from each other (16 of 21 pairs) (Table 5.4).

Pairwise θ values were illustrated on a neighbor-joining tree using Splitstree (Fig. 5.3) (Huson and Bryant, 2006; Saitou and Nei, 1987), alongside NJ-trees using shared allele distances and the Cavalli-Sforza chord distances. There was agreement between the three NJ-trees, with all arable sites except for one falling on the same branch of the tree. In contrast, the urban sites appeared on separate branches, reflecting the greater genetic differentiation described above. The respective positions of the arable sites within the NJ tree were consistent with their geographic locations.

Although pairs of urban sites were shown to be significantly genetically differentiated, the Bayesian-clustering methods of Structure best assigned individuals to a single genetic cluster (Pritchard *et al.*, 2000). The Evanno plots showed the posterior probability to slope downward from K = 1 rather than rising to an asymptote, suggesting that K = 1 was most likely (Evanno *et al.*, 2005; Earl and vonHoldt, 2012). It was possible that population substructure, although evident from a consideration of pairwise θ values, was not detected by Structure because of limited sample sizes and relatively low levels of genetic differentiation (Smith and Wang, 2014). Alternatively, one cluster may have been the uppermost level of hierarchical structure, with further levels of structure being present. A cluster number of one was also suggested for all 10 repeat runs when the Bayesian-clustering methods of Geneland were implemented (Guillot *et al.*, 2005). Further levels of hierarchical substructure were not detected when Structure was run separately for urban and arable sites, nor when the urban populations with significant inbreeding were removed.

Table 5.4 Pairwise θ values below the diagonal and significance levels for those pairwise θ values above the diagonal. Asterisks indicate significant differentiation and 'ns' indicates a non-significant differentiation between that pair.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**U7**	**U8**	**U9**	**U10**	**U11**	**U12**	**U13**
A1	-	ns	ns	ns	ns	*	ns	ns	*	*	*	*	*
A2	0.008	-	ns	ns	ns	ns	*	*	*	*	*	ns	*
A3	0.008	0.002	-	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
A4	0.008	0.004	0.000	-	ns	ns	ns	*	*	ns	ns	ns	*
A5	0.014	0.012	0.003	0.004	-	ns	ns	ns	ns	ns	ns	*	ns
A6	0.023	0.022	0.019	0.016	0.011	-	*	ns	*	*	*	*	ns
U7	0.031	0.026	0.021	0.018	0.023	0.028	-	ns	ns	*	*	*	*
U8	0.021	0.028	0.023	0.025	0.019	0.018	0.019	-	*	*	*	*	*
U9	0.041	0.027	0.025	0.030	0.029	0.021	0.019	0.026	-	*	ns	*	*
U10	0.031	0.025	0.018	0.020	0.013	0.022	0.023	0.033	0.029	-	ns	*	*
U11	0.045	0.029	0.022	0.020	0.022	0.029	0.034	0.055	0.038	0.030	-	ns	*
U12	0.051	0.031	0.029	0.034	0.035	0.038	0.040	0.051	0.030	0.046	0.035	-	*
U13	0.025	0.023	0.026	0.024	0.020	0.020	0.032	0.027	0.032	0.029	0.044	0.039	-

Asterisks indicate significance level: ns = non-significant, $p < 0.0006$ *

Figure 5.3 Neighbor-joining trees produced using inter-population genetic distance measures showing a) Pairwise θ values, b) DAS, shared allele distances and c) Cavalli-Sforza chord distances. Arable sites are circled.

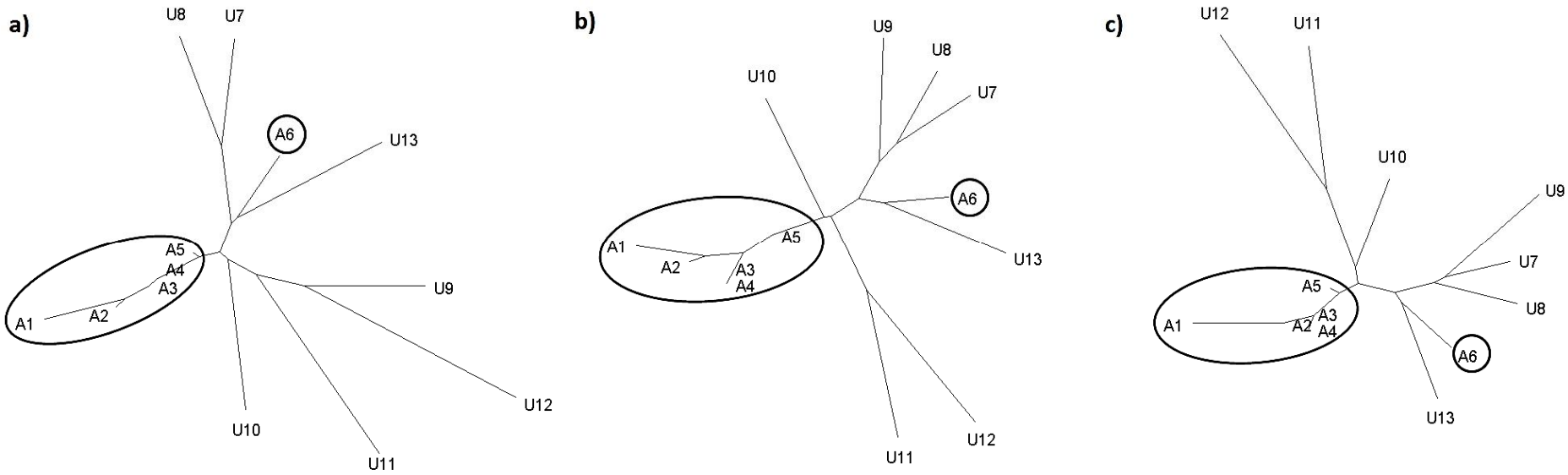
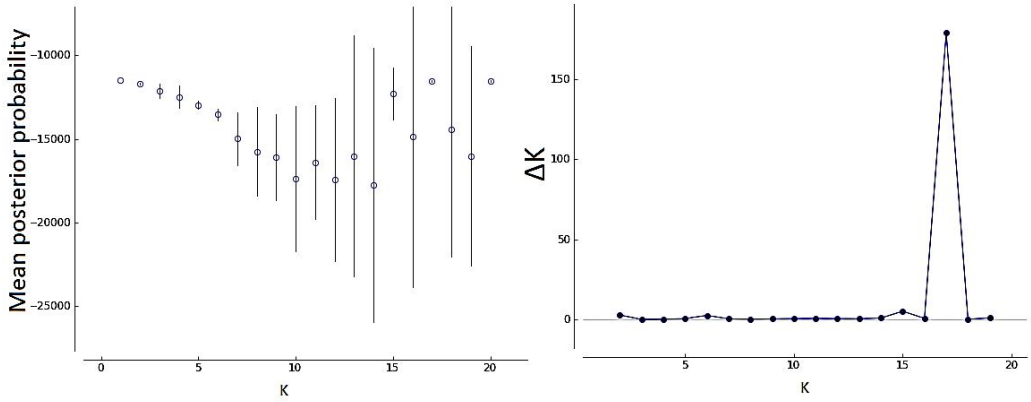


Figure 5.4 Evanno plots showing a) the plot of the posterior probabilities against number of clusters. b) ΔK against the number of clusters. Figures taken from Structure Harvester.



5.3.5 Factors affecting gene flow

Following observation of differences in genetic differentiation under urban and arable land use, the influence of landscape effects on gene flow between sites was tested (Table 5.5). In the arable landscape, overland distance showed the greatest correlation with pairwise θ values, whereas in the urban landscape, measures that accommodated possible differences in land quality were most highly correlated with pairwise θ values (Table 5.5). When simple Mantel tests were used, the correlation between Euclidean distance or overland distance and genetic distance was significant within the arable habitat but not in the urban habitat (Table 5.5). By partialling out Euclidean distance, to control for its effect, the primary influence of overland distance in arable habitat was demonstrated to be likely. The distance along roads, railways or rivers did not significantly explain genetic differentiation patterns in either habitat (Table 5.5).

Habitat was classified as either suitable or less suitable and this metric significantly explained arable genetic differentiation when simple Mantel tests were used, but probably because this distance was almost akin to Euclidean distance in arable habitat. When overland and Euclidean distance were partialled out, to control for their effect, this land quality metric was no longer significantly correlated with genetic differentiation. For urban habitat, metrics that accommodated differences in land quality had higher correlation coefficients than the alternative distances investigated (Table 5.5). Simple Mantel tests suggested that these land quality metrics had an almost significant effect on genetic differentiation but when Euclidean and overland distance were partialled out, they were less significant.

Table 5.5 Distance metrics investigated alongside the Spearman's correlation coefficient and p-values from simple and partial Mantel tests (Euclidean and overland distance partialled out respectively).

Distance metric	Arable (N = 15)		Urban (N = 21)	
	r	p-value	r	p-value
Simple Mantel tests				
Euclidean distance	0.550	0.015*	0.271	0.126
Overland distance	0.639	0.004**	0.271	0.125
Road network distance	0.318	0.111	0.271	0.131
Roads, rivers, railways distance	0.318	0.109	0.271	0.131
Land quality: suitable or less suitable	0.532	0.014*	0.305	0.110
Land quality: Parameters 1	-	-	0.348	0.076
Land quality: Parameters 2	-	-	0.348	0.069
Land quality: Parameters 3	-	-	0.342	0.075
Land quality: Parameters 4	-	-	0.346	0.073
Land quality: Parameters 5	-	-	0.326	0.074
Partial Mantel tests (| Euclidean distance)				
Overland distance	0.407	0.020*	-0.039	0.563
Road network distance	-0.218	0.785	0.150	0.259
Roads, rivers, railways distance	-0.218	0.781	0.150	0.256
Land quality: suitable or less suitable	0.285	0.068	0.142	0.259
Land quality: Parameters 1	-	-	0.319	0.124
Land quality: Parameters 2	-	-	0.319	0.122
Land quality: Parameters 3	-	-	0.303	0.140
Land quality: Parameters 4	-	-	0.249	0.178
Land quality: Parameters 5	-	-	0.227	0.186
Partial Mantel tests (| Overland distance)				
Euclidean distance	0.163	0.197	0.040	0.443
Road network distance	-0.625	0.978	0.150	0.257
Roads, rivers, railways distance	-0.625	0.979	0.150	0.260
Land quality: suitable or less suitable	-0.105	0.553	0.142	0.258
Land quality: Parameters 1	-	-	0.319	0.122
Land quality: Parameters 2	-	-	0.319	0.123
Land quality: Parameters 3	-	-	0.303	0.137
Land quality: Parameters 4	-	-	0.250	0.180
Land quality: Parameters 5	-	-	0.227	0.188

Asterisks indicate significance level: $p < 0.05$ *, $p < 0.01$ *

5.4 Discussion

The aim of this chapter was to investigate whether arable and urban habitat provided barriers to gene flow for wood mice. The local genetic diversity and gene flow between wood mouse populations was compared between urban and arable habitat, two of the most extensively human-modified habitats inhabited by wildlife. Insignificant genetic differentiation between pairs of arable sites implied that arable habitat was not a barrier to gene flow. Using three inter-population genetic distance metrics gene flow was demonstrated to be greater between arable than urban sample sites. Genetic diversity measures were calculated and compared for the two land use types, and a higher genetic diversity was reported for arable habitat with varying levels of significance. In arable habitat, patterns of genetic differentiation between populations were best explained by overland distance but this distance poorly explained patterns in urban habitat. Instead, metrics that accommodated differences in land quality better explained patterns of urban gene flow but with non-significant correlations. The work of this chapter represents the first direct genetic comparison of urban and arable sample sites using a randomised sample design and it develops the understanding of population genetic structure of wild species within these systems.

Despite the generalist ecology of wood mice (Macdonald *et al.*, 2000) and their ability to readily use portions of urban habitat, such as parkland, cemeteries and domestic gardens (Baker *et al.*, 2003), populations sampled from urban sites separated by 4.2 km on average were more genetically differentiated than in arable habitat. Other studies considering abundant mammals within cities have also reported genetic differentiation between populations, which could have implications when seeking to control pest species existing within metapopulations (Gardner-Santana *et al.*, 2009). For example, almost all of 14 white-footed mouse populations sampled throughout New York City were from genetically distinct groups (Munshi-South and Kharchenko, 2010), urban populations of striped field mice were more differentiated than populations sampled in lakeside woodland in Poland (Gortat *et al.*, 2013), Norway rats sampled at 11 sites in Baltimore belonged to three genetically distinct groups (Gardner-Santana *et al.*, 2009), and genetic differentiation was greater between two urban red fox sampling sites in Zurich than between three rural populations separated by man-made barriers or rivers (Wandeler *et al.*, 2003). Despite genetic isolation-by-distance patterns being common for mammals (Aars *et al.*, 2006; Pope *et al.*, 2006; Gauffre *et al.*, 2008), the analysis of correlations between several distance metrics and genetic differentiation suggested

that metrics accommodating land quality performed better than Euclidean or overland distance. This is in line with the finding that undisturbed urban habitats with vegetative cover have been shown to be preferred, with small mammal capture success declining with distance from natural or semi-natural patches (Dickman and Doncaster, 1989; Baker *et al.*, 2003). As suggested by Landguth *et al.* (2012), perhaps with greater sampling, a greater number of alleles, or greater allele variation, a stronger correlation between land quality metrics and genetic differentiation would have been detected.

Arable habitat, in contrast, had significantly lower inter-population genetic differentiation, suggesting that gene flow between arable sites was higher than in urban habitat. Five of the six arable sites fell on the same branch of all three NJ-trees, compared to the urban sites which were on separate branches. This suggested that arable habitat did not act as a significant barrier to wood mouse movement and gene flow. Similarly, Gauffre *et al.* (2008) found that common voles sampled over a 500 km^2 area belonged to a single genetic group, and the authors proposed that agricultural disturbance may promote widespread gene flow above baseline levels. Gene flow was lowest through agricultural habitat and highest through urban habitat for white-footed mice in a recently published study (Marrotte *et al.*, 2014). Directly testing the hypothesis that agricultural disturbance promotes gene flow above background levels would be challenging, since agricultural practices are not easily manipulated on large scales but the results of this chapter are not inconsistent with the hypothesis. Consistent with the suggestion that agricultural processes promote dispersal, are the numerous reports of seasonal habitat preference shifts between crop and margins for this species (Ylönen *et al.,* 1991; Macdonald *et al.*, 2000; Ouin *et al.,* 2000; Butet *et al.*, 2006), including the findings of Chapter 2. However, although arable habitat may be sufficiently connected to counteract genetic differentiation for this generalist species, the landscape is likely to appear more fragmented to species with specialist requirements.

Previous studies have suggested that roads, as common man-made features within both landscapes, may hinder gene flow (Richardson *et al.*, 1997) but additional work has also suggested that the margin habitat of these features (and presumably also river and railway margins), provide useful habitat in otherwise unfavourable landscapes (Bellamy *et al.*, 2000). These habitats could therefore act both positively and negatively on gene flow. However, in the present study there was no evidence that the route of gene flow occurred predominantly via road verge corridors, since there was no significant correlation between along road or road,

river and railway distances and genetic differentiation, while other distance metrics explained the variation in differentiation better. Likewise, there was no evidence that roads significantly impeded gene flow with populations from either side of the major trunk road in Dundee appearing together on NJ tree branches. To improve confidence in this finding it would be useful to survey a selection of the road, river and railway features included in the OS rasters, to determine the extent to which these features explicitly provided suitable habitat for wood mice. The OS rasters could perhaps be edited to incorporate information about habitat quality and the question then revisited. Additional sampling may also prove useful in detecting weaker correlations.

The overall pattern of lower genetic diversity in urban sites than arable sites and the evidence of inbreeding in some of these are consistent with reduced gene flow (Frankham, 2005). Allele richness was found to differ significantly between urban and arable habitat but heterozygosity and private allele richness were lower in urban habitat but not significantly, suggesting that the influence of fragmentation on genetic diversity was not extreme. As for Munshi-South and Kharchenko (2010), the heterozygosity of urban populations was reasonably high and above expected levels (above 0.9 in all cases). Previously, studies have reported lower genetic diversity in urban habitat for some species (Hitchings and Beebee, 1998; Wandeler *et al.*, 2003) but not for others (Chiappero *et al.*, 2011; Gortat *et al.*, 2013). One possible explanation for conflicting findings is that, when fragmentation occurs, there is a lag between genetic differentiation and the decline of genetic diversity (Keyghobadi *et al.*, 2005). An alternative explanation for unexpectedly high genetic diversity is that genetic material may be supplemented from adjacent arable habitat around the city or as a result of human-assisted translocation, as is anecdotally common for pest species.

In conclusion, in this chapter, arable habitat was demonstrated to be sufficiently connected to permit gene flow between wood mouse populations. In contrast, urbanisation acted to fragment the landscape for this species, with sampled populations being significantly differentiated. The results provided additional insight into the landscape genetics of this wild species within arable and urban habitats.

Chapter 6

General discussion

6.1 Addressing the thesis aims

This thesis examined the landscape genetics of a small mammal species in arable habitat, in an attempt to provide a basis for future agricultural landscape genetics studies. Two questions were of particular interest. Firstly, the possibility of both spatial *and* temporal variation in fine scale genetic structure was explored for the most abundant small mammal species in the region. Secondly, the extent to which urban and arable habitats provided barriers to gene flow was considered. Small mammals were chosen as the study organism because they have been called model organisms, they are abundant and common, they are easy to sample at appropriate scales and they may have important ecological roles themselves (Section 1.5).

To identify a suitable study species, in Chapter 2, an example arable field site was surveyed. The most abundant small mammal species was found to be the wood mouse. As for other studies investigating this species, during the growing season, wood mice were found to make greater use of cropped habitat than margin habitat but they switched to making greater use of margins after crop harvesting (Pollard and Relton, 1970; Ylönen *et al.,* 1991; Macdonald *et al.,* 2000; Ouin *et al.,* 2000; Butet *et al.,* 2006). In contrast, the second most abundant species, the bank vole, continuously made greatest use of more stable marginal habitat. Previously, wood mice were shown to have both fine scale and larger scale genetic structure in non-arable habitats (Booth *et al.*, 2009) but the landscape genetics of this species had not been previously examined in arable systems.

To address landscape genetics questions, in Chapter 3, a method was developed for extracting DNA from wood mouse hair samples and for genotyping individuals using microsatellite markers. The key insights obtained from the landscape genetics studies of Chapter 4 and 5 are discussed below.

6.2 Key insights

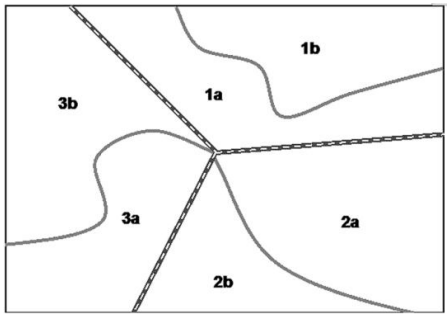
6.2.1 Fine scale population genetic structure

In Chapter 4, fine scale genetic structure was demonstrated for wood mice living within arable habitat. This finding of fine scale population genetic structure is likely to be also applicable across other habitat types, in addition to arable.

Three genetically distinct wood mouse clusters were identified across a 42 hectare site in Chapter 4. Additionally, there was significant spatial overlap between the clusters. Bayesian-clustering software has previously been demonstrated to be useful in describing population structure at various scales, from landscape and regional through to country scale (Booth *et al.*, 2009; Ingvarsson and Olsson, 1997; De Barro, 2005; Arens *et al.*, 2006; Gauffre *et al.*, 2008). More recently, it was used to identify fine scale population genetic structure in wood mice in pastoral and woodland habitat, with sample sites separated by 850m (Booth *et al.*, 2009), and for common vole populations separated by 330–2560 m (Schweizer *et al.*, 2007). Given the lack of obvious geographic barriers to gene flow at the sample site, as for Booth *et al.* (2009) and Schweizer *et al.* (2007), structuring by social processes appeared to be a plausible explanation for genetic structure at this scale.

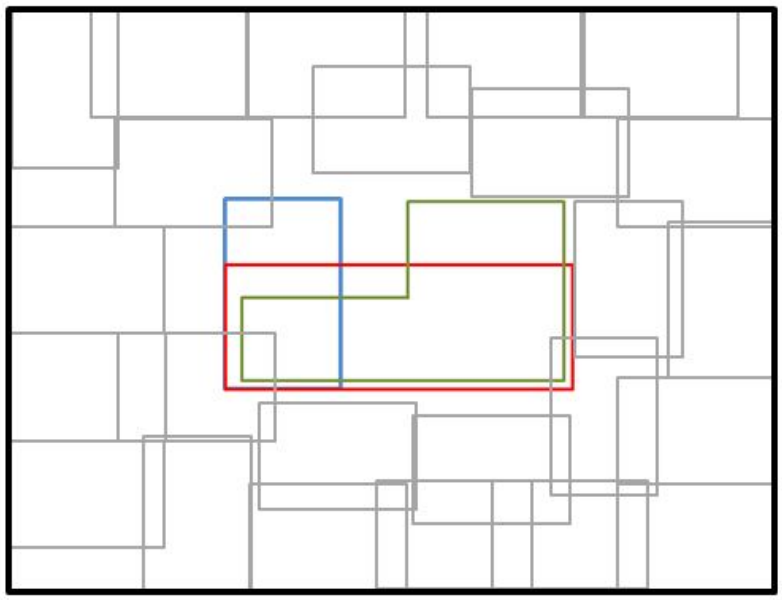
Relationships between fine scale and larger scale population structure have been described using various metapopulation models (Evanno *et al.*, 2005). As described in Chapter 4 and Chapter 5, Evanno *et al.* (2005) developed a method for determining the uppermost level of hierarchical population genetic structure across sampled areas. They also discussed two alternative models of population structuring: the island model and the contact zone model (Kimura and Weiss, 1964; Slatkin and Voelm, 1991; Evanno *et al.*, 2005). The main difference between the three models was in their within and between population migration rates (Evanno *et al.*, 2005). Several studies have demonstrated hierarchical population structure patterns in genetic variation and the hierarchical model is thought to frequently describe observed population structure patterns (Chapuisat *et al.*, 1997; Giles *et al.*, 1998, Bouzat and Johnson, 2004; Dionne *et al.*, 2008; Balkenhol *et al.*, 2014). A figure derived from Balkenhol *et al.* (2014) is given in Figure 6.1 to illustrate the hierarchical genetic structuring pattern that the authors suggested for cougars (*Puma concolor*).

Figure 6.1 Figure taken from Balkenhol *et al.* (2014) illustrating a possible hierarchical population genetic structure for sampled cougars (*Puma concolor*). The dashed lines illustrate an uppermost level of hierarchical structure with further substructuring at a lower level.



A similar hierarchical model could explain wood mouse population structure, although spatial overlap between populations must be incorporated. Fine scale genetic structure was evident in Chapter 4, when three genetically distinct groups were identified across a small area with some spatial overlap. However, when populations were sampled across a larger spatial scale, individuals were assigned to a single genetic cluster (i.e. one was the uppermost level of structure). The possible hierarchical model explanation of wood mouse population structure is illustrated by a schematic in Figure 6.2. This could be confirmed by additional sampling at various scales.

Figure 6.2 Schematic showing the possible hierarchical population genetic structure of wood mouse populations in the arable region. Populations sampled in Chapter 4 are given by the coloured shapes (cluster A = blue, cluster B = red, cluster C = green) and two of these were noted to be transient. A continuation of this pattern is assumed and possible non-sampled fine scale genetic clusters are shown with grey lines. The black line denotes a possible higher level of hierarchical structure, perhaps identified in Chapter 5.



6.2.2 Temporal variation in genetic structure

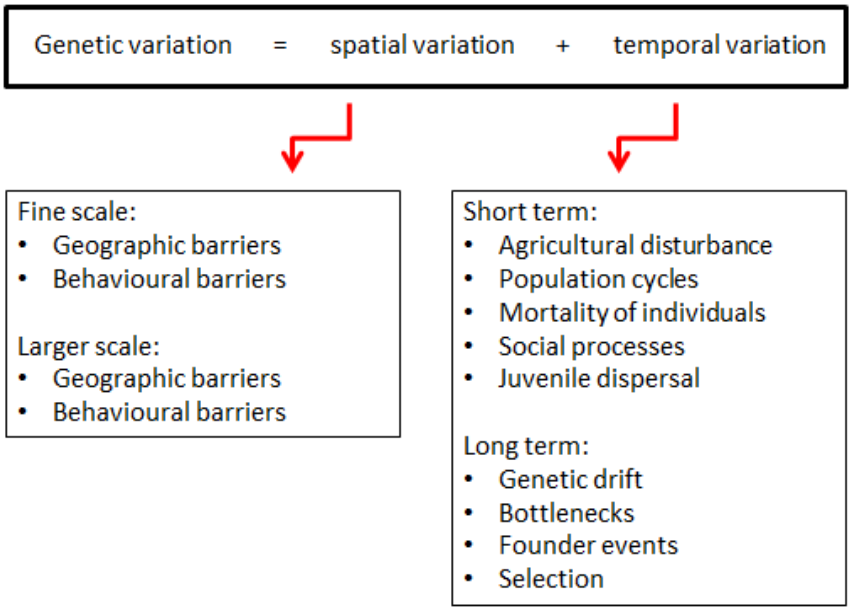
Perhaps the most important finding was that spatial genetic structure varied over a short time scale. Short scale temporal variation had not been demonstrated previously for agricultural habitat but it has occasionally been investigated for other habitat types and situations. For example, the effect of the exploitation of animal populations (i.e. hunting), population cycles and juvenile dispersal have been explored (Nussey *et al.*, 2005; Berthier *et al.*, 2006; Schweizer

et al., 2007; Piertney *et al.*, 2008; Ehrich *et al.*, 2009; Pilot *et al.*, 2010; Nichols *et al.*, 2012; Rikalainen *et al.*, 2012).

Here, short scale temporal variation was hypothesised given that fine scale population structure was previously demonstrated for wood mice (Booth *et al.*, 2009) and given the evident change in the spatial arrangement of wood mouse individuals demonstrated in Chapter 2. This finding suggests that, in addition to the influence of agricultural disturbance on the spatial distribution of individuals and the persistence of individuals, disturbance may also drive changes in spatial genetic structure.

For wood mice and perhaps for other wild species occupying arable or other disturbed habitats, genetic variation may be better understood by separating it into spatial *and* temporal components, and the respective processes which operate at different scales within each component. This is illustrated by the schematic in Figure 6.3. Temporal variation has often been neglected by ecological genetic studies (Heath *et al.*, 2002; Nussey *et al.*, 2005; Nichols *et al.*, 2012) but it has been identified as being an area that requires further consideration (Balkenhol *et al.*, 2009; Anderson *et al.*, 2010; Cushman and Landguth, 2010b; Landguth *et al.*, 2010; Balkenhol and Landguth, 2011). In support, the results of Chapter 4 suggested that only by incorporating a temporal component in the sampling regime, could an accurate understanding of wood mouse fine scale population structure in arable habitat be obtained.

Figure 6.3 Schematic showing how genetic variation could be divided into spatial and temporal components, and listing a range of possible factors that might influence each component of genetic variation in the fine scale, larger scale, short term and long term.



Future studies, especially in disturbed habitats, should consider accommodating possible temporal variation in genetic structure by sampling at multiple time points as demonstrated in Chapter 4. If not possible due to time or financial restraints, then as a less satisfactory alternative, sampling could take place over a short interval, during the growing season when spatial genetic structure was demonstrated to remain stable for this species. This was the method used in Chapter 5 due to time constraints. Extrapolation between time points should be made with caution, and samples collected over longer time periods should ideally not be grouped and assumed to represent a single genetic snapshot. It is also possible that this finding could extend to other habitat types that are regularly disturbed.

6.2.3 Arable habitat: a barrier to gene flow?

Previously landscape genetics studies qualitatively or quantitatively addressing the influence of agricultural habitat on population genetic structure have given mixed results. For some species, agricultural habitat has been suggested to act as a barrier to gene flow (Cegelski *et al.*, 2003; Coulon *et al.*, 2004; Lindsay *et al.*, 2008), but for other species this did not appear to be the case (Johansson *et al.*, 2005; Purrenhage *et al.*, 2009; Schmidt *et al.*, 2009). In Chapter 5, there was no evidence that arable habitat acted as a barrier to gene flow between wood mouse populations. Samples sites were not significantly differentiated in terms of Weir and Cockerham's pairwise θ fixation index and in comparison to urban habitat, samples sites were less differentiated. Additionally, all arable sample sites, except for one, fell on the same branch in the NJ tree. It would appear that agricultural habitat does not act as a barrier to gene flow for all species and Gauffre *et al.* (2008) have suggested that gene flow could even be promoted in arable habitat due to extreme and frequent agricultural disturbance, which could force the regular dispersal of individuals. The extent to which arable habitat acts as a barrier is likely to depend on the habitat requirements of the species under examination.

The use of a comparative approach for addressing this question was demonstrated in Chapter 5. As discussed, this approach is helpful in controlling for genetic differentiation between sample sites due to social organisation. Since genetic structure was demonstrated at a fine scale in Chapter 4, and for other studies (Schweizer *et al.*, 2007; Booth *et al.*, 2009), a comparative approach provides confidence that levels of genetic differentiation measured, are not solely due to social organisation processes. Studies should be interpreted with caution if they report that agricultural habitat acts as a barrier to gene flow having only sampled in this habitat type.

6.2.4 A caveat: specialists versus generalists

In Chapter 2, wood mice were shown to be more generalist in their habitat use than bank voles. Bank voles were more restricted to field margin habitat and, possibly as a result of this, their spatial arrangement remained constant through the harvest period. In contrast, wood mice made use of the majority of available crop types and margins but switched their preference from crop to margin after harvest, alongside a shift in the spatial arrangement of individuals. Habitat use differences were therefore evident and results cannot be easily extrapolated from wood mice to bank voles.

These habitat use differences are likely to have genetic implications. Since the spatial arrangement of bank voles remained more stable throughout the cropping cycle, one hypothesis would be that their genetic structure may also remain more stable over time. Furthermore, since bank voles were more restricted to the more stable margins, it could be hypothesised that gene flow would occur preferentially along suitable margin habitat, and perhaps greater fragmentation between distant sample sites would be observed, because this habitat type is not always well connected.

This distinction between specialist and generalist species may also partly explain the mixed results from previous studies that questioned the role of agricultural habitat as a barrier to gene flow. This may have been best demonstrated by Goldberg and Waits (2010) study, who showed that agricultural habitat provided least resistance to gene flow for Columbia spotted frogs and more resistance to long-toed salamanders, whose gene flow occurred preferentially along moisture gradients (presumably more habitat specialist for this reason). Common voles, like wood mice, are also known to make use of agricultural habitat, especially during population outbreaks (Zapletal *et al.,* 2001), and therefore, it is perhaps unsurprising that a single genetic cluster was identified by Gauffre *et al.* (2008) over a 500 km^2 area.

Care should be taken when extrapolating results from one species to another. The results presented here are likely to be most applicable to other generalist species, that like wood mice, are vagile and make use of cropped habitat but even then, habitat requirements should be carefully considered.

6.4 Future directions

To further investigate the landscape genetics of wild species in arable habitat, several future research topics would be useful.

6.4.1 Landscape genetics: specialists versus generalists

Moving forward with arable landscape genetics, it could be useful to draw some generalities across species because it would be costly and time consuming to repeat studies for a large selection of species. Chapter 2 highlighted the alternative habitat use strategies of a generalist (wood mice) and a specialist (bank voles) in arable habitat and a distinction between these two broad groups could be useful. It is important to understand the landscape genetics of those with more specialist habitat requirements so that these species can be accommodated in biodiversity conservation measures.

Given the alternative habitat use strategies, it is necessary to investigate the possibility of temporal variation in spatial genetic structure for a more specialist species, such as the bank vole. In the study region, the number of bank voles captured may not be enough to identify subtle fine scale population structure so an alternative study site may need to be chosen. An alternative study for addressing this question for more specialist species may be able to use common shrews or an amphibian or bird species, after choosing an appropriate spatial scale for these organisms. Species that specialise on stable portions of habitat within arable landscapes would be hypothesised to have a more stable genetic structure but this remains to be tested.

Specialist species would also be hypothesised to have more fragmented populations and gene flow would be predicted to occur preferentially along suitable habitat. It may be possible to predict landscape genetic patterns by using more traditional ecological surveying approaches to determine habitat requirements of specialist species.

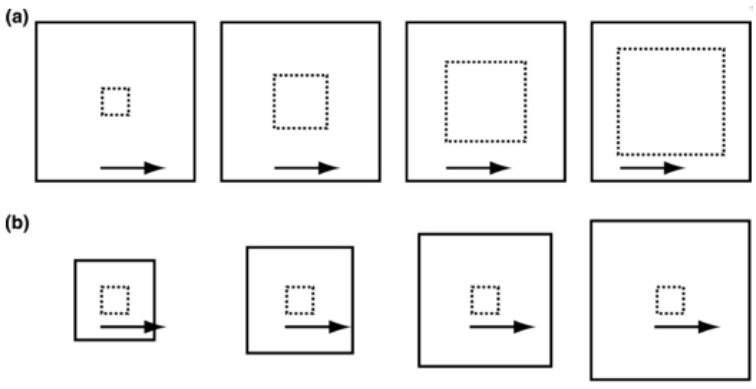
6.4.2 The temporal component of genetic variation

Given the results of Chapter 4, more research into the temporal component of genetic variation would be insightful. Additional research into factors determining the detection of any short term temporal genetic variation would also be useful, for example, scale, sample size and the magnitude of genetic differentiation.

The relative importance of spatial and temporal processes influencing genetic variation will change with spatial scale. This has been acknowledged by several authors who have highlighted, mostly with reference to various spatial scales, that sampling should take place at the scale at which possible processes structuring genetic variation are likely to operate (Cushman and Landguth, 2010b; Manel *et al.*, 2010). The need for appropriate sampling regimes at an appropriate spatial scale has been empirically demonstrated by Murphy *et al.*(2010), who reported that, for western toads (*Bufo boreas*), factors related to connectivity provision varied in importance with spatial scale. Appropriate temporal scales for sampling should be given further consideration.

As two aspects of scale, several authors have suggested that grain size and the extent of sampling in landscape genetic studies are influential in the detection of landscape genetic patterns (Figure 6.4) (Anderson *et al.* 2010; Cushman and Landguth, 2010b). The grain describes the smallest unit of sampling, whereas the extent describes the total area sampled (Anderson *et al.*, 2010). These two aspects of spatial scale can be translated to temporal scales; grain would describe the smallest unit of sampling in space and extent could relate to the length of time over which sampling occurs. Both of these were shown to have an effect on landscape genetic analyses by Cushman and Landguth (2010b), when spatial genetic patterns across a landscape were simulated and the grain and extent of sampling altered to determine the effect on the pattern–process interaction. These two aspects in a temporal context should be considered in future landscape genetics studies.

Figure 6.4 Schematic taken from Anderson *et al.* (2010) illustrating the difference between a) grain and b) spatial extent. The dotted lines show the grain of sampling and the solid line shows the extent for a sequence of sampling scales.



The exact influence of a mismatch between the temporal scale of sampling and the scale of processes structuring genetic variation is largely unknown (Anderson *et al.*, 2010; Cushman and Landguth, 2010b) but it could have various non-intuitive effects on the detection of patterns in genetic variation. For example, there could be a point at which a spatial process (such as geographical separation) is overwhelmingly dominant, leading to it being solely responsible for an uppermost level of hierarchical genetic structure, resulting in temporal variation by short term processes not being detected. Additionally, Manel *et al.* (2010) suggested that inadequate sampling at the appropriate scales could alter the noise to signal ratio, meaning that genetic patterns would be less obvious and difficult to interpret. Other outcomes of a sampling-process scale mismatch are possible and further work would be needed to identify the effect of temporal variation on detecting both spatial and temporal genetic variation patterns at different scales.

Additional factors that could affect the possibility of detecting temporal variation in genetic structure are sample size and the extent of genetic differentiation between populations. Smith and Wang (2014) recently used simulations to demonstrate that the ability of Bayesian-clustering software to detect subtle genetic clusters was influenced by these two factors. Adequate sampling, guided by Smith and Wang's (2014) study, at the appropriate grain would improve the chance of detecting patterns but it could be helpful if future studies could acknowledge temporal differentiation in investigations about appropriate sample sizes.

To gain a more comprehensive understanding of the relative importance of temporal and spatial genetic variation and factors determining the detection of a temporal component to genetic variation, additional landscape genetic field studies at multiple spatial and temporal scales could be insightful. For demonstration purposes, further simulation studies are also likely to be helpful since noise can be more readily controlled and variables can be altered easily, without the need for time and cost intensive field sampling at multiple scales.

6.4.3 Relative effects of agricultural processes on genetic variation

The possible role of agricultural disturbance in altering spatial genetic structure was discussed in Chapter 4, since the shift in spatial arrangement of individuals occurred alongside the harvesting of crops. Initially, it would be useful to introduce a control site to address this definitively.

If agricultural disturbance was one of the main drivers behind the temporal variation in spatial genetic structure observed in Chapter 4, it would be interesting to compare the relative disturbance effects of different agricultural management processes, such as ploughing, harvesting, sowing, and application of pesticides and fertiliser. It would also be insightful to examine the extent to which different features of adequate field margins can offset temporal variation, for example, margin width.

In Chapter 4, the fate of cluster A, B and C differed and it would be interesting to determine the extent to which temporal variation in population structure is random or deterministic. If it was partly deterministic, then factors predicting the outcome of disturbance would be worth investigating. This research would be challenging and costly since intensive sampling at multiple field sites would be required in order to compare different treatments. The timing of agricultural processes would also need to be tightly controlled on a large scale so that the effect of disturbance processes acting on a larger scale, did not affect the genetic structure of the field site being examined.

6.4.4 Agri-environment schemes and landscape genetics

There has been some move to consider whether agri-environment schemes improve the connectivity of agricultural landscapes (Donald and Evans, 2006). Such schemes may influence the landscape genetics of wild species in arable habitat, and the extent to which they might improve connectivity for both specialist and generalist species deserves additional consideration. However, because agri-environment schemes are often short term, there may be a lag between implementation of schemes and their influence on landscape genetics being detected, since lags in genetic variation have been demonstrated to occur (Keyghobadi *et al.*, 2005). This would make it challenging to attribute certain landscape genetic patterns to agri-environment schemes. However, Landguth *et al.* (2010) recently used individual-based simulations to demonstrate that the effects of removing a barrier to gene flow could be detected in as few as 15 generations for species with high dispersal capabilities; approximately 3–4 years for wood mice.

To test the possible influence of agri-environment schemes on landscape genetics, knowledge of the schemes being implemented across a landscape scale would be necessary and this would require extensive discussions with landowners. Schemes implemented in different farms and locations are also likely to vary in their quality and in the benefits they provide to

wildlife, which could generate noise around underlying landscape genetics patterns, again making it challenging to ascribe cause and effect.

Landscape genetics work combined with radio-telemetry studies and other field-based ecological methods may be most insightful for the above reasons, such as the methods described by Tew *et al.* (2000), Reid *et al.* (2007) and Fuentes-Montemayor *et al.* (2011) for considering the use of agricultural habitat by a selection of mammal species.

6.5 Outlook

This thesis has demonstrated that a landscape genetics research approach can provide key insights into the ecology of wild species living in agricultural habitats. The landscape genetics approach enabled novel questions to be investigated, temporal variation in fine-scale population structure could be examined and the connectivity of landscapes could be appraised. The present thesis focused on small mammals but the results obtained may have more generally applicable implications, providing insight for future landscape genetics work. It is hoped that the work of this thesis, will encourage greater use of landscape genetics approaches for improving the understanding of the ecology of wild species in human-modified agricultural habitat.

References

Aars J, Dallas JF, Piertney SB, Marshall F, Gow JL, Telfer S, Lambin X (2006) Widespread gene flow and high genetic variability in populations of water voles *Arvicola terrestris* in patchy habitats. Molecular Ecology, 15, 1455-1466.

Aavik T, Edwards PJ, Holderegger R, Graf R, Billeter R (2012) Genetic consequences of using seed mixtures in restoration: A case study of a wetland plant *Lychnis flos-cuculi*. Biological Conservation, 145, 195-204.

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. Trends in Ecology and Evolution, 23, 327-337.

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program Structure. Molecular Ecology Resources, 8, 1219-1229.

Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James PMA, Rosenberg MS, Scribner KT, Spear S (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. Molecular Ecology, 19, 3565-3575.

Applied Biosystems (2002) Applied Biosystems 3730/3730xl. DNA Analyzers. Sequencing Chemistry Guide. Accessed online (15.10.13): http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_041258.pdf Accessed 14th December 2013

Arens P, Bugter E, van't Westende W, Zollinger R, Stronks J, Vos CC, Smulders MJM (2006) Microsatellite variation and population structure of a recovering Tree frog (*Hyla arborea* L.) metapopulation. Conservation Genetics, 7, 825-835.

Arif IA, Khan HA, Shobrak M, Al Homaidan AA, Al Sadoon M, Al Farhan AH (2010) Measuring the genetic diversity of Arabian Oryx using microsatellite markers: implication for captive breeding. Genes and Genetic Systems, 85, 141-145.

Arnaud JF (2003) Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. Landscape Ecology, 18, 333-346.

Aronson RB, Precht WF (2006) Conservation, precaution and Caribbean reefs. Coral Reefs, 25, 441-450.

Arosa ML, Ramos JA, Quintanilla LG, Brown D (2010) First report of fern (*Culcita macrocarpa*) spore consumption by a small mammal (*Apodemus sylvaticus*). Mammalian Biology, 75, 115-121.

Askew NP, Searle JB, Moore NP (2007) Agri-environment schemes and foraging of barn owls *Tyto alba*. Agriculture, Ecosystems and Environment, 118, 109-114.

Baker PJ, Ansell RJ, Dodds PAA, Webber CE, Harris S (2003) Factors affecting the distribution of small mammals in an urban area. Mammal Review, 33, 95-100.

Baker KH, Hoelzel AR (2013) Evolution of population genetic structure of the British roe deer by natural and anthropogenic processes (*Capreolus capreolus*). Ecology and Evolution, 3, 89-102.

Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner HH (2009) Identifying future research needs in landscape genetics: where to from here? Landscape Ecology, 24, 455-463.

Balkenhol N, Landguth EL (2011) Simulation modelling in landscape genetics: on the need to go further. Molecular Ecology, 20, 667-670.

Balkenhol N, Holbrook JD, Onorato D, Zager P, White C, Waits LP (2014) A multi-method approach for analyzing hierarchical genetic structures: a case study with cougars *Puma concolor*. Ecography, 37, 552-563.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11, 155-165.

Baraibar B, Westerman PR, Carrión E, Recasens J (2009) Effects of tillage and irrigation in cereal fields on weed seed removal by seed predators. Journal of Applied Ecology, 46, 380-387.

Barr KR, Lindsay D, Athrey G, Lance RF, Hayden TJ, Tweddale SA, Leberg PL (2008) Population structure in an endangered songbird: maintenance of genetic differentiation despite high vagility and significant population recovery. Molecular Ecology, 17, 3628-3639.

Bartmann S, Gerlach G (2001) Multiple paternity and similar variance in reproductive success of male and female wood mice (*Apodemus sylvaticus*) housed in an enclosure. Ethology 107, 889-899.

Bates FS, Harris S (2009) Does hedgerow management on organic farms benefit small mammal populations? Agriculture, Ecosystems and Environment, 129, 124-130.

Batzli GO, Pitelka FA (1971) Condition and diet of cycling populations of the California vole, *Microtus californicus*. Journal of Mammalogy, 52, 141-163.

Begg GS, Wishart J, Young MW, Squire GR, Iannetta PPM (2012) Genetic structure among arable populations of *Capsella bursa-pastoris* is linked to functional traits and in-field conditions. Ecography, 35, 446-457.

Bellamy PE, Shore EF, Ardeshir D, Treweek JR, Sparks TH (2000) Road verges as habitat for small mammals in Britain. Mammal Review, 30, 131-139.

Berckmoes V, Scheirs J, Jordaens K, Blust R, Backeljau T, Verhagen R (2005) Effects of environmental pollution on microsatellite DNA diversity in wood mouse (*Apodemus sylvaticus*) populations. Environmental Toxicology and Chemistry, 24, 2898-2907.

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

Berthier K, Charbonnel N, Galan M, Chaval Y, Cosson JF (2006) Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. Molecular Ecology, 15, 2665-2676.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology and Evolution, 18, 503-511.

Bolliger J, Lander T, Balkenhol N (2014) Landscape genetics since 2003: status, challenges and future directions. Landscape Ecology, 29, 361-366.

Boone GC, Tinklin R (1988) Importance of hedgerow structure in determining the occurrence and density of small mammals. Aspects of Applied Biology, 16, 73–78.

Booth W, Montgomery WI, Prodöhl PA (2007) Polyandry by wood mice in natural populations. Journal of Zoology, 273, 176-182.

Booth W, Montgomery WI, Prodöhl PA (2009) Spatial genetic structuring in a vagile species, the European wood mouse. Journal of Zoology, 279, 219-228.

Bouzat JL (2001) The population genetic structure of the Greater Rhea (*Rhea americana*) in an agricultural landscape. Biological Conservation, 99, 277-284.

Bouzat JL, Johnson K (2004) Genetic structure among closely spaced leks on a peripheral population of lesser prairie-chickens. Molecular Ecology, 13, 499-505.

Brady MJ, McAlpine CA, Possingham HP, Miller CJ, Baxter GS (2011) Matrix is important for mammals in landscapes with small amounts of native forest habitat. Landscape Ecology, 26, 617-628.

Brook BW, Sodhi NS, Bradshaw CJA (2008) Synergies among extinction drivers under global change. Trends in Ecology and Evolution, 23, 453-460.

Brown RW (1999) Margin/field interfaces and small mammals. Aspects of Applied Biology, 54, 203-210.

Brown PR, Huth NI, Banks PB, Singleton GR (2007) Relationship between abundance of rodents and damage to agricultural crops. Agriculture Ecosystems and Environment, 120, 405-415.

Butet A, Paillat G, Delettre Y (2006) Seasonal changes in small mammal assemblages from field boundaries in an agricultural landscape of western France. Agriculture, Ecosystems and Environment, 113, 364-369.

Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. Molecular Ecology, 23, 843-856.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Evolution, 32, 550-570.

Cegelski CC, Waits LP, Anderson NJ (2003) Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. Molecular Ecology, 12, 2907-2918.

Chambers JL, Garant D (2010) Determinants of population genetic structure in Eastern Chipmunks (*Tamias striatus*): the role of landscape barriers and sex-biased dispersal. Journal of Heredity, 101, 413-422.

Chantrey J, Meyer H, Baxby D, Begon M, Bown KJ, Hazel SM, Jones T, Montgomery WI, Bennett M (1999) Cowpox: reservoir hosts and geographic range. Epidemiology and Infection, 122, 455-460.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. Molecular Ecology, 18, 2746-2765.

Chappell MJ, LaValle LA (2011) Food security and biodiversity: can we have both? An agroecological analysis. Agriculture and Human Values, 28, 3-26.

Chapuisat M, Goudet J, Keller L (1997) Microsatellites reveal high population viscosity and limited dispersal in the ant *Formica paralugubris*. Evolution, 51, 475-482.

Chia W, Savakis C, Karp R, Pelham H, Ashburner M (1985) Mutation of the *Adh* gene of *Drosophila melanogaster* containing an internal tandem duplication. Journal of Molecular Biology, 186, 679-688.

Chiappero MB, Panzetta-Dutari GM, Gómez D, Castillo E, Polop JJ, Gardenai CN (2011) Contrasting genetic structure of urban and rural populations of the wild rodent C*alomys musculinus* (Cricetidae, Sigmodontinae). Mammalian Biology, 76, 41-50.

Churchfield S, Hollier J, Brown VK (1991) The effects of small mammal predators on grassland invertebrates, investigated by field exclosure experiment. Oikos, 60, 283-290.

Churchfield S, Hollier J, Brown VK (1997) Community structure and habitat use of small mammals in grasslands of different successional age. Journal of Zoology, 242, 519-530.

Clark JE, Hellgren EC, Parsons JL, Jorgensen EE, Engle DM, Leslie Jr DM (2005) Nitrogen outputs from fecal and urine deposition of small mammals: implications for nitrogen cycling. Oecologia, 144, 447-455.

Convention on Biological Diversity (2014) Accessed online (06.06.2014): http://www.cbd.int/

Coulon A, Cosson JF, Angibault JM, Cargnelutti B, Galan M, Morellet N, Petit E, Aulagnier S, Hewison AJM (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. Molecular Ecology, 13, 2841-2850.

Cushman SA, Landguth EL (2010a) Spurious correlations and inference in landscape genetics. Molecular Ecology, 19, 3592-3602.

Cushman SA, Landguth EL (2010b) Scale dependent inference in landscape genetics. Landscape Ecology, 25, 967-979.

Cushman SA, Wasserman TN, Landguth EL, Shirk AJ (2013) Re-evaluating causal modelling with Mantel tests in landscape genetics. Diversity, 5, 51-72.

Dalgaard T, Hutchings NJ, Porter JR (2003) Agroecology, scaling and interdisciplinarity. Agriculture, Ecosystems and Environment, 100, 39-51.

De Barro PJ (2005) Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region revealed using microsatellite markers. Molecular Ecology, 14, 3695-3718.

DEFRA (2013) Guidance on complying with the rules for Nitrate Vulnerable Zones in England for 2013 to 2016, Accessed online (06.06.2014): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/261371/pb14050-nvz-guidance.pdf

DEFRA (2014) Common Agricultural Policy (CAP) reform. Accessed online (20.06.14): http://rpa.defra.gov.uk/rpa/index.nsf/15f3e119d8abcb5480256ef20049b53a/7c4f7cd0d3afc7e380257c7b005ac070/$FILE/43169830.pdf/CAP%20reform%20-%20What%20you%20need%20to%20know%20now.pdf

Do C, Waples RS, Peel D, Macbeth GM, Tillet BJ, Ovenden JR (2014) NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources, 14, 209-214.

Dreier S, Redhead JW, Warren IA, Bourke AFG, Heard MS, Jordan WC, Sumner S, Wang J, Carvell C (2014) Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. Molecular Ecology, 23, 3384-3395.

Dewoody J, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. Molecular Ecology Notes, 6, 951-957.

Dickman CR, Doncaster CP (1987) The ecology of small mammals in urban habitats. I. Populations in a patchy environment. Journal of Animal Ecology, 56, 629-640.

Dickman CR, Doncaster CP (1989) The ecology of small mammals in urban habitats. II. Demography and dispersal. Journal of Animal Ecology, 58, 119-127.

Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic Salmon: the interaction of gene flow and local adaptation. Molecular Ecology, 17, 2382-2396.

Dobrowolski MP, Tommerup IC, Blakeman HD, O'Brien PA (2002) Non-Mendelian inheritance revealed in a genetic analysis of sexual progeny of *Phytophthora cinnamomi* with microsatellite markers. Fungal Genetics and Biology, 35, 197–212.

Donald PF, Evans AD (2006) Habitat connectivity and matrix restoration: the wider implications of agri-environment schemes. Journal of Applied Ecology, 43, 209-218.

Earl DA, vonHoldt BM (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conservation Genetics Resources, 4, 359-361.

EDINA Digimap Ordnance Survey Service, accessed online, (2013-2014): http://digimap.edina.ac.uk/digimap/home

Ehrich D, Yoccoz NG, Ims RA (2009) Multi-annual density fluctuations and habitat size enhance genetic variability in two northern voles. Oikos, 118, 1441-1452.

Elkinton JS, Liebhold AM, Muzika RM (2004) Effects of alternative prey on predation by small mammals on gypsy moth pupae. Population Ecology, 46, 171-178.

Emaresi G, Pellet J, Dubey S, Hirzel AH, Fumagalli L (2011) Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. Conservation Genetics, 12, 41-50.

Essbauer S, Schmidt J, Conraths FJ, Friedrich R, Koch K, Hautmann W, Pfeffer M, Wölfel R, Finke J, Obler GD, Ulrich R (2006) A new Puumala hantavirus subtype in rodents associated with an outbreak of *Nephropathia epidemica* in South-East Germany in 2004. Epidemiology and Infection, 134, 1333-1344.

Ersts PJ (2014) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation, accessed online (05.02.14): http://biodiversityinformatics.amnh.org/open_source/gdmg

Etherington TR (2011) Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. Methods in Ecology and Evolution, 2, 52-55.

European Commission (2013) The common agricultural policy (CAP) and agriculture in Europe – Frequently asked questions. http://europa.eu/rapid/press-release_MEMO-13-631_en.htm Accessed 10 December 2013

European Commission (2014a) Nagoya Protocol, accessed online (06.06.14): http://ec.europa.eu/environment/biodiversity/international/abs/index_en.htm

European Commission (2014b) CAP Reforms, accessed online (06.06.14): http://europa.eu/rapid/press-release_MEMO-13-621_en.htm

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

Fischer C, Thies C, Tscharntke T (2011) Small mammals in agricultural landscapes: Opposing responses to farming practices and landscape complexity. Biological Conservation, 144, 1130-1136.

Fitzgibbon CD (1997) Small Mammals in Farm Woodlands: The Effects of Habitat, Isolation and Surrounding Land-Use Patterns. Journal of Applied Ecology, 34, 530-539.

Fletcher H, Hickey I (2012) BIOS Instant Notes in Genetics, 4th Edition. London and New York, Tayler and Francis.

Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK, Helkowski JH, Holloway T, Howard EA, Kucharik CJ, Monfreda C, Patz JA, Prentice IC, Ramankutty N, Snyder PK (2005) Global consequences of land use. Science, 309, 570-574.

Ford RG (1983) Home range in a patchy environment: optimal foraging predictions. American Zoologist, 23, 315-326.

Frankham R (2005) Genetics and extinction. Biological Conservation, 126, 131-140.

Frantz AC, Pope LC, Carpenter PJ, Roper TJ, Wilson GJ, Delahay RJ, Burke T (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. Molecular Ecology, 12, 1649-1661.

Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? Journal of Applied Ecology, 46, 493-505.

Fuentes-Montemayor E, Goulson D, Park KJ (2011) Pipistrelle bats and their prey do not benefit from four widely applied agri-environment management prescriptions. Biological Conservation, 144, 2233-2246.

Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. Molecular Ecology, 6, 861-868.

Galpern P, Manseau M, Hettinga P, Smith K, Wilson P (2012) allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. Molecular Ecology Resources, 12, 771-778.

Galpern P, Manseau M, Hettinga P, Smith L, Wilson P (2014) The allelematch package for R: Supplementary documentation with tutorials. Accessed online (05.01.14): http://cran.r-project.org/web/packages/allelematch/vignettes/allelematchSuppDoc.pdf

Gardner-Santana LC, Norris DE, Fornadel CM, Hinson ER, Klein SL, Glass GE (2009) Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (*Rattus norvegicus*). Molecular Ecology, 18, 2766-2778.

Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008) Spatial genetic structure of a small rodent in a heterogeneous landscape. Molecular Ecology, 17, 4619-4629.

Gentili R, Abeli T, Rossi G, Li M, Varotto C, Sgorbati S (2010) Population structure and genetic diversity of the threatened quillwort *Isoëtes malinverniana* and implication for conservation. Aquatic Botany, 93, 147-152.

Gerlach G, Bartmann S (2002) Reproductive skew, costs, and benefits of cooperative breeding in female wood mice (*Apodemus sylvaticus*). Behavioural Ecology, 13, 408-418.

Giles BE, Lundqvist E, Goudet J (1998) Restricted gene flow and subpopulation differentiation in *Silene dioica*. Heredity, 80, 715-723.

Gockel J, Harr B, Schlötterer C, Arnold W, Gerlach G, Tautz D (1997) Isolation and characterization of microsatellite loci from *Apodemus flavicollis* (Rodentia: Muridae) and *Clethrionomys glareolus* (Rodentia: Cricetidae). Molecular Ecology, 6, 597-599.

Goldberg CS, Waits LP (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. Molecular Ecology, 19, 3650-3663.

Goldstein DB, Schlotterer C (1999) Microsatellites: Evolution and Applications. Oxford, Oxford University Press.

Goossens B, Waits LP, Taberlet P (1998) Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. Molecular Ecology, 7, 1237-1241.

Gortat T, Rutkowski R, Gryczynska-Siemiatkowska A, Kozakiewicz A, Kozakiewicz M (2013) Genetic structure in urban and rural populations of *Apodemus agrarius* in Poland. Mammalian Biology, 78, 171-177.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity, 86, 485-486.

Green R (1979) The ecology of wood mice (*Apodemus sylvaticus*) on arable farmland. Journal of Zoology, 188, 357-377.

Greene RA, Reynard C (1932) The influence of two burrowing rodents, *Dipodomys spectabilis spectabilis* (kangaroo rat) and *Neotoma albigula albigula* (Pack Rat), on desert soils in Arizona. II Physical effects. Ecology, 13, 359-363.

Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. Molecular Ecology Notes, 5, 712-715.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. Methods in Ecology and Evolution, 4, 336-344.

Gurnell J (1978) Seasonal changes in numbers and male behavioural interaction in a population of wood mice, *Apodemus sylvaticus*. Journal of Animal Ecology, 47, 741-755.

Halle S (1988) Avian predation upon a mixed community of common voles (*Microtus arvalis*) and wood mice (*Apodemus sylvaticus*). Oecologia, 75, 451-455.

Hansson L (1987) Dispersal routes of small mammal at an abandoned field in central Sweden. Ecography, 10, 154-159.

Harper JL (1974) Agricultural ecosystems. Agro-Ecosystems, 1, 1-6.

Harr B, Musolf K, Gerlach G (2000) Characterization and isolation of DNA microsatellite primers in wood mice (*Apodemus sylvaticus*, Rodentia). Molecular Ecology, 9, 1664-1665.

Harris SA, Yalden DW (2008) Mammals of the British Isles: Handbook, 4th revised edition. The Mammal Society, Southampton.

Heath DD, Busch C, Kelly J, Atagi DY (2002) Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). Molecular Ecology, 11, 197-214.

Hedges SB (2002) The origin and evolution of model organisms. Nature Reviews Genetics, 3, 838-849.

Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. Biological Conservation, 142, 14-32.

Heroldová M, Bryja J, Zejda J, Tkadlec E (2007) Structure and diversity of small mammal communities in agriculture landscape. Agriculture, Ecosystems and Environment, 120, 206-210.

Heroldová M, Tkadlec E (2011) Harvesting behaviour of three central European rodents: Identifying the rodent pest in cereals. Crop Protection, 30, 82-84.

Hitchings SP, Beebee TJC (1998) Loss of genetic diversity and fitness in common toad (*Bufo bufo*) populations isolated by inimical habitat. Journal of Evolutionary Biology, 11, 269-283.

Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology 14(2): 599-612

Hole DG, Perkins AJ, Wilson JD, Alexander IH, Grice F, Evans AD (2005) Does organic farming benefit biodiversity? Biological Conservation, 122, 113-130.

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

Howe JG, Grant WE, Folse LJ (1982) Effects of Grazing by *Sigmodon hispidus* on the Regrowth of Annual Rye-Grass (*Lolium perenne*). Journal of Mammalogy, 63, 176-179

Huitu O, Norrdahl K, Korpimäki E (2004) Competition, predation and interspecific synchrony in cyclic small mammal communities. Ecography, 27, 197-206.

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

Ingvarsson PK, Olsson K (1997) Hierarchical genetic structure and effective population sizes in *Phalacrus substriatus.* Heredity, 79, 153-161.

Isbell F, Calcagno C, Hector A, Connolly J, Harpole WS, Reich PB, Schere-Lorenzen M, Schmid B, Tilman D, van Ruijven J, Weigelt A, Wilsey BJ, Zavaleta ES, Loreau M (2011) High plant diversity is needed to maintain ecosystem services. Nature, 477, 199-202.

Jacob J, Brown JS (2000) Microhabitat use, giving-up densities and temporal activity as short- and long-term anti-predator behaviours in common voles. Oikos, 91, 131-138.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23, 1801-1806.

James A, Gaston KJ, Balmford A (2001) Can we afford to conserve biodiversity? BioScience, 51, 43-52.

Janova E, Heroldova M, Konecny A, Bryja J (2011) Traditional and diversified crops in South Moravia (Czech Republic): Habitat preferences of common vole and mice species. Mammalian Biology, 76, 570-576.

Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N (2011) Inferring landscape effects on dispersal from genetic distances: how far can we go? Molecular Ecology, 20, 692-705.

Johansson M, Primmer CR, Sahlsten J, Merilä J (2005) The influence of landscape structure on occurrence, abundance and genetic diversity of the common frog, *Rana temporaria*. Global Change Biology, 11, 1664-1679.

Justus J, Colyvan M, Regan H, Maguire L (2009) Buying into conservation: intrinsic versus instrumental value. Trends in Ecology and Evolution, 24, 187-191.

Kalinowski S (2005) HP-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes, 5, 187-189.

Keller D, Van Strien MJ, Holderegger R (2012) Do landscape barriers affect functional connectivity of populations of an endangered damselfly? Freshwater Biology, 57, 1373-1384.

Keyghobadi N, Roland J, Matter SF, Strobeck C (2005) Among and within patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proceedings of the Royal Society B, 272, 553-560.

Kijlstra A, Meerburg B, Cornelissen J, De Craeye S, Vereijken P, Jongert E (2008) The role of rodents and shrews in the transmission of *Toxoplasma gondii* to pigs. Veterinary Parasitology, 156, 183-190.

Kikkawa J (1964) Movement, activity and distribution of the small rodents *Clethrionomys glareolus* and *Apodemus sylvaticus* in woodland. Journal of Animal Ecology, 33, 259-99.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics, 49, 561.

Kiviniemi K, Telenius A (1998) Experiments on adhesive dispersal by wood mouse: seed shadows and dispersal distances of 13 plant species from cultivated areas in southern Sweden. Ecography, 21, 108-116.

Kleijn D, Sutherland WJ (2003) How effective are European agri-environment schemes in conserving and promoting biodiversity? Applied Ecology, 40, 947-969.

Kleijn D, Baquero RA, Clough Y, Díaz M, De Esteban J, Fernández F, Gabriel D, Herzog F, Holzschuh A, Jöhl R, Knop E, Kruess A, Marshall EJP, Steffan-Dewenter I, Tscharntke T, Verhulst K, West TM, Yela JL (2006) Mixed biodiversity benefits of agri-environment schemes in five European countries. Ecology Letters, 9, 243-254.

Korpimäki E, Brown PR, Jacob J, Pech RP (2004) The Puzzles of Population Cycles and Outbreaks of Small Mammals Solved? BioScience, 54, 1071-1079.

Lada H, Thomson JR, MacNally R, Taylor AC (2008) Impacts of massive landscape change on a carnivorous marsupial in south-eastern Australia: inferences from landscape genetics analysis. Journal of Applied Ecology, 45, 1732-1741.

Lambin Z, Bretagnolle V, Yoccoz NG (2006) Vole population cycles in northern and southern Europe: Is there a need for different explanations for single pattern? Journal of Animal Ecology, 75, 340-349.

Landguth EL, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. Molecular Ecology, 19, 4179-4191.

Landguth EL, Fedy BC, Oyler-McCance SJ, Garey AL, Emel SL, Mumma M, Wagner HH, Fortin MJ, Cushman SA (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Molecular Ecology Resources, 12, 276-284.

Lens L, Dongen SV, Glabusera P, Schenck T, Matthysen E, Van De Casteele T (2000) Developmental instability and inbreeding in natural bird populations exposed to different levels of habitat disturbance. Journal of Evolutionary Biology, 13, 889-896.

Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America, 15, 237-240.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions, and mutational mechanisms: a review. Molecular Ecology, 11, 2453-2465.

Life Technologies, Gene Mapper, available at http://www.lifetechnologies.com/order/catalog/product/4475073?CID=search-4475073

Lindsay D, Barr KR, Lance RF, Tweddale SA, Hayden TJ, Leberg PL (2008) Habitat fragmentation and genetic diversity of an endangered, migratory songbird, the golden-cheeked warbler (*Dendroica chrysoparia*). Molecular Ecology, 17, 2122-2133.

Maas J, Verheij RA, Groenewegen PP, de Vries S, Spreeuwenberg P (2006) Green space, urbanity, and health: how strong is the relation? Journal of Epidemiology and Community Health, 60, 587-592.

Macdonald DW, Tew TE, Todd IA, Garner JP, Johnson PJ (2000) Arable habitat use by wood mice (*Apodemus sylvaticus*). 3. A farm-scale experiment on the effects of crop rotation. Journal of Zoology, 250, 313-320.

Macdonald DW, Tattersall FT (2001) Britain's mammals – the challenge for conservation. The Wildlife Conservation Research Unit, University of Oxford.

Macdonald DW, Tattersall FH, Service KM, Firbank LG, Feber RE (2007) Mammals, agri-environment schemes and set-aside – what are the putative benefits? Mammal Review, 37, 259-277.

Makova KD, Patton JC, Krysanov EY, Chesser RK, Baker RJ (1998) Microsatellite markers in wood mouse and striped field mouse (genus *Apodemus*). Molecular Ecology, 7, 247-255.

Makova KD, Nekrutenko A, Baker RJ (2000) Evolution of microsatellite alleles in four species of mice (Genus *Apodemus*). Journal of Molecular Evolution, 51, 166-172.

Mallorie HC, Flowerdew JR (1994) Woodland small mammal population ecology in Britain: a preliminary review of the Mammal Society survey of wood mice *Apodemus sylvaticus* and bank voles *Clethrionomys glareolus*, 1982–87. Mammal Review, 24, 1-15.

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MC, Scribner KT, Bonin A, Fortin MJ (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Molecular Ecology, 19, 3760-3772.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology and Evolution, 18, 189-197.

Marchi C, Andersen LW, Damgaard C, Olsen K, Jensen TS, Loeschcke V (2013) Gene flow and population structure of a common agricultural wild species (*Microtus agrestis*) under different land management regimes. Heredity, 111, 486-494.

Marino PC, Westerman PR, Pinkert C, van der Werf W (2005) Influence of seed density and aggregation on post-dispersal weed seed predation in cereal fields. Agriculture, Ecosystems and Environment, 106, 17-25.

Marrotte RR, Gonzalez A, Millien V (2014) Landscape resistance and habitat combine to provide an optimal model of genetic structure and connectivity at the range margin of a small mammal. Molecular Ecology, 23, 3983-3998.

Mcdonald RI, Kareiva P, Forman RTT (2008) The implications of current and future urbanisation for global protected areas and biodiversity conservation. Biological Conservation, 141, 1695-1703.

McGarigal K, Cushman SA, Neel MC, Ene E (2002) Fragstats: Spatial pattern analysis program for categorical maps. University of Massachusetts, Amherst, accessed online (01.04.14): www.umas.edu/landeco/research/fragstats/fragstats.html

Meirmans PG (2012) The trouble with isolation by distance. Molecular Ecology, 21, 2839-2846.

Michel N, Burel F, Butet A (2006) How does landscape use influence small mammal diversity, abundance and biomass in hedgerow networks of farming landscapes? Acta Oecologica, 30, 11-20.

Millán de la Peña N, Butet A, Delettre Y, Paillat G, Morant P, Le Du L, Burel F (2003) Response of the small mammal community to changes in western French agricultural landscapes. Landscape Ecology, 18, 265-278.

Montgomery WI (1989) Population regulation in the wood mouse *Apodemus sylvaticus*. II. Density dependence in spatial distribution and reproduction. Journal of Animal Ecology, 58, 477-494.

Montgomery WI, Dowie M (1993) The distribution and population regulation of the wood mouse *Apodemus sylvaticus* on field boundaries of pastoral farmland. Journal of Applied Ecology, 30, 783-791.

Montgomery WI, Lundy MG, Reid N (2012) 'Invasional meltdown': evidence for unexpected consequences and cumulative impacts of multispecies invasions. Biological Invasions, 14, 1111-1125.

Morton D, Rowland C, Wood C, Meek L, Marston C, Smith G, Simpson IC (2011) Final report for LCM2007 – the new UK land cover map. CS Technical Report No 11/07 NERC/Centre for Ecology & Hydrology 108pp. (CEH project number: C03259).

Munshi-South J, Kharchenko K (2010) Rapid, pervasive genetic differentiation of urban white-footed mouse (*Peromyscus leucopus*) populations in New York City. Molecular Ecology, 19, 4242-4254.

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. Ecology, 91, 252-261.

Natural England (2013) Land Management Developments, accessed online (06.06.2014): http://www.naturalengland.org.uk/ourwork/farming/funding/developments.aspx

Nichols HJ, Jordan NR, Jamie GA, Cant MA, Hoffman J (2012) Fine-scale spatiotemporal patterns of genetic variation reflect budding dispersal coupled with strong natal philopatry in a cooperatively breeding mammal. Molecular Ecology, 21, 5348-5362.

Nussey DH, Coltman DW, Coulson T, Kruuk LEB, Donald A, Morris SJ, Clutton-Brock TH, Pemberton J (2005) Rapidly declining fine-scale spatial genetic structure in female red deer. Molecular Ecology, 14, 3395-3405.

Obrtel R, Zejda J, Holisova V (1978) Impact of small rodent predation on an overcrowded population of *Diprion pini* during winter. Folia Zoologica, 27, 97-110.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2012) vegan: community ecology package. R package version. Accessed online (02.01.14): http://CRAN.R-project.org/package=vegan

Oliveras de Ita A, Oyama K, Smith TB, Wayne RK, Mila B (2012) Genetic evidence for recent range fragmentation and severely restricted dispersal in the critically endangered Sierra Madre Sparrow, *Xenospiza baileyi*. Conservation Genetics, 13, 283-291.

Olsen MT, Andersen LW, Dietz R, Teilmann J, Härkönen T, Siegismund HR (2014) Integrating genetic data and population viability analyses for the identification of harbour seal (*Phoca vitulina*) populations and management units. Molecular Ecology, 23, 815-831.

Ouin A, Paillat G, Butet A, Burel F (2000) Spatial dynamics of wood mouse (*Apodemus sylvaticus*) in an agricultural landscape under intensive use in the Mont Saint Michel Bay (France). Agriculture, Ecosystems and Environment, 78, 159-165.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics, 147, 1943-1957.

Panzacchi M, Linnell JDC, Melis C, Odden M, Odden J, Gorini L, Andersen R (2010) Effect of land-use on small mammal abundance and diversity in a forest–farmland mosaic landscape in south-eastern Norway. Forest Ecology and Management, 259, 1536-1545.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6, 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics, 28: 2537-2539.

Perry JN, Winder L, Holland JM (1999) Red-blue plots for detecting clusters in count data. Ecology Letters, 2, 106-113.

Piertney SB, Lambin X, Maccoll ADC, Lock K, Bacon PJ, Dallas JF, Leckie F, Mougeot F, Racey P, Redpath S, Moss R (2008) Temporal changes in kin structure through a population cycle in a territorial bird, the red grouse *Lagopus lagopus scoticus.* Molecular Ecology, 17, 2544-2551.

Pilot M, Dąbrowski MJ, Jancewicz E, Schtickzelle N, Gliwicz J (2010) Temporally stable genetic variability and dynamic kinship structure in a fluctuating population of the root vole *Microtus oeconomus*. Molecular Ecology, 19, 2800-2812.

Pollard E, Relton J (1970) Hedges. V. A study of small mammals in hedges and cultivated fields. Journal of Applied Ecology, 7, 549-557.

Pope LC, Domingo-Roura X, Erven K, Burke T (2006) Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. Molecular Ecology, 15, 371-386.

Pretty J, Sutherland WJ, Ashby J, Auburn J, Baulcombe D, Bell M, Bentley J, Bickersteth S, Brown K, Burke J, Campbell H, Chen K, Crowley E, Crute I, Dobbelaere D, Edward-Jones G, Funes-Monzote F, Godfrat HCJ, Griffon M, Gypmantisiri P, Haddad L, Halavatau S, Herren H, Holderness M, Izac AM, Jones M, Koohafkan P, Rattan L, Lang T, McNeely K, Mueller A, Nisbett N, Noble A, Pingali P, Pinto Y, Rabbinge R, Ravindranath NH, Rola A, Roling N, Sage C, Settle W, Sha JM, Shiming L, Simons T, Smith P, Strzepeck K, Swaine H, Terry E, Tomich TP, Toulmin C,

Trigo E, Twomlow S, Vis JK, Wilson J, Pilgrim S (2010) The top 100 questions of importance to the future of global agriculture. International Journal of Agricultural Sustainability, 8, 219-236.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Purrenhage JL, Niewiarowski PH, Moore FBG (2009) Population structure of spotted salamanders (*Ambystoma maculatum*) in a fragmented landscape. Molecular Ecology, 18, 235-247.

R Development Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Rayfield B, Fortin MJ, Fall A (2010) The sensitivity of least-cost habitat graphs to relative cost surface values. Landscape Ecology, 25, 519-532.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.

Recce KS, Ribeiro WL, Gaffney PM, Carnegie RB, Allen SK (2004) Microsatellite marker development and analysis in the Eastern Oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. Journal of Heredity, 95, 346-352.

Reid N, McDonald RA, Montgomery WI (2007) Mammals and agri-environment schemes: hare haven or pest paradise? Journal of Applied Ecology, 44, 1200-1208.

Richardson JH, Shore RF, Treweek JR, Larkin SBC (1997) Are major roads a barrier to small mammals? Journal of Zoology, 243, 840-846.

Rikalainen K, Aspi J, Galarza JA, Koskela E, Mappes T (2012) Maintenance of genetic diversity in cyclic populations – a longitudinal analysis in *Myodes glareolus*. Ecology and Evolution, 2, 1491-1502.

Robinson RA, Sutherland WJ (2002) Post-war changes in arable farming and biodiversity in Great Britain. Journal of Applied Ecology, 39, 157-176.

Rocha RG, Ferreira E, Leite YLR, Fonseca C, Costa LP (2011) Small mammals in the diet of barn owls, *Tyto alba* (Aves: Strigiformes) along the mid-Araguaia river in central Brazil. Zoología, 28, 709-716.

Rodríguez-Ramilo ST, Wang J (2012) The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. Molecular Ecology Resources, 12, 873-884.

Rogers LM, Gorman ML (1995) The diet of the wood mouse *Apodemus sylvaticus* on set-aside land. Journal of Zoology, 235, 77-83.

Rosenzweig ML (1987) Habitat selection as a source of biological diversity. Evolutionary Ecology, 1, 315-330.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources, 8, 103-106.

Rudd H, Vala J, Schaefer V (2002) Importance of backyard habitat in a comprehensive biodiversity conservation strategy: a connectivity analysis of urban green spaces. Restoration Ecology, 10, 368-375.

Ryšavá-Nováková M, Koubek P (2009) Feeding habits of two sympatric mustelid species, European polecat *Mustela putorius* and stone marten *Martes foina*, in the Czech Republic. Folia Zoologica, 58, 66-75.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation. Nature, 392, 491-494.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4, 406-425.

Salamolard M, Butet A, Leroux A, Bretagnolle V (2000) Responses of an avian predator to variations in prey density at a temperate latitude. Ecology, 81, 2428-2441.

Sander AC, Purtauf T, Wolters V, Dauber J (2006) Landscape genetics of the widespread ground-beetle *Carabus auratus* in an agricultural region. Basic and Applied Ecology, 7, 555-564.

Schmidt T, Arens P, Smulders MJM, Billeter R, Liira J, Augenstein I, Durka W (2009) Effects of landscape structure on genetic diversity of *Geum urbanum L.* populations in agricultural landscapes. Flora, 204, 549-559.

Schweizer M, Excoffier L, Heckel G (2007) Fine-scale genetic structure and dispersal in the common vole (*Microtus arvalis*). Molecular Ecology, 16, 2463-2473.

Scottish Agricultural Science Agency, Hay RKM, Russell G, Edwards TW (2000) Crop Production in the East of Scotland, accessed online (06.06.2014): http://www.sasa.gov.uk/sites/default/files/Crop%20Production%20in%20East%20of%20Scotland.pdf

Scottish Natural Heritage (2014) Licences for shrews and water voles, accessed online (06.06.2014): http://www.snh.gov.uk/protecting-scotlands-nature/species-licensing/mammal-licensing/shrew-vole/

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9, 615–629.

Seto KC, Fragkias M, Güneralp B, Reilly MK (2011) A meta-analysis of global urban land expansion. PLoS ONE, 6(8), e23777.

Shore RF, Meek W, Sparks TH, Pywell RF, Nowakowski M (2000) Small mammals in Countryside Stewardship field margin strips. Aspects of Applied Biology, 58, 417-420.

Shore RF, Meek WR, Sparks TH, Pywell RF, Nowakowski M (2005) Will Environmental Stewardship enhance small mammal abundance on intensively managed farmland? Mammal Review, 35, 277-284.

Slatkin M, Voelm L (1991) *FST* in hierarchical island model. Genetics, 127, 627–629.

Smirnov VS, Tokmakova SG (1971) Preliminary data on the influence of different numbers of voles upon the forest tundra vegetation. Annales Zoologici Fennici, 8, 154-156.

Smith HE, Macdonald DW, Feber R, Plesner Jensen S, Younes M (1993) A report to English Nature on the Wytham Farm Experiment.

Smith O, Wang J (2014) When can noninvasive samples provide sufficient information in conservation genetics studies? Molecular Ecology Resources, 14, 1011-1023.

Smith AL, Bull CM, Gardner MG, Driscoll DA (2014) Life history influences how fire affects genetic diversity in two lizard species. Molecular Ecology, 23, 2428-2441.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity, 82, 561-73.

Southern HN, Lowe VPW (1982) Predation by Tawny owls (*Strix aluco*) on bank voles (*Clethrionomys glareolus*) and wood mice (*Apodemus sylvaticus*). Journal of Zoology, 198, 83-102.

Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Molecular Ecology, 19, 3576-3591.

Stanton JM, McNicol CD, Steele V (1998) Non-manual lysis of second-stage *Meloidogyne* juveniles for identification of pure and mixed samples based on the polymerase chain reaction. Australasian Plant Pathology, 27, 112-115.

Stewart WA, Dallas JF, Piertney SB, Marshall F, Lambin X, Telfer S (1999) Metapopulation genetic structure in the water vole, *Arvicola terrestris*, in NE Scotland. Biological Journal of the Linnean Society, 68, 159-171.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. Heredity, 98, 129-142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? Molecular Ecology, 19, 3496-3514.

Swenson JJ, Franklin J (2000) The effects of future urban development on habitat fragmentation in the Santa Monica Mountains. Landscape Ecology, 15, 713-730.

Taberlet P, Waits LP, Luikart G (1999) Non-invasive genetic sampling: look before you leap. Trends in Ecology and Evolution, 14, 323-327.

Tattersall FH, Macdonald DW, Hart BJ, Manley WJ, Feber RE (2001) Habitat use by wood mice (*Apodemus sylvaticus*) in a changeable arable landscape. Journal of Zoology, 255, 487-494.

Tattersall FH, Macdonald DW, Hart BJ, Manley W (2004) Balanced dispersal or source–sink – do both models describe wood mice in farmed landscapes? Oikos, 106, 536-550.

Tew TW, Macdonald DW (1993) The effects of harvest on arable wood mice *Apodemus sylvaticus.* Biological Conservation, 65, 279-283.

Tew TE, Todd IA, Macdonald DW (2000) Arable habitat use by wood mice (*Apodemus sylvaticus*). 2. Microhabitat. Journal of Zoology, 250, 305-311.

The Geneland Development Group (2012) Population genetic and morphometric data analysis using R and the GENELAND program, accessed online (05.02.14): http://www2.imm.dtu.dk/~gigu/Geneland/Geneland-Doc.pdf

The Scottish Government (2014) The new common agricultural policy in Scotland, accessed online (20.06.14): http://www.scotland.gov.uk/Resource/0045/00453526.pdf

Tilman D, Fargione J, Wolfe B, D'Antonio C, Dobson A, Howarth R, Schindler D, Schlesinger WH, Simberloff D, Swackhamer D (2001) Forecasting agriculturally driven global environmental change. Science, 292, 281-284.

Todd IA, Tew TE, Macdonald DW (2000) Arable habitat use by wood mice (*Apodemus sylvaticus*). 1. Macrohabitat. Journal of Zoology, 250, 299-303.

Van De Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. Molecular Ecology, 10, 1539-1549.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 4, 535-538.

Vander Wall SB (2003) Effects of seed size of wind-dispersed pines (*Pinus*) on secondary seed dispersal and the caching behavior of rodents. Oikos, 100, 25-34.

Vargas V, Valensuela-Galván D, Alcalá RE (2012) Is genetic structure of the southern pygmy mouse *Baiomys musculus*(Cricetidae) related to human-induced spatial landscape heterogeneity in a tropical dry forest? Genetica, 140, 287-295.

Waits LP, Liukart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology, 10, 249-256.

Wandeler P, Funk SM, Largiadèr CR, Gloor S, Breitenmoser U (2003) The city-fox phenomenon: genetic consequences of a recent colonization of urban habitat. Molecular Ecology, 12, 647-656.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. Molecular Ecology, 19, 2605-2608.

Watts CHS (1968) The Foods Eaten by Wood Mice (*Apodemus sylvaticus*) and Bank Voles (*Clethrionomys glareolus*) in Wytham Woods, Berkshire. Journal of Animal Ecology, 37, 25-41.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358-1370.

Westerman PR, Hofman A, Vet LEM, van der Werf W (2003) Relative importance of vertebrates and invertebrates in epigeaic weed seed predation in organic cereal fields. Agriculture, Ecosystems and Environment, 95, 417-425.

White TA, Searle JB (2008) The colonization of Scottish islands by the common shrew, *Sorex araneus* (Eulipotyphla: Soricidae). Biological Journal of the Linnean Society, 94, 797-808.

Whittingham MJ (2007) Will agri-environment schemes deliver substantial biodiversity gain, and if not why not? Applied Ecology, 44, 1-5.

Wolton RJ (1985) The ranging and nesting behaviour of wood mice, *Apodemus sylvaticus* (Rodentia: Muridae), as revealed by radio-tracking. Journal of Zoology, 206, 203-222.

Woodmansee RG, Dodd JL, Bowman RA, Clark FE, Dickinson CE (1978) Nitrogen budget of a shortgrass prairie ecosystem. Oecologia, 34, 363-376.

Ylönen H, Altner HJ, Stubbe M (1991) Seasonal dynamics of small mammals in an isolated woodlot and its agricultural surroundings. Annales Zoologici Fennici, 28, 7-14.

Zapletal M, Obdržálková D, Pikula J, Zejda J, Pikula J Jr, Beklová M, Heroldová M (2001) Common vole *Microtus arvalis* (Pallas, 1778) in Czech Republic. Academic Press CERM.

Zhou Y, Newman C, Xu W, Buesching CD, Zalewski A, Kaneko Y, Macdonald DW, Xie Z (2011) Biogeographical variation in the diet of Holarctic martens (genus *Martes*, Mammalia: Carnivora: Mustelidae): adaptive foraging in generalists. Journal of Biogeography, 38, 137-147.