Population genetic structure of a recovering otter (*Lutra lutra*) population in the UK.

Geoffrey I Hobbs

Cardiff University

A dissertation submitted to Cardiff University in candidature for the degree of Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The aim of this study was to identify the genetic structure of the expanding UK otter population. To do this I undertook detailed analysis of methodologies from the emerging field of landscape genetics. I compared the Bayesian Clustering methodologies, culminating in recommendations on how to interpret the results of the available software. Further to this I devised a novel progressive partitioning method, incorporating GIS (geographical information systems) to allow the clustering results of the different software packages to be compared and combined, producing a more robust interpretation of clustering results. The effect of landscape features on otter movement was explored using GIS by mapping individuals on cost grids of landscape features and identifying the degree to which dispersal is influenced by the landscape, by correlating effective distance with genetic distance. Inspiration was taken from recent advances in landscape genetics and required the development of these techniques to achieve the aims of the project; as a result this thesis also contributes to the advancement of this field of research.

This study identified that there are four regional otter populations in the UK with little or no gene flow between them. The recovering otter populations in the strongholds of North England, Wales and Borders and Southwest England appear not to be contributing to expansion of the once fragmented, unviable population in Central England. This population has been subject to captive bred re-introductions by the Otter Trust. Despite the apparent success of the reintroductions, questions have arisen about the origin of the released individuals and their conservation implications. Further sub-structuring was identified in all of the regional populations and potential reasons explored. The Wales and Borders region was singled out for further analysis, to identify the influence of landscape features on the genetic structure. The highly urban areas of southeast Wales appeared to be acting as a barrier to dispersal between sub-regions. Correlations between genetic and effective distances (created from resistance-to-movement surfaces) suggest that upland habitat and slope contribute to the genetic sub-structuring; the Cambrian and Brecon Beacon mountain ranges act as permeable barriers, restricting the amount of gene flow and help to create the identified sub-regions.

Acknowledgements

I would like to thank all the people who have contributed to this thesis in one way or another. Support and guidance came from Mike, Liz and Fred mu PhD supervisors from whom I have learnt much. Thanks goes to Mike for giving me snippets of knowledge and wisdom to help me on my way and special thanks to Liz who has gone above and beyond, with lots of encouragement some constructive some just mean, and advising me impeccable. Thanks goes also to lab members old and new who were willing to share their knowledge, give advice and be great friends; particular thanks goes to Andy, Pati, Gabi, Fairus, Amy, Muya, Dan, Carlos, Renata, Loys, Wendy, Rhys, Milena, everyone called Jo, Jeff, Viola, the Mafaldas, Xiang Jiang, Mirelle, Dave S, Dave B and last but not least Timbo, and all the others I have missed out.

I would like to thank the Environment Agency, for their funding of the Cardiff University Otter Project as well as CCW and UK Wildlife Trusts for their collection of otter carcasses. Kevin Dawson and Gilles Guillot for advice on their softwares. Special thanks to the Cardiff University 'CONDOR' supervised by James Osbourne. I owe a lot to my parents and family for their support while they waited for me to get a proper job, and most importantly I would like to thank Jude for putting up with me and for her strength and support in the last couple of years during the PhD madness.

Table of Contents

List of Tables and Figures

Chapter 1

Introduction

1.1. Introduction

There is currently little doubt that a crisis is overtaking our planet with a significant global decline in biodiversity which has precipitated a major species extinction event (Myers & Knoll 2001). Whilst the earth has undergone major extinctions in the past that were driven by random catastrophes and major climate fluctuations, the current crisis is occurring largely due to human activities (Myers & Knoll 2001). Humans have dominated the earth's ecosystems, transforming much of the worlds' land surface and utilising over half of the worlds available surface freshwater (Vitousek *et al.* 1997). The growth of the human population and development have had many effects on biodiversity, from the conversion of natural habitats, exploitation of species, introductions of invasive species, war, pollution and urbanisation (Botkin *et al.* 2007).

In our modern and crowded world, medium and large carnivores are among the most challenging taxonomic groups to conserve (Mech 1995). Their large home ranges, greater food demands and predatory nature can all bring them into conflict with humans (Linnell *et al.* 2001). Europe once provided a range of natural habitats for large carnivores but they have been driven from many areas, experiencing years of persecution and loss of habitat, most now surviving on the edges of their former range in fragmented habitats.

In much of Europe the natural and semi natural habitat that is left is located in protected areas. These areas were initially set aside for hunting grounds by rulers of the countries and have developed into national parks, mostly on land owned by national and regional governments, however, a greater diversity of ownership and an expansion of protected areas has occurred since the implementation of Natura 2000. Further national responsibility remains due to international obligations (Bonn, Ramsar and World Heritage Conventions and the Convention on Biological Diversity) and European obligations under Natura 2000, the Convention on the Conservation of European Wildlife and Natural Habitats (Berne Convention).

The legal protection of remaining habitats has provided refuges for many threatened species, and after favourable legislation some carnivore populations have started to

increase (Boitani 2000; Linnell *et al.* 2001). Carnivores have started to recolonise areas partly as a result of reintroductions but also through natural expansion (Breitenmoser 1998) (e.g. wolf (*Canis lupis*) [Lucchini *et al.* 2002; Valière *et al.* 2003]; Eurasian Lynx (*Lynx lynx*) [Cop & Frkovic 1998]. Connectivity between these fragmented populations is very important as gene flow is considered necessary for the viability of small isolated populations (Mills & Allendorf 1996). The expansion of the species range and identification of habitat corridors is therefore essential to link fragmented genetically isolated populations and re-establish gene flow.

Many carnivore species are secretive and difficult to study, and as a result, field observations alone are not sufficient to adequately investigate populations. Therefore conservation biologists, wildlife managers and other scientists are turning to genetic analysis to aid their understanding of conservation strategies. Population genetic techniques provide tools to answer many questions in ecology, conservation and wildlife management. The use of genetic analysis by isolating DNA from hair, faeces or animal tissues has allowed researchers to study wild populations *in situ*, and has been used in studies investigating population structure, reintroductions, expansions and invasive colonization (Lucchini *et al.* 2002; Fabri *et al.* 2007; Zalewski *et al.* 2009). The ability to identify genetic patterns of ongoing recolonisation and population expansion has an important role in practical conservation biology.

Landscape genetics allows the identification of population structure and enables the correlation of genetic variation with landscape features, so facilitating the identification of barriers to and routes of dispersal (Manel *et al.* 2003; Holderegger & Wagner 2006). Investigating the genetic patterns of ongoing population expansion, recolonisation from refugia, and invasive events will have important roles for conservation biology and help construct predictive models for future spread (Lucchini *et al.* 2002; Fabri *et al.* 2007; Zalewski *et al.* 2009).

In the UK, large carnivores such as bears and wolves were persecuted to extinction several centuries ago. A number of medium-sized carnivores remain in much reduced numbers. The wild cat of Scotland (*Felis silvestris grampia*) is critically

endangered (Kitchener *et al.* 2005), and while pine martens (*Martes martes*) (Kyle *et al.* 2003) and polecats (*Mustela putorius*) (Birks 1997) are expanding from their main ranges in Scotland and Wales, respectively, their expansion will inevitably lead to conflict with people. The Eurasian otter *(Lutra lura)*, one of the UK's largest remaining carnivores is also making a comeback under legislative protection, and conservation aimed at protecting both the species and its habitat. The otter, with no natural predators and limited human persecution, provides an ideal opportunity to study natural recolonisation events in carnivores, and to test various hypotheses and techniques which may later be applied to other large carnivores.

As well as acting as bio-indicators of ecosystem health (Ruiz-Olmo *et al.* 1998; Basu *et al.* 2007), an increased presence of otters will allow people to become accustomed to large carnivores. This may help change public attitude (Schwartz *et al.* 2003), softening of opinions and a realization of the intrinsic value of wildlife, including large predators (White *et al.* 1997; Schwartz *et al.* 2003). Potentially this may help create an acceptance for other large carnivores when they return to areas where they have been absent in recent years (Valière *et al.* 2003; Kaczensky *et al.* 2004).

1.2. The Eurasian otter (*Lutra lutra*)

Otters are members of the Mustelidae, the most diverse and numerous family within Order Carnivora. There are thirteen species of otter found worldwide, grouped into the subfamily *Lutrinae* (Koepfli & Wayne 1998; Kruuk 2006), of which the Eurasian otter (*Lutra lutra*) has the broadest distribution, extending from the west coast of Ireland to Japan, and from Arctic Finland to North Africa and Indonesia (Chanin 1985). Otters are found from sea level up to 4120m in Tibet (Mason & Macdonald 1986) and up to 2000m in Spain (Ruiz-Olmo *et al.* 1998). Throughout this thesis, 'otter(s)' refers to the Eurasian otter except where otherwise specified.

Otters utilise diverse aquatic habitats, including lakes and bogs, rivers and streams, and coastal areas. Specialising on aquatic prey, they feed primarily on fish, but are an opportunistic predator supplementing their diet with amphibians, small mammals, birds and invertebrates depending on the availability of prey (Jedrzejewska 2001; Clavero *et al.* 2003; Bonesi & Macdonald 2004; Lanszki & Sallai 2006). Although feeding primarily in water, otters spend three quarters or more of their time on land

(Durbin 1998). They require suitable terrestrial breeding and resting sites, commonly referred to as holts. These may be tunnels under waterside trees, or more open 'nests' in dense vegetation such as reed beds. They can be found in large cities and can tolerate some disturbance (Durbin 1998; Bedford 2009).

Otters generally live in habitats which can be described as linear, with territories along water bodies (Ruiz-Olmo *et al.* 2001b; Bonesi & Macdonald 2004) that are 10-20km in length depending on the quality of the habitat and resources (40km in length in poor quality habitat, or less where prey is more readily available (Kruuk 2006; Bedford 2009). Male territories tend to overlap those of several females, but otters are primarily solitary (except where females are with cubs; Kruuk 2006). Direct contact is therefore rare, and it is thought that communication is mainly achieved by means of scent signals (Kruuk 1992). As well as resident animals, part of the population is transitory (Kruuk 2006). Otters have a low life expectancy for their size (3-4 years; Gorman *et al.* 1998, Simpson 1998), a low reproductive rate and high mortality, and are vulnerable to human disturbances especially road kill. There may therefore be a large turnover and frequent change in territorial boundaries (Kruuk 2006).

1.3. Population history

The Eurasian otter has declined significantly throughout its European range (Barbosa *et al.* 2003) and in the UK this occurred particularly during the late 1950's and early 1960's, throughout much of Wales, England and the Scottish borders (Coxon *et al.* 1999; Conroy & Chanin 2000; Mason & Macdonald 2004). By the mid 1970's the UK population was largely confined to strongholds in parts of Scotland, Northern Ireland, mid and west Wales and south west England (Jones & Jones 2004) with a small remnant population in East Anglia (Jessop & Cheyne 1992). Europe-wide declines led to the species being listed in the (IUCN 1990) red list of threatened species as either vulnerable or endangered throughout much of its current range (Ruiz-Olmo *et al.* 2001b), although its status has now been revised to 'near threatened' on the IUCN Red List of endangered species (2004, 2008).

1.4. Otter population history in the UK

1.4.1. Decline of the UK otter population

The otter was thought to be distributed throughout most of the UK in the 18th century (Jefferies 1989). During the latter half of the 18th century they were persecuted, hunted for sport, as a result there were local extinctions (Jefferies 1989). In the 19th century persecution became more efficient with the formation of otter hunts with hounds and more accurate guns, and the start of the industrial revolution led to rivers becoming increasingly polluted; both led to greater declines in otters (Jefferies 1989). Despite this, hunt records from the early 20th century showed a stable otter population, however by the mid 20th century hunt records indicate a decline in the otter numbers (Jefferies 1989). Loss of riparian habitat, water pollution, fish traps, road traffic accidents and general disturbance all contributed to this decline in otter populations in the UK (Jefferies 1989; Jefferies & Hanson 2001; Mason & Macdonald 2004).

In the late 1950s there was a sharp nationwide decline in otters attributed to the use of toxic organochlorine insecticides, (particularly dieldrin) (Jefferies 1989; Conroy & Chanin 2000, Mason & Macdonald 2004). Ironically, it was records from the otter hunts that first drew attention to the marked decrease in otter numbers throughout Britain in 1957 (Chanin & Jefferies 1978). The widespread nature of the decline, across Wales, Scotland and England, suggested a man-made cause rather than a disease epizootic. Spatial trends support this view and the decline is comparable with the agricultural practices of Britain, with the most dramatic reductions in the South and East of the UK, where arable farming dominates, and less severe declines in the west and north where farming is typically more pastoral (Chanin & Jefferies 1978; Strachen & Jeffries 1996).

Hunting records showed that voluntary bans in 1962 on the use of aldrin/dieldrin on spring sown cereals in some areas coincided with a reduced rate of population decline (Strachen & Jefferies 1996). In the west of Britain hunt records showed a population recovery just 2 years after a ban on the use of dieldrin in sheep dips (Jefferies & Hanson 2001). Some areas of Britain continued to use seed dressing to control wheat bulb fly (*Leptohylemia coarctata*) until a mandatory ban in 1975

(Jefferies & Hanson 2001), (e.g. East Anglia, Lincolnshire, east Midlands south-east Scotland) and in these areas otters became locally extinct (Crawford *et al.* 1979, Lenton *et al.* 1980).

The first otter surveys of Wales, Scotland and England (Crawford *et al.* 1979; Green & Green 1980; Lenton *et al.* 1980) confirmed that at that time the otter population was absent or sparsely distributed in much of lowland and central England (Crawford *et al.* 1979).

1.4.2. Otter population recovery

Detailed monitoring programmes have shown that since the late 1970's there has been a slow expansion of the otter population in the UK (Conroy & Chanin 2000). In England, Wales and Scotland otter surveys confirm that there has been an increase in otter distribution (Crawford 2003, Jones & Jones 2004, Strachan 2007), with recolonisation rates exceeding Biodiversity Action Plan (BAP) targets in Wales (Jones & Jones 2004).

The population expansion has largely been a result of natural re-colonisation, however in the Anglian area of England the increase has been aided by the success of populations that have received introduced animals (Crawford 2003). In this area of central and southern England, the species had been absent, or very rare (Strachen & Jefferies 1996), despite the availability of potentially suitable habitat (Jessop & Cheyne 1992). Small isolated populations that existed were considered likely to disappear completely (Jessop & Cheyne 1992). In response, a re-introduction project was established by the Otter Trust, which released 117 otters between 1983 and 1999 to increase the population, add genetic diversity and link fragmented non-viable populations (Jessop & Cheyne 1992). The increase in otter numbers in this area is believed to be due to both the success of the reintroduction project and by natural dispersal, from the west (south west England and the Welsh borders) and from the north (Scotland) (Coxon *et al.* 1999; Conroy & Chanin 2000).

1.5. Population monitoring: using spraints

In the UK otters are protected under Schedule 5 of the WCA 1981 and Schedule 2 of the Conservation (Natural Habitats etc) Regulations 1994 (Regulation 38). A national

Action Plan for the Otter was prepared by the UK Biodiversity Steering Group in 1995, part of this plan is to monitor populations and distribution of otters throughout the UK, including local surveys to monitor the expansion of fringe populations. As a result of its legal protection there are logistical and ethical problems which can hamper data collection. Otters live at low densities and are often nocturnal or crepuscular, so their study is not straightforward and monitoring techniques encounter many difficulties (Ruiz-Olmo *et al.* 2001b).

Spraints (faeces) are the clearest signal that otters inhabit a river system and their identification is the most frequently used technique in Europe for detecting the presence, abundance or relative abundance of otters. Otters leave spraints in visible spots (e.g. stones, rocks, tree-trunks) and in predictable places (e.g. under bridges, at junctions of rivers, in basins) which facilitates survey work. This allows the possibility to differentiate between positive and negative sites and to count the number of signs (Ruiz-Olmo *et al.* 2001b; Hung *et al.* 2004; Prigioni *et al.* 2005). Over the past 25 years detecting spraints has become the standard survey method and has been used on a large scale for the national surveys of Britain and Ireland (Chanin 2003) where otter presence is based on the percentage of positive sites. A disadvantage of using just spraints as an indicator is that they can only prove an otter has visited a particular site, but they cannot prove that it has not.

Alternative studies have been conducted using radio-tracking to monitor otter movement, focusing mainly on space use i.e. range sizes and rates of travel (Sjoasen 1997). This requires the trapping of individuals, which may be problematic due to the low capture rate, small population sizes, potential for injuries caused by handling and is illegal without a licence due to its endangered species status (Mills *et al.* 2000). Radio-tracking has been successful, but is more suited to monitoring introduced and translocated individuals, providing data without the risk associated with trapping wild animals (Sjoasen 1997).

Despite the limitation of using spraints for assessing otter populations they are the best evidence of the presence of this nocturnal, highly secretive animal (Kruuk 2006). Mason & Macdonald (2004) tested the method of predicting abundance of otters from spraints, using river catchments where colonisation by otters was assisted

by the release of a known number of captive animals. These authors showed that there was a relationship between the number of otters, the number of sprainting sites and spraint density. Although this method cannot be used to determine the exact number of otters present, it does provides evidence that the number of positive sites and the intensity of sprainting can be used to give a broad estimation of the performance of the otter population.

The Mason and MacDonald (2004) method is useful for monitoring otters in a single catchment, however, these are relatively large animals and can travel several kilometers of river a day, with a home range likely to extend over tens of kilometers (Chanin 2003). A single otter is therefore capable of marking many kilometers which poses some difficulties to monitoring (Ruiz-Olmo *et al.* 2001b; Chanin 2003).

Genetic analysis using non-invasive samples, such as faeces, allows DNA to be recovered from otter spraints and the genetic identity of individuals to be characterised, providing an abundance of information on the population (Chanin 2003; Dallas 2003; Hung *et al.* 2004). A positive identification provides the location of an individual at a particular point in space and time, but provides no information on whether it is resident or transient, adult or juvenile. A distinction must be made between areas otters frequently use and occupy, from areas through which otters move quickly (Ruiz-Olmo *et al.* 2001b). A pilot study was performed by Coxon *et al.* (1999) which allowed the identification of a minimum number of individuals; repeated identification allowed the calculation of home range sizes for one individual. There are problems however, associated with the use of spraints. For example, the collection of spraints involves a lot of effort, not only in the field to collect the spraint, but also in the lab taking many hours per DNA profile (Chanin 2003). Another limitation of this technique is the difficulty of obtaining a sufficient quantity and quality of DNA from spraints (Dallas *et al.* 2003; Hung *et al.* 2004), otter spraints are notoriously difficult to extract DNA from and must be collected fresh otherwise they may become degraded and unusable (Chanin 2003). Also genotyping of DNA from faeces is prone to several problems. Due to the scarcity of the template DNA, stochastic amplification of only one out of two alleles at a locus can cause 'allelic dropout'. Artifacts are sometimes generated during amplification to produce a 'false allele', and sometimes a 'counterfeit' or third allele is produced.

Contaminant DNA can cause serious problems when the target DNA is rare and may lead to mistyping of the genotype (Hung *et al.* 2004). These errors need to be detected and resolved and this can mean repeating the DNA amplification independently several times in order to obtain reliable genotypes (Taberlet *et al.* 1997; Dallas *et al.* 2003; Hung *et al.* 2004) making amplification from spraints laborious and expensive.

The otter population in Britain is recovering and as a result the likelihood of encounters with humans has increased. Over the last 15 – 20 years, mortality due to road traffic accidents has increased, to become one of the most important causes of death of otters in most European countries (Hauer *et al.* 2002; Philcox *et al.* 1999). Genotyping of otters from tissue is a much more reliable detection method and the collection of these carcasses and subsequent genetic analysis provide a unique tool for monitoring the otter population.

1.6. Otter population fragmentation and its genetic consequences

Dallas *et al.* (2002) studied the genetic structure of British otter populations using microsatellite markers. They had two major findings, that "*populations in Scotland, regarded as continuous according to distributions of signs, were to some extent genetically subdivided and populations in mainland Scotland showed a strong pattern of isolation by distance (IBD)...*" and "*populations in southern Britain regarded as biologically equivalent to those in Scotland contained significantly reduced levels of microsatellite polymorphism*". Statistical assignment tests performed by Dallas *et al.* (2002) suggest there was no gene flow between populations in Scotland, Wales and SW England at the time of study.

The different levels of microsatellite polymorphism shown by Dallas *et al.* (2002) were associated mainly with the discontinuity between populations in mainland Scotland, and those in Wales and SW England. It was unclear whether the reduced microsatellite polymorphism in Wales and SW England was the result of recent or long-term population fragmentation (Dallas *et al.* 2002). It was suspected that the reduced polymorphism reflected a long history of low effective population size rather than recent declines (Dallas *et al.* 2002). However, assessment of the loss of variability was hampered by the lack of information about the genetic composition of

the same populations prior to their fragmentation and bottleneck (Pertoldi *et al.* 2001).

Recent *versus* long-term fragmentation was addressed by Pertoldi *et al.* (2001) in Denmark, where otter populations have undergone similar declines to the UK. A comparison between microsatellite DNA variation in samples from the contemporary otter population and historical (museum) specimens collected between 1880 and 1960, showed surprisingly few signs of a recent bottleneck (Pertoldi *et al.* 2001). The study also showed that some geographical subdivision was present in historical specimens. There were indications of a drastic population decline, but this was shown to have happened on a time scale covering hundreds or thousands of years, not during the last few decades. It was concluded that northern European otter populations generally exhibit low genetic variability, due either to post-glacial founder events or a decline which started *ca.* 2,000-3,000 years ago. These findings support Dallas *et al.*'s (2002) hypothesis that the low genetic variation found in the otter populations of the UK is the result of historical rather than recent population declines.

It is nonetheless important that the long-term viability of UK otter population is likely to depend upon recolonisation and the establishment of corridors for gene flow between isolated populations. Mitigation should therefore be considered against the potentially negative effects of population fragmentation.

1.7. Cardiff University Otter Project – CUOP

Since 1992, the Environment Agency has funded the collection and post mortem examination of otters found dead in Wales and England. Post mortem examinations are carried out at Cardiff University (1994-present), and at the Veterinary Investigation Centre in Cornwall (1988-2007). Measurements are taken and tissue samples are retained; data and samples are used for a number of research objectives (Simpson 1998; Bradshaw & Slater 2002, Chadwick 2006, Sheppard-Smith et al. 2009) or archived. Importantly, the exact geographic location (grid reference) where the otter was found is recorded by the collector and provides the opportunity for spatial analyses. Muscle tissues from these individuals were used for DNA extraction and further genetic analysis.

1.8. Molecular approaches

Microsatellites consist of tandemly repeated units, generally less than 5bp (base pairs) in length such as (TG)n or (ATT)n (Bruford & Wayne 1993). These repeat units are often highly polymorphic with many different alleles segregating in a population. Due to their attributes they have been used in many different areas of study ranging from ancient and forensic DNA studies, to population genetics and conservation/management of biological resources (Jarne & Lagoda 1996; Zhivotosky & Feldman 1995; Zane *et al.* 2002). Locus-specific PCR primers are designed to recognise sequences flanking the tandem repeats (Bruford *et al.* 1996). Literature searches provided information on primers that had already been described for the Eurasian otter and 21 loci were chosen for the work in this thesis from these papers (Dallas & Piertney 1998; Dallas *et al.* 1999: Huang *et al.* 2005) to produce microsatellite genotypes of individuals. This number was reduced to 15 after rarefaction analysis was used to identify the combination of loci which most efficiently recovered accurate relatedness and genetic diversity estimates (Altmann *et al.* 1996; Smith *et al.* 1997; Kays *et al.* 2000). These fifteen loci were allocated into three PCR multiplexes for further analysis. More information about optimisation of PCRs Rarefaction anlysis and development of multiplex PCRs can be found in chapter 2) and in the Appendix (Hobbs *et al.* 2006).

1.9. Landscape genetics

"The collection of genetic data from many individuals of known geographic origin, in combination with recently developed statistical tools, potentially allows the identification of spatial genetic patterns"

(Manel *et al.* 2003).

Landscape genetics is an emerging field in and involves the combination of molecular ecology and landscape ecology (Manel *et al.* 2003; Holderegger & Wagner 2006). This approach enables spatial mapping of allele frequencies and potential correlation of microevolutionary processes such as genetic drift, gene flow and selection with landscape or environmental features (Manel *et al.* 2003; Berthier *et al.* 2005).

Landscape genetics studies are rapidly increasing in number, due both to advances in molecular genetic tools, and the development of increasingly powerful statistical approaches, using coalescent simulation (Beaumont & Rannala 2004), individual assignment (e.g. Faubet *et al.* 2007) and methods to estimate kinship (e.g. Goodnight & Queller 1999; Konovalov *et al.* 2004).

1.10. Bayesian clustering

In this study Bayesian clustering techniques were used; which use individual multilocus genotypes derived from multiple microsatellite markers to assign individuals to clusters, on the assumption that markers are in Hardy Weinberg and linkage equilibrium within each randomly mating subpopulation (Pearse & Crandall 2004; Manel *et al.* 2005; Latch *et al.* 2006). Bayesian Clustering techniques are described in more detail and compared in Chapter 2.

Bayesian Clustering techniques are used with increasing frequency in the population genetic literature, however, they have not been thoroughly compared using a wild georeferenced data set. In Chapter 2 samples were used from the Wales and Borders area to compare the ability of four of these programs to estimate the number of populations (K) and to compare how similar they are in their assignment of individuals to identified populations. In Chapter 3 three of the Bayesian Clustering techniques were used to identify population structure of the entire UK dataset and investigated a novel progressive partitioning technique to identify populations and compare and combine the outputs of each of the softwares to get the most robust clustering solution.

1.11. The use of geographical information systems (GIS) in landscape genetics.

Geographical information systems (GIS) are powerful packages and often underutilised in conservation genetics. They can be used in conjunction with statistical tests to visualise spatial genetic patterns, by overlaying landscape variables and genetic data (Manel *et al.* 2003). An important feature of this approach is that it allows the user to visualise the distribution of individuals and their population assignments (as derived from Bayesian clustering for example). Being able to see how two populations are distributed on a map may give the researcher the opportunity to identify cryptic genetic discontinuities (barriers to gene flow) across

populations which have no obvious cause and can identify secondary contact between previously isolated populations. This spatial delineation of genetic discontinuities within a species potentially allows the user to define operational units, important for management purposes (Manel *et al.* 2003) and to identify isolated populations, routes of and barriers to dispersal. This technique has been implemented in the following chapters, where Bayesian clustering has been used to assign individual otters to populations and GIS has been used to display these populations and provide the reader with an understanding of their distribution, also allowing the user to compare population assignments between programs.

Typically in a natural continuous population where dispersal is limited, Isolation by Distance (IBD) arises. IBD occurs because levels of gene flow tend to decrease with increasing geographical distance, resulting in increasing genetic differentiation between individuals (Broquet *et al.* 2006). Genetic distance can be identified between populations by using Wright's F_{ST} (1951), and by analyzing the pair-wise estimates of genetic differences between individuals (Rousset 2000). Correlations between genetic and geographic distances matrices using Mantel tests have been used in a number of studies (Berthier *et al.* 2005; Broquet *et al.* 2006; Diniz-Filho 2008; Latch *et al.* 2008) and provide an important method for identifying IBD (Wright 1943). Linear geographic distance separating populations however, may have less influence on creating and maintaining genetic structure than features of the environment that affect dispersal (e.g. slope, roads and other climatic and topographical features) (Kozak *et al.* 2008). GIS technology can also test informative hypotheses concerning the effect of landscape structure on the movement of organisms and how organisms perceive habitat connectivity (Holderegger & Wagner 2006). Many landscape genetic studies have used GIS-based data to show that genetic distance is influenced to a great extent by topography, habitat type and other parameters (Cushman *et al.* 2006; Kozak *et al.* 2008; Perez-Esona *et al.* 2008; Zalewski *et al.* 2009). Landscape genetics aims to understand which factors are structuring genetic variation at both the population and individual levels (Manel *et al.* 2003; Storfer *et al.* 2007) by integrating the genetic, biological and environmental variation with spatial statistics (Dionne *et al.* 2008).

In Chapter 4, the Wales and Borders otter population data set is used to illustrate this approach, and to identify whether selected environmental variables correlate with genetic distance and hence gene flow / dispersal of individual otters. Cushman *et al.* (2006) provide a good example of this technique in a study on black bears (*Ursus americanus*), a factorial, multimodel approach is used to evaluate alternative hypotheses and identify the combination of environmental factors in a landscape that allow connectivity between individuals / populations and appear to drive gene flow.

Connectivity between populations depends not only on the landscape structure but also on the mobility of the organism (Adriaensen *et al.* 2003). These factors in combination give rise to the concept of landscape connectivity, defined by Taylor *et al.* (1993) as 'the degree to which a landscape facilitates or impedes movement among resource patches'. It is this interaction that may strongly shape evolutionary processes by affecting dispersal and thereby effective movements (i.e. movement followed by successful reproduction), which can drive gene flow across a landscape (Coulon *et al.* 2006). Landscape features that influence effective movement can be identified by studying gene flow in relation to landscape structure. In order to do this a cost can be applied to a landscape feature, the magnitude of which depicts how much it impedes or facilitates movement of individuals of that species.

1.12. Aims of this thesis

Using genetic data available from otter carcasses found and collected in the UK since 1994, the genetic structure of remnant and newly established populations will be investigated. This information can be used to analyse the origin, rate and direction of recolonisation into formerly vacant regions using spatial genetic analysis and population assignment tests (e.g. Piry *et al.* 2004).

Bayesian clustering techniques are untested on georeferenced datasets in wild populations and this study will be among the first to use such a population dataset. Chapter 2 aims to use the population dataset from the Wales and Borders area to compare and assesses the performance of and to evaluate the inferred genetic structure produced using approaches implemented in the software STRUCTURE, PARTITION, BAPS and GENELAND in a landscape context.

Chapter 3 aims to identify population structure within a second dataset, representing a much larger proportion of the UK (England, Wales and part of Scotland), to identify population units and the degree of gene flow between them and to investigate the relative contributions of expansions from otter strongholds and population reinforcement by captive bred otters on the population in lower central England. The results of multiple Bayesian clustering methodologies will be compared and combined using a novel progressive partitioning method. The population clustering identified is expected to represent the distribution of the known stronghold areas but to also unveil cryptic substructure within these regions. Where sub-structuring exists, GIS will be used to visualise spatial genetic patterns, to identify whether genetic boundaries are associated with physical obstacles such as roads and other landscape features.

Finally (Chapter 4) exploratory analysis is used to test the influence of landscape features on the genetic structure of the otter population in the Wales and borders region. Using recently developed techniques in landscape genetics (looking for correlations between landscape features and gene flow at different spatial scales), the effects of 5 landscape features on gene flow are explored. The landscape features are hypothesised to facilitate (Rivers, Broadleaf woodland) and resist (Slope, Upland habitat, Anthropogenic factors) dispersal.

Chapter 2

Examining the Robustness of Bayesian Clustering Algorithms at a Fine Spatial Scale Using a Recovering Otter (*Lutra lutra*) Population in the Wales and Borders Area

2.1 Abstract

Landscape genetic analysis was applied to a large georeferenced genetic sample of otters (*Lutra lutra*) that were collected as a result of road casualties in Wales, UK. The performance of a number of recently developed spatially explicit and non-spatial Bayesian clustering approaches (implemented in the programs STRUCTURE, PARTITION, BAPS and GENELAND) was assessed and compared using this dataset. The programs were compared in terms of inference of the number of populations/clusters (K) and assignment of individual genotypes to populations. Results of population assignment were compared using interpolation maps of posterior probabilities and by assignment similarity coefficient. Two presumed migrant individuals were shown to affect the estimation of K in some programs but not in others, however, estimation of K agreed among most programs when these individuals were removed from the analysis. The patterns resulting from the interpolation maps were remarkably similar between the methods with the exception of the BAPS4 NON-SPATIAL model, despite only 49 of 216 individuals being assigned to the same population cluster by all programs. The results suggest that more than one method should always be used to produce optimal partitions, with the choice of programs depending on the type of dataset used and the research objectives.

2.2 Introduction

Traditional methods for identifying populations and characterising genetic differentiation among populations relied upon *a priori* groupings of individuals, however, identifying populations in advance may be undesirable due to potential bias arising from cryptic spatial structure and unidentified migrants (Rousset 2000; Sumner *et al.* 2001; Manel *et al.* 2003). Analytical techniques using Bayesian Clustering Algorithms avoid the need to assign individuals to populations in advance. Instead, they use individuals as the study unit and use each individual's multilocus genotype to assign clusters on the assumption that markers are in Hardy Weinberg and linkage equilibrium within each randomly mating subpopulation (Manel *et al.* 2003). These Bayesian methodologies allow complex questions to be addressed, using sometimes computationally intensive simulations of the coalescent (Beaumont & Rannala 2004) to approximate the posterior probabilities of population genetic parameters (Pearse & Crandall 2004). There are a number of these methods emerging, but there is little guidance for researchers' as to what programs to use and in what combination. STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003) is the most commonly used, probably as a result of being the first such program, now with a proven track record it has become the standard software of choice. Newer models have been released and their authors tend to compare them against STRUCTURE (Guillot *et al.* 2005a; Dawson & Belkhir 2001; Corander *et al.* 2004; Chen *et al.* 2007) and with each other. Some reviews (e.g. Pearse & Crandall 2004; Manel *et al.* 2005; Latch *et al.* 2006) describe and compare these techniques, noting that these methods are relatively untested, and that comparative analyses are lacking, particularly for 'real' datasets. This has been addressed; both Rowe and Beebee (2007) and Frantz *et al.* (2006, 2009) used three methods on their respective wild population datasets and found that each model gave differing estimates of the number of clusters (K). This demonstrates that whilst these programs identify population structure using the same principles, they can differ in their clustering results, and even differ between runs of the same program. It is recommended therefore that multiple runs be performed for each program, and that several programs are compared; the most likely clustering solution is the one that most consistently occurs (Pearse & Crandall 2004; Latch *et al.* 2006; Chen *et al.* 2007).

Comparing several Bayesian Clustering programs is time consuming and complicated, and despite these recommendations there are studies that still rely on the results of just one program (Crompton *et al.* 2008; Crosby *et al.* 2008; Johansson *et al.* 2008). Increasingly sample datasets are georeferenced and more recently developed Bayesian Clustering Algorithms are able to incorporate these spatial data into the analysis, on the assumption that some spatial dependence is present among individuals (Guillot *et al.* 2005b). As a result information on how individuals are spatially organized is added *a priori*. The use of priors however, may seriously bias the parameter space searches of these programs (Mank & Avise 2004), and creates dilemmas as to the strength of prior evidence, and how heavily it should be weighted (Mank & Avise 2004). In some cases a dataset may be insufficient to override such *a priori* assumptions; the use of informed priors can result in assignments that merely recover the information given in the priors (Mank & Avise 2004). To err on the side of caution many studies run non-spatial alongside spatial analysis and compare the outputs (Croteau *et al.* 2007; Barnett *et al.* 2008; Barbara *et al.* 2009). There are few studies that use three or more programs and those that do tend to be reviews of Bayesian Clustering Algorithms, for example Latch *et al.* (2006) compared the relative performance of three non-spatial Bayesian clustering programs (STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003), PARTITION (Dawson & Belkhir 2001) and BAPS (Corander *et al.* 2003) using a simulated dataset. Chen *et al.* (2007) added to this study using the same dataset to compare programs that included spatial coordinates (GENELAND (Guillot *et al.* 2005a), GENECLUST (François *et al.* 2006) and TESS (Chen *et al.* 2007)). Chen *et al.* (2007) found that Bayesian clustering programs using spatial data are as reliable as nonspatial Bayesian clustering programs, particularly when the number of polymorphic loci available to the study is limited. Spatial Bayesian clustering appears to work best when populations are separated by simple shaped boundaries with no recent gene-flow (Chen *et al.* 2007), whist non-spatial Bayesian clustering (STRUCTURE) outperformed others when the shape of the contact zone became irregular. To add to the complication there has been much debate about the effect of isolation by distance (IBD) (Wright 1943), and whether these programs are actually identifying true clusters or are artificially detecting structures emerging from uneven sampling along a cline (Serre & Pääbo 2004; Rosenberg *et al.* 2005; Frantz *et al* 2009).

The authors of STRUCTURE concede that there may be difficulties detecting structures if IBD is present (Pritchard & Wen 2003). To combat this Frantz *et al.* (2006) recommend using spatial data in the analysis and found that genetic clusters identified by BAPS 4.1 spatial were the most biologically meaningful out of three models tested, and that the model was robust when faced with isolation-by-distance relationships in the genetic data set. Chen *et al.* (2007) found that all the Bayesian clustering methods they tested that included spatial data as a prior could identify a cline, however they found STRUCTURE, despite not incorporating spatial data, showed the best estimation of a cline to the actual allele frequencies. The ability of the program to identify true clusters and not artificial clusters along a cline can also depend on study design and the number of markers (Corander *et al.* 2004; Serre & Pääbo 2004; Rosenburg *et al.* 2005). Frantz *et al.* (2009) recommend caution when interpreting results of populations characterised by IBD as this can lead to an overestimation of genetic structure and the identification of erroneous population units.

These previous assessments of performance have used simulated datasets that conformed to the assumption of genetic and demographic equilibrium, an assumption that is usually violated by real populations. It is therefore of importance to investigate the identification of, and assignments to clusters of a wild geo-referenced dataset. In this scenario the true number of populations and their geographical partitions are unknown; therefore selection of the best program will depend on known population history, which may be confounded by the possibility of cryptic barriers to dispersal, therefore the interpretation of the resulting clusters and assignments should be made carefully from several programs. The current study aims to compare some of the more widely used software programs available for Bayesian clustering (STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003), PARTITION (Dawson & Belkhir 2001), BAPS (Corander *et al.* 2003; 2004) and spatially explicit landscape genetics software, here GENELAND (Guillot *et al.* 2005a)) to identify population structure in a wild geo-referenced dataset.

2.2.1. Differences between Bayesian clustering software

These programs all follow the Bayesian framework described above, and are expected to provide similar outcomes. Differences between the models are summarised below:

(1) Estimation of (*K*): STRUCTURE infers the likelihood of a number of clusters (populations) based on an *ad hoc* method using the posterior probability for different numbers of putative populations specified by the user (Manel *et al.* 2005) whereas BAPS, GENELAND and PARTITION simultaneously assess the likelihoods for a range of *K* values up to a maximum (up to and including the total number of individuals) specified by the user.

(2) Assumptions of ancestry: STRUCTURE, BAPS4 and GENELAND allow individuals to be specified as being of mixed ancestry, proportionally assigning an individual's genome into clusters, while PARTITION and BAPS 2 assume all individuals to be of pure ancestry.

(3) Use of spatial information: STRUCTURE, PARTITION and BAPS 2 take no consideration of geographic location during analysis. GENELAND and BAPS4 have the option of using spatial coordinates to favour partitions that are spatially related. A new version of STRUCTURE is now available that can incorporate spatial information (Hubisz *et al.* 2009) but this was not available at the time analyses were conducted.

Finally, GENELAND and BAPS4 produce spatial plots when spatial data are provided, while STRUCTURE, PARTITION and BAPS 2 do not. To enable visual comparison, the outputs (assignment of individuals to different populations/ posterior probability) can then be mapped using Geographical Information Systems (GIS) such as ArcMap v 9.2 (ESRI 2007).

2.2.2. Aims

The current study utilises a large georeferenced sample of multilocus genotypes from the population of the Eurasian otter, *Lutra lutra,* in Wales, UK. The aim of the study was to compare and assesses performance and evaluate the inferred genetic structure in a landscape context produced using approaches implemented in the software STRUCTURE, PARTITION, BAPS and GENELAND.

2.3 Methods

Over the past two decades, the Environment Agency along with other regional organisations have recorded the geographical location and collected otter road casualties throughout England and Wales. Post mortem examinations have been conducted by Cardiff University Otter Project (England and Wales, 1994-ongoing), or by the Wildlife Veterinary Investigation Centre (southern England, 1988-2007). Muscle samples have been removed from most individuals and stored in ethanol at -20°C.

As a first step a preliminary study was carried out on 100 individual otter samples chosen randomly from Wales and bordering catchments to establish molecular methodologies. Once the PCR techniques were optimised 216 otters from the Wales and Borders area were used for genetic analysis.

2.3.1. DNA extraction

DNA was extracted from muscle tissue, using the QIAGEN DNeasy tissue kit following the 'isolation of total DNA from animal tissues' protocol (QIAGEN, #65906).

2.3.2. Primers

Using primers that have been designed for the Eurasian otter, the genotypes of individuals were identified for 22 loci. The microsatellite loci used were lut435, 453, 457, 604, 615, 701, 715, 717, 733, 782, 818, 832, 833 (Dallas & Piertney 1998) lut902 (Dallas *et al.* 1999) and 04OT02, 04OT04, 04OT05, 04OT07, 04OT14, 04OT17, 04OT19 and 04OT22 (Huang *et al.* 2005). (Following preliminary analyses, the number of loci will be reduced using rarefaction analysis, see section 2.3.4).

2.3.3. Multiplex design

For more efficient analysis, four PCR multiplex groups were designed and optimised. The Forward primers of each primer pair were labelled with a fluorescent dye (Ned,

Hex or Fam). The dye used to label each primer was chosen as part of the design of the multiplex group which also took into account the allele sizes, to ensure that each locus was distinct. Two multiplex groups contained five primer pairs and two contained six. PCR reactions were conducted with a QIAGEN Multiplex PCR kit following the 'amplification of microsatellite loci using multiplex PCR' protocol (QIAGEN, #206143). Amplification of DNA extracts was performed using a GeneAmp® PCR system 9700 (Applied Biosystems) in 6.5 µl reactions containing DNA template, 1x QIAGEN Multiplex PCR Master Mix (containing HotStarTaq® DNA polymerase, Multiplex PCR buffer (contains 3 mM $MgCl_2$) and dNTP Mix), 10x Primer Mix (0.2 µM of each primer) and sterile water). The PCR profile was identical for each multiplex and included an initial denaturation step of 95 °C for 15 mins, 29 cycles with 94 °C for 30 s, 58 °C for 90 s and 72 °C for 1 mins and a final extension of 60 °C for 30 mins. PCR products were analysed using an ABI PRISM® 3100 Genetic Analyser (Applied Biosystems) and gel analysis was performed using the software Genescan v 3.7 and Genotyper version 3.6 (Applied Biosystems).

2.3.4. Rarefaction analysis

A random sub-sample of 100 otters from the Wales and Borders region were genotyped for all 21 loci for rarefaction analysis using the methods described above. These genotypes were input into the program POPASSIGN version 4.3a (S.M. *Funk*, Zoological Society of. London, as used in Utami *et al.* 2002, and Goossens *et al.* 2003) to conduct rarefaction analysis. Rarefaction analysis aims to identify the combination of loci which most efficiently recover data, enabling accurate relatedness and genetic diversity estimation (Altmann *et al.* 1996; Smith *et al.* 1997; Kays *et al.* 2000). In POPASSIGN, relatedness is assessed by simulating first order relative datasets based on the observed allele frequencies, estimating 'Queller & Goodnight (1989) relatedness' (*R*) using the simulated data, and repeating the process for all possible combinations of loci to be used. Standard errors are generated by permuting loci without replacement. The number of loci was increased by addition without replacement until all 22 loci were selected (Girman *et al.* 1997; Kays *et al.* 2000). This procedure was repeated 1000 times. The mean difference in relatedness estimate *R* for different numbers of loci and jackknifed standard errors

were calculated as the average of absolute differences in *R* values calculated between steps (Altmann *et al.* 1996).

2.3.5. Genotyping the Wales and Borders dataset

Based on rarefaction analysis 15 loci were chosen in three multiplexes for the rest of the study (Table 2.1), reducing the number of PCRs, fragment analysis runs and thus the cost of the analysis per individual without detracting from the results. Multiplexes were run under the same conditions as the preliminary study, however, primer concentration differed for some primers (Table 2.1). During the study fragment analysis switched from the ABI PRISM® 3100 Genetic Analyser (Applied Biosystems) to the ABI 3130 and gel analysis was performed using the software Genemapper ™ (Applied Biosystems). Allelic ladders were created with known sized DNA fragments and run alongside analysis on the new machine and software. Differences in allele sizes were identified and the original dataset was adapted and merged with the new scoring regime.

2.3.6. Summary genetics

Genotypic distribution for conformance with Hardy–Weinberg equilibrium (HWE) was tested using GENEPOP 3.4 (Raymond & Rousset 1995) with all probability tests based on the Markov chain method using 1,000 de-memorization steps, 100 batches and 1,000 iterations per batch. The levels of genetic diversity within the population were estimated by calculating observed (Ho) and expected (He) heterozygosities, and the average number of alleles (A) using the GDA software (version 1.1; Lewis & Zaykin 2001) (Table 2.1).

Table 2.1. Properties of 15 microsatellite loci used in the current study. The loci in each multiplex were amplified in a single PCR. (*A*), number of alleles; (*He*) expected heterozygosity; (*Ho*) observed heterozygosity. Summary statistics are for the Wales and Borders dataset with a sample size of 216 individuals. Loci marked with an asterisk deviated from Hardy-Weinberg proportions at the α 0.05 level.

	Dye	Annealing (Tm)	Primer mix (m)	Number of alleles (A)	Allele size range (bp)	Expected Heterozygosity (He)	Observed heterozygosity (Ho)
Multiplex 1		58°C					
Lut435*	Fam		0.16µm	4	123-135	0.45	0.42
Lut453	Hex		0.2µm	5	119-133	0.29	0.25
04OT05	Hex		0.2µm	5	171-187	0.69	0.67
lut717*	Ned		0.2µm	5	175-199	0.43	0.37
04OT22*	Fam		0. 16µm	7	142-166	0.54	0.49
Multiplex 2		58°C					
lut604*	Fam		0. 16µm	5	127-137	0.63	0.53
lut733*	Fam		0. 12µm	7	156-182	0.46	0.44
lut615	Fam		0. 16µm	7	216-229	0.63	0.56
lut902*	Hex		0.2µm	5	145-170	0.66	0.59
lut782	Ned		0.2µm	4	161-192	0.48	0.46
Multiplex 3		58°C					
lut818	Fam		0. 16µm	7	154-188	0.67	0.66
lut701	Fam		0.16µm	3	201-210	0.43	0.44
lut833	Hex		0. 16µm	6	154-174	0.73	0.70
lut715	Hex		0. 16µm	5	199-216	0.57	0.55
lut832*	Ned		0.2µm	5	181-197	0.36	0.29

2.3.7. Genotyping error rate

Microsatellite genotyping errors might greatly bias results, the rate of these errors should always be assessed, even when working with good-quality tissue samples (Bonin *et al.* 2004; Hoffman & Amos 2005). All the data were double-checked in order to identify and eliminate errors that had occurred during data entry and scoring of alleles by hand. For each multiplex 21-27 samples (9.7-12.5%) were chosen randomly from this database, and re-genotyped. Allelic mismatches were counted by comparing these genotypes to the previous ones (Bonin *et al.* 2004; Hoffman & Amos 2005). Error rates were summarized as the number of errors per allele, i.e. the number of incorrect alleles divided by the total number of alleles. Individuals with missing data were removed from the analysis leaving 216 individual genetic profiles consisting of 15 loci.

2.3.8. Bayesian clustering analysis

The genotypes for 216 individuals at 15 loci were analysed for population structure using the four programs STRUCTURE, BAPS, PARTITION and GENELAND. In general, for all programs the author's guidelines were followed and default values taken where applicable, when spatial information could be used both spatial and nonspatial analysis were conducted.

2.3.8.a. STRUCTURE

Analysis was performed in STRUCTURE version 2.1 (Pritchard *et al.* 2000; Falush *et al.* 2003; Pritchard & Wen 2003) modified to take advantage of the distributed computing software CONDOR (Litzkow *et al.* 1998) and therefore enabling a large number of iterations for the Markov Chain Monte Carlo (MCMC) algorithm. 1,000,000 replicates were performed following a burn-in of 100,000 replicates, using the admixture model and assumed correlated allele frequencies. STRUCTURE was run with the parameter set for K from one to seven, with five independent runs of each K. The estimated log probability of data Pr(X|K) has been used to estimate the most likely number of clusters (Pritchard *et al.* 2000). However, this method has been recently augmented by that recommended by Evanno *et al.* (2005) which uses the second order rate of change of the likelihood function with respect to K. Both approaches were used here.

2.3.8.b. PARTITION

For analysis using PARTITION (version 2; Dawson & Belkhir 2001), a maximum number of source populations was assumed at $K = 7$ (to ensure that this number is greater than the expected number of populations). 100,000 iterations in the Markov Chain were used with a burn-in length of 1,000. The prior distribution of population allelic diversity was set to ($\theta = 1$) and a uniform prior probability distribution on K ($\mu = 1$). The Bayes factor is a likelihood ratio where a value greater than one provides evidence favouring the existence of a single random mating (panmictic) population, against the alternative of multiple source populations. PARTITION uses an agglomerative hierarchical clustering algorithm (exact linkage) to construct a binary

tree (Dawson & Belkhir 2001). When displayed as a co-assignment dendrogram, individuals can be defined into groups based on visual inspection.

2.3.8.c. BAPS

Investigation of population structure was conducted using BAPS (Version 4.13; Corander *et al.* 2003, 2004), and BAPS (Version 2). BAPS (Version 2) uses an exact Bayesian analysis by enumerative calculation when the known number of populations is nine or less and uses a MCMC algorithm for estimation when the number of populations exceeds nine. BAPS 2 clustering was done at an individual level using a similar method to PARTITION. The default value *n* (number of individuals) was used as the number of clusters *k* in the initialisation. The basic clustering fit model was chosen with 100,000 iterations with a burn-in of 50,000 and thinning of 5. The best visited partitions are produced and the results can also be displayed as a dendrogram.

BAPS4 uses a stochastic optimisation algorithm to infer the posterior mode of the genetic structure (Corander & Marttinen 2006). BAPS4 genetic mixture analysis was carried out by clustering at the individual level, using both spatial and non-spatial models. The non spatial model assumes that the prior distributions for clusterings are uniform having at most *K* clusters (Corander *et al.* 2003, 2004). The spatial model uses the individual geo-references to assign a biologically relevant non-uniform prior distribution over the space of clustering solutions, thus expecting the underlying clusters to be spatially smooth, with at most *K* clusters (Corander *et al.* 2003, 2004). The program was run with a vector of values for the maximum number of clusters (*K*) with five replicates of $K = 5$, 10 and 15. After all the *K* values were processed, the stored results were merged based on the logML values with the best 10 partitions displayed. BAPS4 can also estimate individual admixture coefficients from each cluster. Admixture analysis is performed using the results from the mixture analysis. The minimum number of individuals in a cluster was set to 1 and default values were used for admixture priors. 100 iterations were used to estimate the admixture coefficients, with 200 reference individuals from each population and 20 iterations to estimate the admixture coefficients for the reference individuals. Analysis using values higher than the default provided similar results (not shown).

2.3.8.d. GENELAND

Finally, analysis was also conducted using GENELAND (R package), which can make use of spatial information to favour partitions that are spatially organised, but can also infer populations without spatial information. Both models were used to aid comparison with other programs. GENELAND can use two allele frequency models, the Dirichlet (λ) distribution which assumes that allele frequencies in different populations are independent (Pritchard *et al.* 2000), and the Falush model where allele frequencies are not independent among populations (Falush *et al.* 2003). The Dirichlet distribution was used here, following the advice of Guillot *et al.* (2005a) using spatial data (Spatial D-model) and non-spatial data (non-spatial D-model). To identify *K*, five independent MCMC runs were performed to verify the consistency of results. The following parameters were used: Priors on *K*-uniform between 1 and 7 with 200,000 MCMC iterations. The parameter describing the amount of uncertainty to spatial coordinates was set at 0.3 (author recommendation due to duplication of some georeferences), with maximum rate of Poisson process fixed at 216 and maximum number of nuclei in the Poisson-Voronoi tessellation of 648. The number of populations were inferred from the modal *K* values of the five runs, and used as a fixed value in a further five MCMC runs with 50,000 MCMC iterations with other parameters identical as for variable *K*. The posterior probabilities of population membership for each individual were calculated using a burn-in of 10,000 and an average was taken for each individual over the 5 runs.

2.3.9. CLUMPP

In Bayesian Clustering the runs of each model may give slightly different solutions (Jakobsson & Rosenberg 2007), CLUMPP v 1.1 (Jakobsson & Rosenberg 2007) a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure was used to calculate the average matrix of ancestry membership. When the optimum number of clusters was calculated for each model, 5 independent runs were performed at that *K* value to identify individual assignments, and the averaged assignment for each population was calculated. CLUMPP (Jakobsson & Rosenberg 2007) also takes into account label-switching before averaging ancestry membership.

2.3.10. GIS: ArcMap v 9.2

The spatial coordinates of the data used in spatial analysis and for GIS were supplied by carcass collectors as UK National Grid references, which were converted to X and Y coordinates with at least a 1km resolution. ArcMap v 9.2 (ESRI 2007) was used to spatially map the posterior probabilities of assignment produced for each individual by each model, with a map of Wales (shape files provided by the Environment Agency) used for reference. The point-based assignment data from each individual was converted to a continuous surface using the interpolation distance weighted function in ArcMap v 9.2. Based on the values at sampled locations interpolation allows the estimation of a parameter at an un-sampled location (Lindley & Walsh 2005), providing a mapped output and thus a visual depiction of spatial patterns.

2.3.11. Detecting clinal variation

There is the possibility that the clusters identified are actually artificial structures produced by clines in allele frequencies. Evidence of clinal variation can be identified by plotting the population adherence sorted by Q (the most likely population for any individual) (Sahlsten *et al.* 2008). Clinal variation may not be possible to identify with only one run, and the average of multiple runs must be used to detect this type of variation (Chen *et al.* 2007). Clinal variation was identified by using the average population membership coefficients created in CLUMPP and displaying them using EXCEL software.

2.4 Results

2.4.1. Rarefaction analysis

The difference between consecutive sampling in the outcome of *R* was expressed as a function of the total number of loci drawn, and showed that mean and variance estimates of relatedness *(R)* stabilised after 15 loci (Figure 2.1.). Therefore 15 loci can be used to provide consistent measures of relatedness.

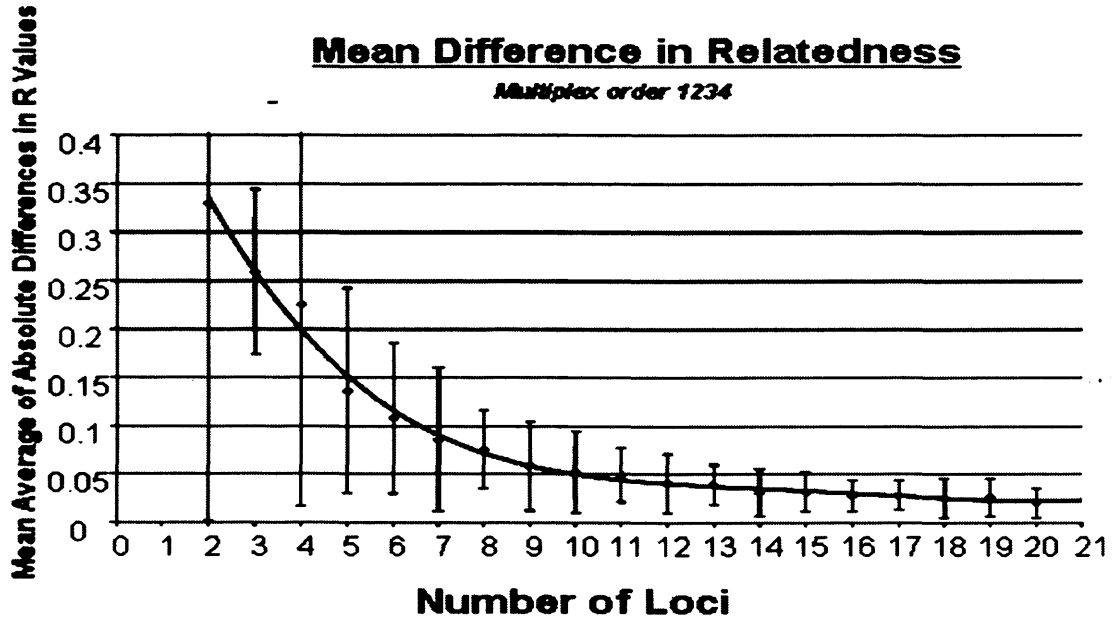


Figure 2.1. The decrease in the mean difference between consecutive relatedness estimates as a function of the number of microsatellite loci analysed.

2.4.2. Genotyping error rate

Error rates were summarized as the number of errors per allele, i.e. the number of incorrect alleles divided by the total number of alleles = 0.035 over all loci. This equates to just seven genetic profiles in the total data set of 216 to have at least one allelic mistyping, which is low enough to not affect the results significantly.

2.4.3. Genetic variability

The microsatellite loci for Wales and Borders otters are polymorphic with an average of 5.3 alleles per locus (maximum = 7, minimum = 3). The Wales and Borders otter population had an average observed heterozygosity (*Ho*) of 0.49 and an average expected heterozygosity (*He*) of 0.53. Significant deviations from HWE were observed at seven out of fifteen loci (Table 2.1).

2.4.4. Number of clusters

Table 2.2. shows the most likely number of clusters inferred using the approaches and software described above. PARTITION and BAPS 2 assign individuals to populations categorically, whilst GENELAND, STRUCTURE and BAPS4 use admixture analysis and allow the assignment of a proportion of the genome to more than one population. It is important to note that STRUCTURE's 'no admixture' model and BAPS4 'mixture analysis', both provide a categorical assignment but these analyses are not used here. Two individuals (UWCRef 433 and 441) strongly influenced the outcome of some of the models (see Table. 2.2. Figure 2.2.) and were assigned to their own clusters in all programs except PARTITION and STRUCTURE. When these individuals were removed there was strong similarity among all approaches with K = 2 being found as the most likely number of populations with the exception of the BAPS4 NON-SPATIAL model (Table 2.2).

Table 2.2. Estimates of the number of otter population clusters (*K*) for each of the Bayesian clustering models with and without individuals 433 and 441.

Model	Estimation of *K*	Estimation of *K* without individuals 433 & 441
PARTITION	2	2
STRUCTURE	2	2
BAPS 2	4	2
BAPS4 SPATIAL	3	2
BAPS4 NON-SPATIAL	7	7
GENELAND SPATIAL D MODEL	4	2
GENELAND NON- SPATIAL D MODEL	5	3 (+ 3 ghost pop)

STRUCTURE infers the number of clusters by comparing the posterior probability for different numbers of putative populations specified by the user (Manel *et al.* 2003). The estimated log probability Pr(X|K) for each cluster was compared using the

method described in Pritchard *et al.* (2001) with an asymptote for Pr(X|Y) at K=2 (-5075.7) and a peak at the highest likelihood of -5026.6 at K=4. The methods of Evanno *et al.* (2005) produced a modal K at K=2 but also produced a much smaller node peak at K=4. K=2 was used for admixture analysis.

The methods used in the analysis by BAPS 2, BAPS4, PARTITION and GENELAND differ in that they directly estimate the number of populations (K) up to a user defined maximum.

PARTITION provides the posterior probability (pp) of each K and computes the Bayes factor as a measure of evidence for (>1) or against (<1) a single panmictic population. For 100,000 iterations the most probable value of K with a posterior probability of 0.382 was 2, which was supported by a Bayes factor <1 (0.064). However, lower iterations produced K = 1. PARTITION failed to make a dendrogram for 100,000 iterations so a dendrogram from 30,000 iterations was used cautiously instead to create individual assignments. BAPS 2 produced a list of the best partitions and their associated posterior probability values, the highest of which (P>0.95) clustered the otter data first into four populations, and then into two populations when individuals UWCRef 433 and 441 were removed.

BAPS4 provided a list of the ten best partitions with an estimate of the correct number of clusters and their associated posterior probability estimates. It provides the option of utilising spatial information in analyses. BAPS4 SPATIAL provided the highest probability for K = 3, while the absence of spatial data produced an estimate of K = 7. Two of the top ten runs also had K = 8 as the best partition. Analysis without individuals 433 and 441 still produced K = 7.

The GENELAND posterior distribution including spatial data gave a mode of K = 4, and without spatial data gave a mode of K = 5. The removal of individuals 433 and 441 gave population estimates of K = 2 for the spatial and K = 3 for the non spatial analysis.

Appendix 2.1 shows the average assignment of individuals to each population for each of the programs.

2.4.5. Spatial patterns

The spatial coordinate data were joined to the posterior probabilities of assignment from each of the models for each individual. The interpolation maps of posterior probabilities produced in GIS provide a visual depiction of the cluster assignment. The maps (Figure 2.2.) show similar patterns of cluster distribution, with a distinct cluster in the South West of Wales (Cluster 1) and most of the remainder in the North and East of the country (Cluster 2). In STRUCTURE, the latter cluster appears to be fragmented, with an area in the Northwest of Wales and another adjoining the English border. A similar division is apparent in the clustering shown by BAPS4 NON-SPATIAL. BAPS4 NON-SPATIAL finds the greatest number of clusters, and as a result has the least spatially defined population structure, and differs markedly from all other outputs. The identification of individuals 433 and 441 as separate groups depended on the model used rather than the use of spatial or non-spatial models and these populations along with additional populations in GENELAND NON-SPATIAL and BAPS4 NON-SPATIAL are identified under Other (Clusters 3-7).

As PARTITION and BAPS 2 assigned individuals categorically, with no admixture, the differences between their maps were a result of the assignment of individuals to different clusters. BAPS 2 displayed a remarkably similar distribution of posterior probability of assignment to STRUCTURE which allows admixture (assignment of individuals to more than one cluster). GENELAND NON-SPATIAL analysis followed a similar pattern of distribution however it also identified a third isolated cluster in the Cardigan Bay region in West Wales. This cluster was also found to some degree by BAPS4 NON-SPATIAL (clusters 3-7). Further inspection of maps created by other methods show that this area is not assigned particularly strongly to any cluster, with individuals in this area showing a large amount of admixture.

GENELAND SPATIAL identified four clusters and produced the most distinct geographical separation between a South West cluster and a North and East cluster. This was a result of more individuals having greater assignment to a particular cluster. Nonetheless it still identified an individual in North Wales with strong assignment to the South West cluster (Figure 2.2.), the additional two clusters are a result of individuals 433 and 441 being assigned to two separate groups with 50% of each of their genomes assigned to each. BAPS4 SPATIAL analysis identified three

clusters; one containing only individuals 433 and 441. The two remaining resembled those defined by GENELAND SPATIAL analysis but the boundary between them was much less pronounced with some individuals showing greater levels of admixture.

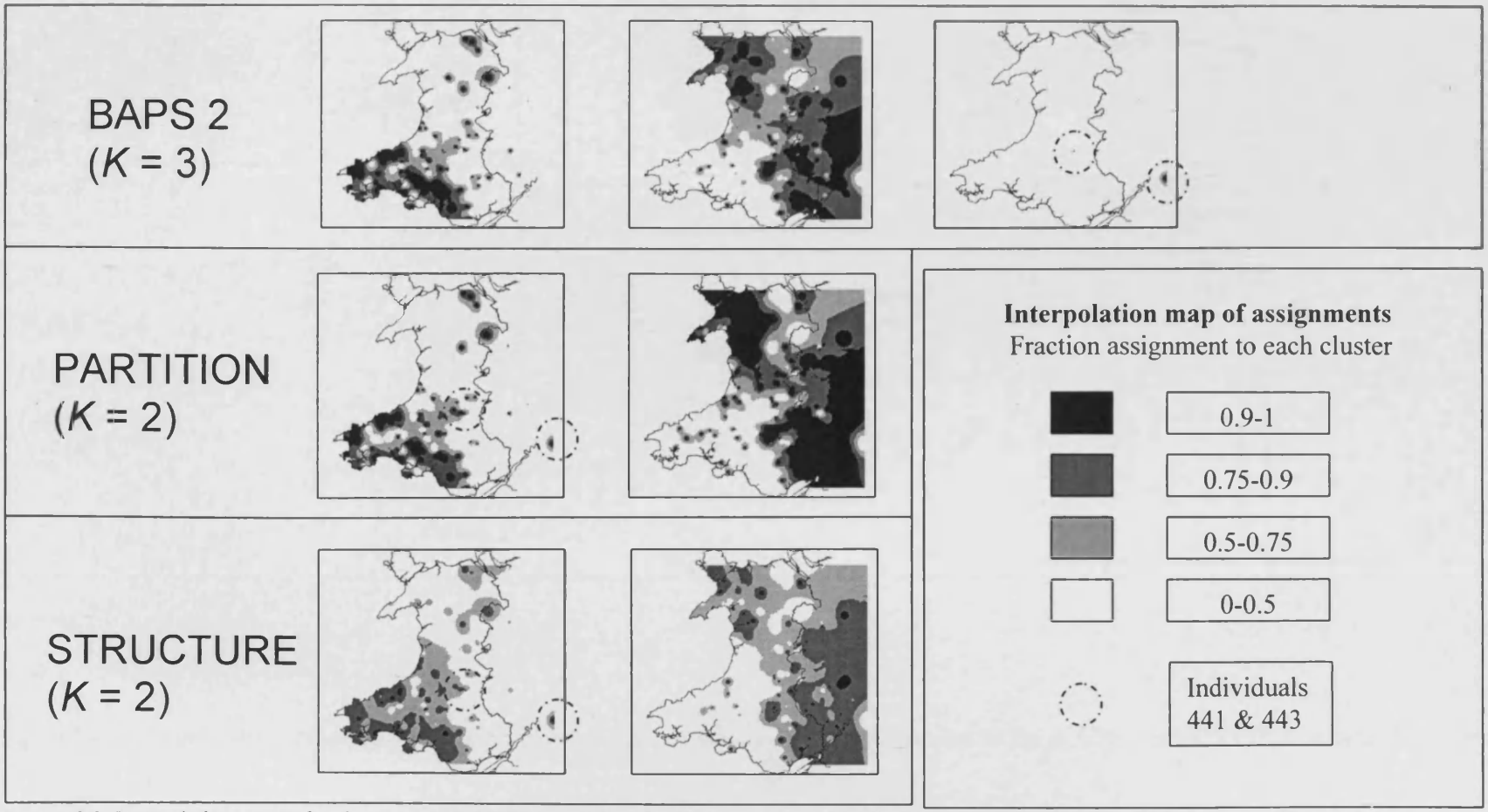


Figure 2.2. Interpolation maps of assignment probabilities, produced for all clusters identified for the estimated optimal K values for each of the models.

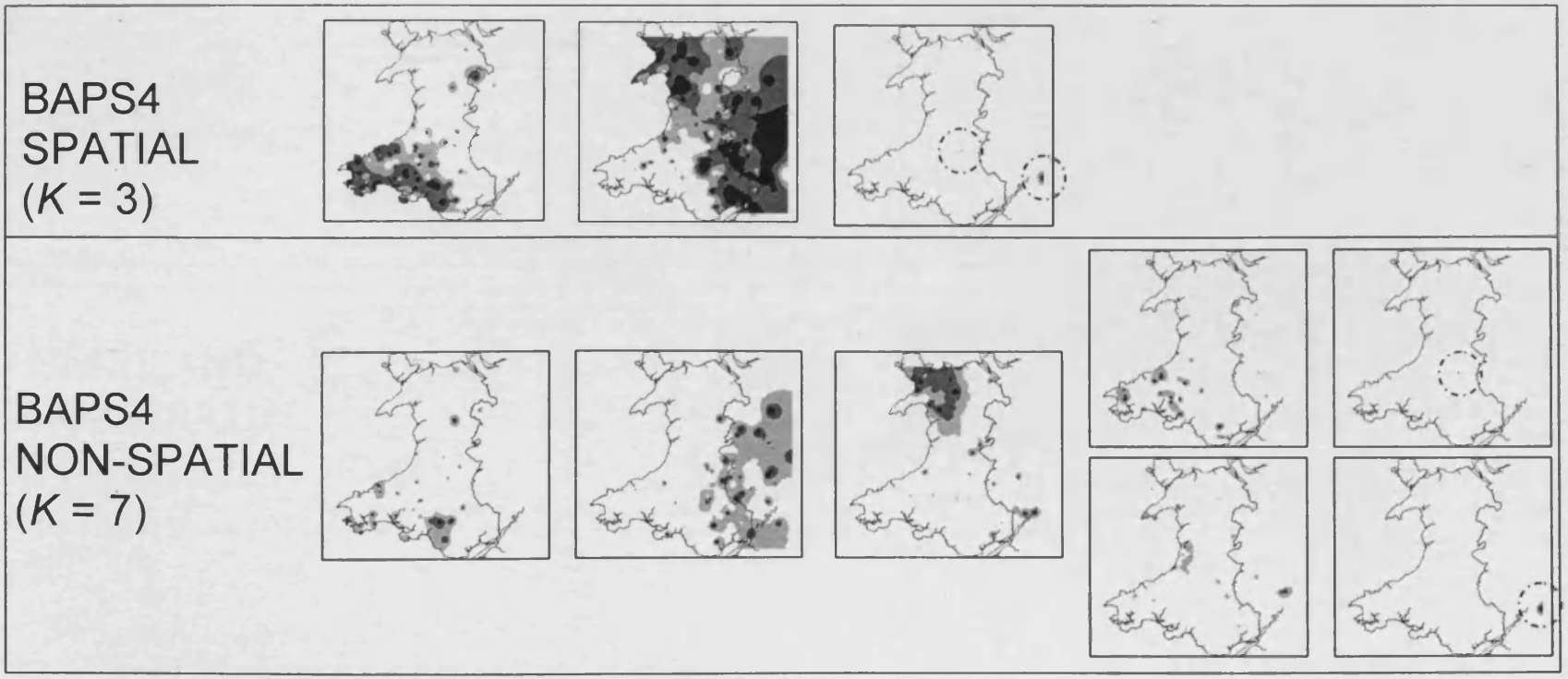


Figure 2.2 (continued). Interpolation maps of assignment probabilities, produced for all clusters identified for the estimated optimal K values for each of the models.

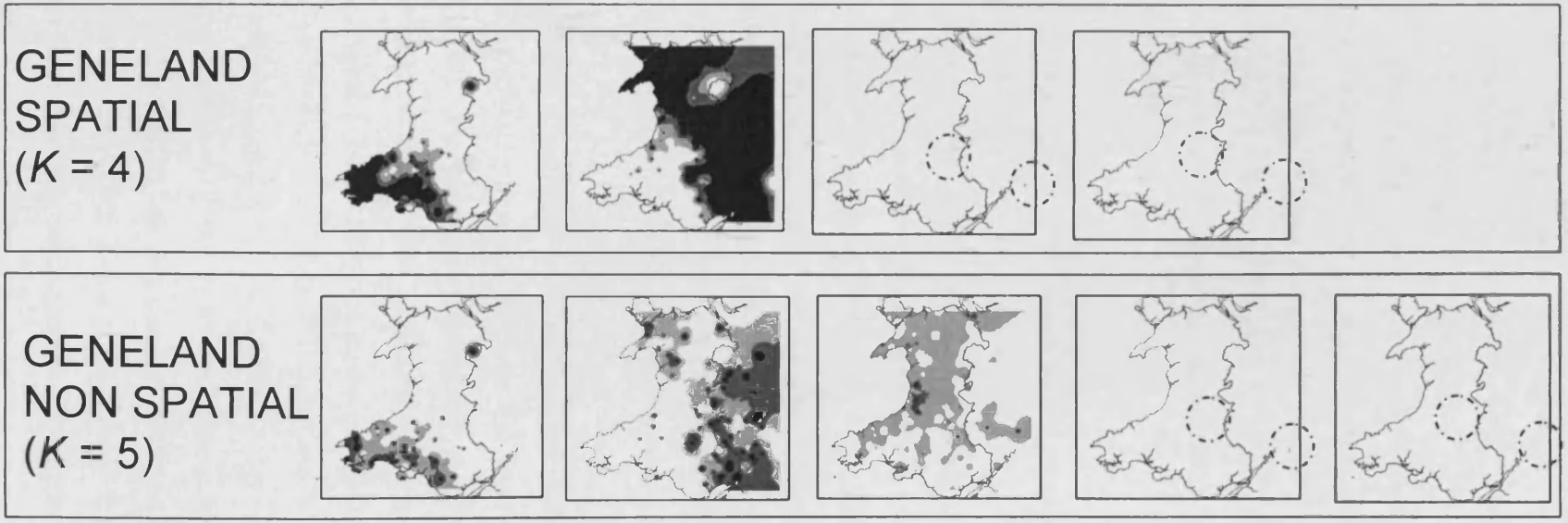


Figure 2.2 (continued). Interpolation maps of assignment probabilities, produced for all clusters identified for the estimated optimal K values for each of the models.

2.4.6. Comparison of assignments among programs

In comparing assignments using each of the programs, two arbitrary thresholds were used. First, an individual was deemed to belong to a cluster if it had > 0.75 of its genome assigned to that cluster. Since PARTITION and BAPS 2 do not account for admixture, an assignment threshold of 0.5 was also used to allow these programs to be compared (Table 2.3). There were two main clusters identified and comparisons of the assignments by the models will concentrate on these two identified clusters; cluster 1 (Southwest Wales) and cluster 2 (North and East of Wales). When the threshold of assignment was 0.75, 38 of 216 individuals (17.6%) were assigned to the same cluster by all seven methods used, with 62 (28.7%) individuals assigned at the 0.5 threshold.

The programs used the same theory to identify population groupings, but due to differences in their methods, assignments of some individuals could differ between programs especially when more programs are used. This is compounded by the fact that some models identify different *K*-values (BAPS4 NON-SPATIAL, GENELAND NON-SPATIAL). Taking this into account, the models were further compared by looking for agreement of individual assignment between any six of the seven methods used. A further 88 individuals (total 126; 58.3%)) were co-assigned by 6/7 methods at the 0.75 threshold (Table 2.4) and an additional 87 individuals (total 149; 70%) were co-assigned by 6/7 methods at the 0.5 threshold (Table 2.5).

Further comparisons were made of the models that agreed on $K = 2$ following the exclusion of individuals 441 and 433 (BAPS4 NON-SPATIAL, and GENELAND NON-SPATIAL were therefore excluded from this comparison) (Table 2.3.B). The similarity of assignments by these models was greater. With 135 (62.5%) and 163 (75.5%) of individuals being assigned to the same cluster at the 0.75 and 0.5 thresholds respectively, increasing to 171 (79%) and 202 (93.5%) individuals when 4 out of the 5 models agreed. This leaves only 11 individuals who have split assignment by 2 or more models to different clusters, these individuals have high levels of admixture by all of the models and as a result do not appear on the 0.75 threshold. It must be noted however, that a significant proportion of individuals would be expected to fall into the same cluster between programs by chance especially at $K = 2$.

Table 2.3. Agreement between models in their assignment of individuals to each cluster at the 0.75 and 0.5 thresholds. For A) all models compared, B) Models with a prediction of K=2 without individuals 433 and 441 were compared.

	Number of individuals			
	A) All 7 models compared		B) Models with prediction of *K* = 2, 5 models compared	
	0.75	0.5	0.75	0.5
All models assign to cluster 1	13	26	48	57
All models assign to cluster 2	25	36	87	106
All but 1 model assign to cluster 1	31	29	13	18
All but 1 model assign to cluster 2	57	58	23	21
Assigned at least twice to two clusters	0	12	0	11

A simple coefficient (Equation 2.1.) was used to index the similarity of assignment of individuals between the programs:

Equation 2.1.

$$\text{Similarity} = \frac{2 \times \text{Number of individuals assigned to pop 1 by both softwares A and B}}{\text{(Total Number of individuals assigned to pop 1 by software A + Total Number of individuals assigned to pop 1 by software B)}}$$

Table 2.4 shows the cumulative total of assignments of individuals to each cluster for each model, at the 0.75 threshold. Those with <0.75 assignment possibly share ancestry between the two clusters, and where $K > 2$ some individuals may be assigned to additional clusters. Table 2.5 shows the similarity coefficient (Equation 2.1.) for assignment of individuals to cluster 1 and cluster 2 at the 0.75 threshold.

When analysis was restricted to those programs that estimated K=2 (Table 2.2); at the 0.75 threshold the mean number of individuals assigned to cluster 1 was 71.8 $\pm$ 10.69, with 121.2 $\pm$ 20.41 assigned to cluster 2.

BAPS 2 and PARTITION were the most similar in their assignment of individuals to these clusters, both in number of individuals assigned and in the agreement of individuals assigned. These methods however do not take into account admixture and

as a result all individuals were assigned to one of the populations and were not discounted due to admixture. Where models identified K = 2 the similarity coefficient between models ranged between 0.75 and 0.93 for cluster 1 and between 0.63 and 0.97 for cluster 2 (Table 2.5). STRUCTURE and GENELAND SPATIAL had the lowest agreement in assignment between all of these programs. Due to the varying K values BAPS4 NON-SPATIAL had low similarity coefficients for the assignment to clusters 1 and 2, although GENELAND NON-SPATIAL had relatively high agreement considering it also assigned individuals to a third population.

For each model, the cumulative total assignment to each cluster at the 0.5 threshold is shown in Table 2.6, and the similarity coefficient between each program for each cluster at the 0.5 threshold are shown in Table 2.7. When the threshold was lowered to 0.5 there was a small increase (average = 0.04) in the similarity in individual assignments between the models that estimate $K = 2$.

Table 2.4. Total number assigned for clusters 1 and 2 at >0.75 assignment threshold

	Total number assigned to each cluster >0.75 assignment		
	Cluster 1	**Cluster 2**	**Total**
PARTITION	79	137	216
STRUCTURE	70	89	159
BAPS 2	86	128	214
BAPS4 SPATIAL	60	114	174
BAPS4 NON-SPATIAL	28	32	60
GENELAND SPATIAL	64	138	202
GENELAND NON-SPATIAL	43	83	126

Table 2.5. Similarity Coefficient for assignment to cluster 1 (>0.75), below diagonal and to cluster 2 (>0.75) above diagonal.

		Coefficient for assignment to cluster 2 (>0. 75)						
		PARTITION	STRUCTURE	BAPS 2	BAPS4 SPATIAL	BAPS4 NON-SPATIAL	GENELAND SPATIAL	GENELAND NON-SPATIAL
Coefficient for assignment to cluster 1 (>0. 75)	PARTITION	X	0.79	0.97	0.91	0.38	0.92	0.75
	STRUCTURE	***0.90***	X	0.67	0.71	0.31	0.63	0.75
	BAPS 2	***0.93***	***0.83***	X	0.88	0.36	0.87	0.75
	BAPS4 SPATIAL	***0.85***	***0.88***	***0.82***	X	0.37	0.81	0.75
	BAPS4 NON-SPATIAL	***0.52***	***0.57***	***0.49***	***0.57***	X	0.23	0.23
	GENELAND SPATIAL	***0.78***	***0.75***	***0.77***	***0.81***	***0.43***	X	0.75
	GENELAND NON-SPATIAL	***0.70***	***0.71***	***0.67***	***0.82***	***0.42***	***0.71***	X

Table 2.6. Total number assigned for clusters 1 and 2 at >0.5 assignment threshold

	Total number assigned to each cluster >0.5 assignment		
	Cluster 1	Cluster 2	Total
PARTITION	79	137	216
STRUCTURE	103	113	216
BAPS 2	86	128	214
BAPS4 SPATIAL	75	138	213 *
BAPS4 NON-SPATIAL	43	45	88
GENELAND SPATIAL	68	146	214
GENELAND NON-SPATIAL	59	90	149

* one individual UWCref702 is assigned 0.46 cluster 1, 0.49 cluster 2, 0.05 cluster 3.

Table 2.7. Similarity Coefficient for assignment to cluster 1 (>0.5), below diagonal and to cluster 2 (>0.5) above diagonal.

		Coefficient for assignment to cluster 2 (>0.5)						
		PARTITION	STRUCTURE	BAPS 2	BAPS4 SPATIAL	BAPS4 NON-SPATIAL	GENELAND SPATIAL	GENELAND NON-SPATIAL
Coefficient for assignment to cluster 1 (>0.5)	PARTITION	X	0.90	0.97	0.97	0.48	0.90	0.79
	STRUCTURE	***0.87***	X	0.82	0.81	0.46	0.77	0.79
	BAPS 2	***0.93***	***0.87***	X	0.92	0.46	0.86	0.78
	BAPS4 SPATIAL	***0.92***	***0.83***	***0.91***	X	0.48	0.93	0.79
	BAPS4 NON-SPATIAL	***0.62***	***0.55***	***0.62***	***0.64***	X	0.31	0.33
	GENELAND SPATIAL	***0.80***	***0.75***	***0.81***	***0.85***	***0.52***	X	0.78
	GENELAND NON-SPATIAL	***0.81***	***0.72***	***0.81***	***0.87***	***0.61***	***0.77***	X

2.4.7. Detecting clinal variation

Clinal variation can be detected by using the average population membership coefficients from multiple runs created in CLUMPP. By plotting the clusters in order of their membership coefficient to each cluster (Figure 2.3.) it can be seen that BAPS 2 and PARTITION (which do not allow admixture) seem to propose two discrete populations, however, where admixture is allowed there appears to be a cline in the assignments especially with STRUCTURE. When spatial data are included as a prior the cline is still present but shows a steeper slope (BAPS4 SPATIAL and GENELAND SPATIAL), with GENELAND SPATIAL having the strongest spatial prior and the steepest clinal gradient. Where $K > 2$ the individuals show a lot of admixture to the clusters, BAPS4 NON-SPATIAL and GENELAND NON-SPATIAL.

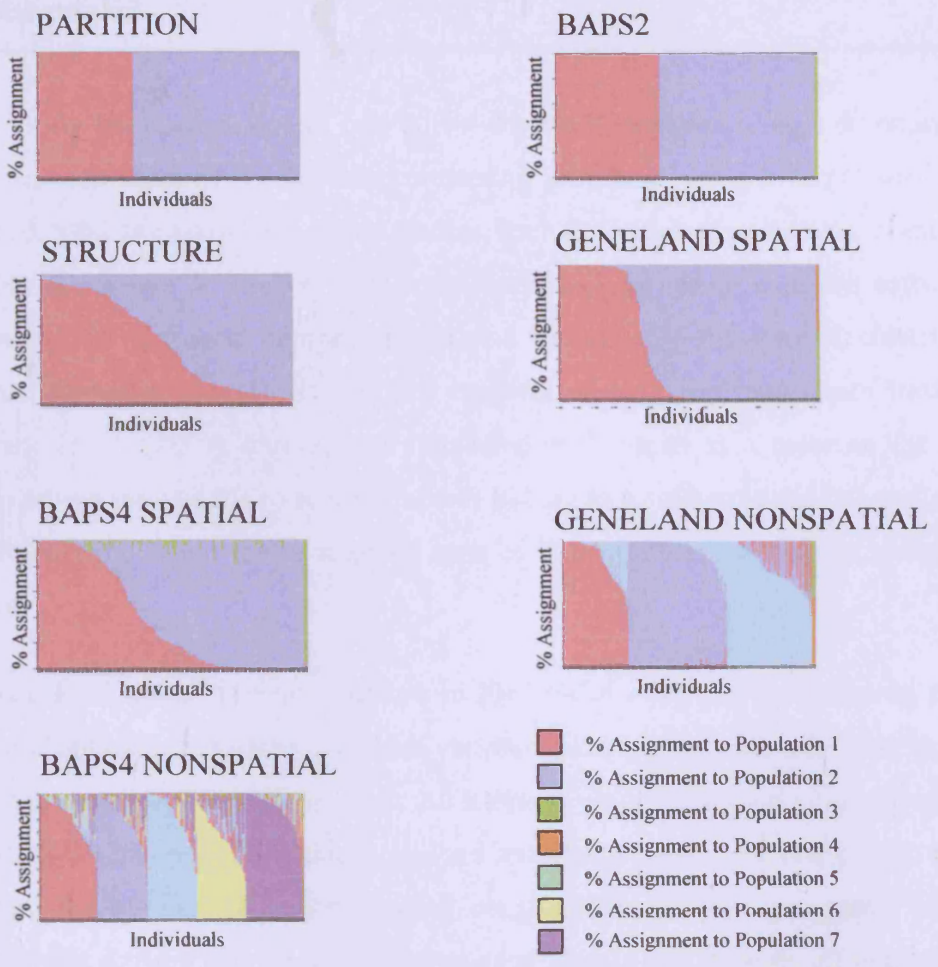


Figure 2.3. Detection of clinal variation. Individuals are plotted in order of their average membership coefficients to each cluster (modified in EXCEL). Each vertical line represents an individual, and the colour composition displays the probability of belonging to each cluster defined by the respective models.

2.5 Discussion

This is, to my knowledge, one of only a few studies to compare a large diversity of approaches implemented by Bayesian clustering programs, using a large, spatially referenced, wild animal dataset. Other studies, such as Coulon *et al.* (2006), Frantz *et al.* (2006) and Rowe & Beebee (2007) which did not set out to compare software have looked for the most straightforward and biologically meaningful clustering solutions, disregarding those that do not conform without analysing them further. Rosenburg *et al.* (2005) investigated 'clusteredness' which is a measure for the extent to which individuals were estimated to belong to a single cluster rather than a combination of clusters; they compared runs of STRUCTURE but did not compare programs.

Examination of spatial genetic structure in the Welsh otter population using four Bayesian clustering algorithms produced variable results, but in general there was a strong consistency between approaches. All methods produced results which strongly suggest that Welsh otters comprise more than one cluster, hence population. On first inspection, the number of clusters varied considerably between programs, but a closer analysis showed that the *K* value in some programs was strongly influenced by just two individuals (UWCref 433 and 441). It is interesting to note that STRUCTURE and PARTITION failed to identify these individuals as deriving from a separate group. Separation of these individuals by the other algorithms could be a result of unique alleles present in these individuals and to unique single-locus genotypes. Over the 15 loci, individual 433 had 2 unique alleles, and had unique single locus genotypes at seven loci, sharing genotypes at each locus with a mean of 14 individuals. Individual 441 had three unique alleles, and seven unique single-locus genotypes sharing genotypes with a mean of 22 individuals. Across the sample set, individuals were found to share a locus genotype with a mean of 61 individuals. Of eight unique alleles within the dataset, these individuals accounted for five. Of 46 unique single-locus genotypes they accounted for twelve. STRUCTURE and PARTITION may not be able to identify these unique individuals when at low numbers during individual level clustering. Further study (Chapter 3) suggests aassignment of these individuals to other UK populations, indicating that BAPS and GENELAND are able to identify migrant individuals from un-sourced populations. A study by Lecis *et al.* (2008) may

be a good example of this phenomenon, examining the American mink population (*Mustela vison*) in Spain they found that a poorly sampled (n=5) geographical area was assigned to its own cluster by GENELAND as expected from the population history, whereas STRUCTURE grouped these samples with another cluster, therefore having direct implications on the management of this invasive species. This highlights the importance of using multiple models for analysis of these types of data.

When the analysis was re-run without individuals 441 and 433, the estimation of K agreed between five of seven methods at $K = 2$ (Table 2.1), inferring that there are two otter population foci in Wales. On analysis of the maps of posterior probability (Figure 2.2.) these foci are located in South West Wales and in Mid and North Wales. This supports a hypothesis that population structure has arisen following a sudden decline in otter numbers in the late 1950's and early 1960's (Coxon *et al.* 1999) as a result of the use of substances such as organochlorines (Conroy & Chanin 2000; Mason & Macdonald 2004), leaving suspected population refugia in Mid and West Wales (Jones & Jones 2004). Comparisons between mapped population structure (Figure 2.2.) and landscape features (not shown) reveal no obvious barriers to dispersal correlating with the two groups. Sub-structuring as a result of differentiation at a lower geographical scale is confirmed by having seven loci deviating significantly from HWE.

The Wales and Borders otter population appears to have low genetic diversity with an average expected heterozygosity (*He*) of 0.53 and an observed heterozygosity (*Ho*) of 0.49 over the 15 loci, which was somewhat lower than the European average *He* = 0.74, *Ho* = 0.55 (Randi *et al.* 2003), and also lower than the island population of Kinmen (China) *He* = 0.61, 0.70 (Hung *et al.* 2004, Huang *et al.* 2005). With the increase in the otter population as a whole in Wales these populations appear to be no longer in isolation with strong evidence for admixture found in some individuals.

Comparison of individual population assignments in Table 2.3 shows that using a 0.75 assignment threshold, population assignment only agrees among all the models for 18% of the 216 individuals. This is important to take into consideration when using Bayesian clustering methods; further analysis using additional models will

increase the probability of the correct assignment (Yeung & Ruzzo 2001; Latch *et al.* 2006). It seems reasonable to assume that the more models which agree in a given individual's assignment, the more probable that assignment is likely to be. However, it also follows that the more programs used, the less likely it is that all individuals will be assigned to the same population by all programs. This can be a result of varying estimations of *K,* but even when models agree on the *K* value assignment can vary between models for certain individuals. For example, in table 2.3, five models with the same estimation of *K* agree on the assignment of 62.5% of individuals, however allowing the relaxation of the parameters so any 4 of the 5 models match, individual assignment agreement increased to 79%. To facilitate comparison of individual assignment between methods, a simple binary similarity coefficient has been used but where more than $K = 2$ populations are inferred additional covariance analysis is advisable. Table 2.5 & 2.7 show the similarity coefficients of the programs for their assignments to cluster 1 and 2, respectively, at the 0.75 and 0.5 thresholds. BAPS4 NON-SPATIAL shows the least correlation to the other methods for both clusters, however this is explained by its estimation of *K* which was different to the other programs. GENELAND NON-SPATIAL also had differing number of clusters and as a result has lower similarity coefficients for individual assignments when compared to the other programs.

Based on their similarity coefficients STRUCTURE, PARTITION, BAPS 2 and BAPS4 SPATIAL assigned a majority of individuals to the same clusters, 82-97% of individuals with 0.75 assignment threshold (Table 2.6) and 83-97% of individuals with a 0.5 assignment threshold (Table 2.7). When these programs are compared to GENELAND SPATIAL the average similarity coefficients were slightly lower ranging from 75-97% at >0.75 threshold and 75-97% at 0.5 threshold. GENELAND SPATIAL assigned more individuals above the 0.75 threshold than any other program that allows admixture, only 12 individuals failed to have assignment greater than the 0.75 threshold (Table 2.4), this perhaps reflects the strong effect of the spatial *prior,* as a result this program must be considered a powerful addition to the landscape genetics toolkit, although care must be taken in its interpretation given the strong spatial dependence.

The clustering patterns resulting from the interpolation maps (Figure 2.2.) were remarkably similar between the methods with the exception of BAPS4 NON-SPATIAL model. The 'core' 38 individuals assigned to the same clusters by all of the programs were usually assigned with the least admixture (see appendix 2.1). Differences exist where individuals were classified differently between programs (on analysis of the posterior probability these were usually assigned the most admixture). However, this was not the rule and some individuals had a large assignment probability to one cluster by one program which was not reflected by all programs. To achieve a higher percentage of correct assignments by all of the models, high levels of genetic differentiation are needed (Latch *et al.* 2006), the variability in the assignments of individuals reflects the low levels of differentiation between the groups. Given the similarity but most importantly the differences between the results of the Bayesian clustering methods it is important to use a variety of these methods to completely understand the population structure especially when genetic variation and/or differentiation is low.

The average population membership coefficients of individuals for programs that allow admixture indicate that there may be evidence of a cline in allele frequencies between the two populations (figure 2.3.). This adds strength to the hypothesis that the otter population was once a panmictic population that became fragmented for reasons mentioned previously. Changes in farming practices, have contributed to the recovery of the two populations and possible gene flow between them. Alternatively there is also the possibility that there were historically two populations, separated by a semi permeable barrier with limited gene-flow between them. There are no obvious landscape features that could be acting as barriers to otter dispersal, considering that otters are mobile carnivores with the ability to disperse large distances, further analysis and investigation of the Wales and the UK otter population will be conducted in future chapters 3 and 4.

Several papers have reviewed Bayesian clustering software, such as Pearse & Crandall (2004), Beaumont & Rannala (2004), Manel *et al.* (2005), Excoffier & Heckel (2006) and Latch *et al.* (2006). Only Latch *et al.* (2006) and Chen *et al.* (2007) compare software performance. Using a simulated dataset, they acknowledge that the performance of the models in identifying population structure depends

mostly on the properties of the data, such as sample size, number of loci and variability of loci. This is not a new conclusion; Rand (1971) stated that to evaluate the performance of a clustering method its results must be compared with either standard results, or the results of another method. When using simulated datasets the former comparison can be undertaken, but when analysis is performed on unknown population datasets at least two methods should be used. The use of multiple approaches has been incorporated into phylogenetic and gene expression analysis where it is recommended that several different algorithms should be used, because agreement among analytical methods is more likely to detect the real as opposed to apparent signal in the data (Yeung & Ruzzo 2001; Datta 2003; Thalamuthu *et al.* 2006). To further emphasise the importance of using more than one approach, each of the following studies favoured the results of different programs: Coulon *et al.* (2006) favoured GENELAND SPATIAL, Frantz *et al.* (2006) favoured BAPS4 SPATIAL whereas Rowe & Beebee (2007) found different programs worked best when studying different discrete populations.

2.6 Recommendations

The optimal choice of program clearly depends on the dataset used and the objectives. Although all the programs showed a similar population structure when mapped, GENELAND SPATIAL D model provided the most resolution between populations, as would be expected if greater weight was given to assignment of individuals based on spatial data. GENELAND SPATIAL should be used where spatial discontinuity is likely to be very important (Chen *et al.* 2007), for example for the identification of migrants between adjacent and non-overlapping populations. The use of this model would therefore appear prudent and an ideal tool for identifying barriers to dispersal. GENELAND, however, requires considerable computational power especially when estimating *K*. However, BAPS4 uses much less computer time. If spatial coordinates are available BAPS4 would be appropriate for a preliminary study, however when the non-spatial option was used it did not perform well in comparison to the other methods with this data set appearing to identify spurious populations, however, it could be delineating partitions at a finer genetic resolution. If no spatial data are available, BAPS 2 may be a quicker option for reliable results; it also produces a dendrogram which may prove useful if looking for relationships among individuals.

PARTITION did not identify individuals 433 and 441 as belonging to a separate group and does not give admixture data, only producing a dendrogram, interpretation of which is time consuming and may lead to human error. The map of posterior probabilities produced in ArcMap v 9.2 (ESRI 2007) using data from PARTITION identified less spatially defined populations than other programs. It is also very computationally intensive. It did however, generate clusters that were consistent with other models, assigning 75-97% of individuals to the same cluster. STRUCTURE, despite being the most widely used method, also did not identify the two very distinct individuals in this dataset, and the reliability of its estimation of *K* has been a matter of debate (e.g. Evanno *et al.* 2005). It is very computationally intensive, and the need to perform replicate runs for each *K* value exacerbates this, but distributed processing offers a solution to this problem, and it remains a powerful method to identify populations.

When using Bayesian clustering approaches it should be acknowledged that each method is likely to give a slightly different result unless the patterns of genetic structure are extremely strong. This is not always the case, and here mean F_{ST} between clusters 1 and 2 displayed in Figure 2.2 ranged between 0.06 and 0.08 (data not shown), although all values were significant ($P < 0.05$; data not shown). Anomalous assignments and disagreement within and between spatial and non-spatial methods are likely. It is therefore clear that more than one approach should be used and that agreement between a combination of spatially explicit and non-spatial programs is powerful evidence of population assignment.

Studies on simulated datasets (Latch *et al.* 2006; Chen *et al.* 2007) show that at low levels of genetic differentiation Bayesian clustering Algorithms could identify the correct number of populations at low levels of genetic differentiation, however they performed poorly in the assignment of individuals to populations at these levels. It is therefore even more important when identifying conservation and management strategies to identify and assign individuals to the correct populations and this depends on our ability to correctly delineate genetically distinct populations and to identify landscape corridors or barriers to gene flow between them.

The identification of clines is also important, however given the fact that at low levels of genetic differentiation these models performed so poorly in their assignments on simulated datasets the resulting evidence for clines may be an artefact of the model. For example STRUCTURE showed strong evidence of a cline between the two populations, however, the models that use spatial priors find less admixed individuals. This could mean that there is too much emphasis put on the spatial prior or that STRUCTURE finds too much admixture in its assignment of individuals when there is only a small amount of genetic differentiation between populations. BAPS4 Spatial places less emphasis on the spatial prior than GENELAND SPATIAL and may be a compromise between these two models.

Therefore a more reliable way must be established that increases the robustness of the clustering solution. At a minimum several clustering methods should be used and compared to estimate the number of populations. In the case of otters in Wales, two

possible clusters were identified with agreement between a majority of models. Individual assignment showed some variance between models, and it is difficult to say which models results should take priority. Many studies use the most biologically plausible answer to explain their results that best fits the known population history. However, to increase the validity of the assignments, a method should be devised that allows the combination of the assignments of multiple models.

This study into Bayesian Clustering Algorithms is a precursor to a fuller study into landscape genetic analysis and a further study with ecological data will be carried out to identify less obvious barriers to dispersal investigating landscape connectivity correlating landscape features with genetic diversity or whether these two groups arose as a result of a panmictic population becoming fragmented due to anthropogenic causes.

Chapter 3

Bayesian Clustering Techniques and Progressive Partitioning to Identify Population Structuring within a Recovering Otter Population in the UK

3.1. Abstract

After a major decline, the UK otter population is now recovering from its known strongholds (northern England, Wales and Borders, southwest England) and from central England where it was once thought to be close to extinction, and as a result has additionally received many reintroductions. Bayesian clustering techniques and GIS are used here to identify the genetic structure of the UK otter population and to map the otter expansion from the known strongholds and identify the contribution of reintroduced otters.

Three Bayesian clustering techniques were used (STRUCTURE, GENELAND SPATIAL, BAPS4 SPATIAL) to estimate the number of populations (K). In addition a novel progressive partitioning approach was tested to identify sub-structuring at various hierarchical levels using a $K = 2$ approach.

Four regional populations (genetically distinct groupings) were identified that reflect known population history. Isolated populations in southwest England and Wales and its borders showed lower levels of genetic diversity. High levels of genetic diversity and unique alleles in the north and central England regions reflect the proximity to genetically diverse Scottish populations and the positive effect of reintroductions respectively.

The progressive partitioning approach provided a detailed clustering analysis, by allowing the comparison and combination of clusters identified by the different Bayesian clustering techniques and avoiding the subjective estimation and choice of K. This method gives a better understanding of the assignments to the final clusters and could be used as a method of identifying spurious clusters along a cline.

Whilst the otter population is increasing the results show little or no sign of population expansion from the stronghold regions into central England. The results reflect the success of reintroductions on population growth in the central England region and the identification of further sub-structuring (11 sub-regions) will provide a tool for management efforts in protecting genetically differentiated, geographically isolated populations.

3.2. Introduction

Wild animal populations are under increasing pressure from anthropogenic factors, leading to fragmentation and isolation. The reduction or absence of gene flow between populations can lead to losses in genetic diversity through effects such as genetic drift; as a result a small amount of gene flow is considered necessary for the viability of small isolated populations (Mills & Allendorf 1996).

Conservation management aims to preserve evolutionary processes and adaptive diversity across the geographic range of a species (Storfer 1996). To do this management plans should aim to preserve the natural network of genetic connections between populations, rather than just preserve isolated populations within that network (Crandall *et al.* 2000), thus ensuring that the processes that maintain adaptive diversity and evolutionary potential are conserved. It is therefore important to identify population units within a species range and the degree of gene flow between them. Defining populations and their geographic boundaries can however be difficult.

In population genetics, estimators of population structure traditionally rely on *a priori* definition (Pearse & Crandall 2004). In conservation biology, study species are often rare or declining, and the populations can become small and fragmented while still appearing continuous. For example, for widespread but elusive species like the otter, the population may appear to be continuous and it can therefore be difficult to delineate population boundaries or find substructure within populations with certainty. In these cases grouping individuals into pre-defined populations may reduce the ability to accurately describe the true population structure (Pearse & Crandall 2004). To combat this problem, recent methods instead look to cluster individuals into groups defined on genetic criteria.

3.2.1. Bayesian clustering algorithms

Bayesian clustering algorithms use individual multilocus genotypes derived from multiple microsatellite markers to assign individuals to clusters, on the assumption that markers are in Hardy Weinberg and linkage equilibrium within each randomly

mating subpopulation (Pearse & Crandall 2004; Manel *et al.* 2005; Latch *et al.* 2006).

Recent reviews of Bayesian clustering approaches (e.g. Pearse & Crandall 2004; Manel *et al.* 2005; Latch *et al.* 2006) describe and compare such techniques, and note that these methods are relatively untested, and that comparative analyses are lacking, particularly for 'real' datasets. The choice of software used in the literature is at the discretion of the researcher and when multiple programs are used the results of the software that best fits the known biology of the species tends to be chosen. There is also some debate about the utility of these models to identify the true number of populations (K). For the software STRUCTURE there are two methods to identify K, the author's own (Pritchard *et al.* 2000) and one which uses the second order rate of change of the likelihood function with respect to K (Evanno *et al.* 2005). Many papers use just one clustering technique and consider this to reflect the true population structure, but because the interpretation of clustering can be subjective (Dubes & Jain 1976; Jain *et al.* 1999) a single clustering algorithm or approach is not adequate to solve every clustering problem (Jain *et al.* 1999). It is advised in the literature to interpret clustering techniques with caution as they are tools of exploration rather than ends in themselves, and it is preferable that several clustering programs are used and compared (Dubes & Jain 1976).

The general clustering literature underpins the recommendation in chapter 2, that to account for varying K values the outputs of several clustering algorithms should be combined to increase the robustness of the solution (Topchy *et al.* 2003). Carmichael *et al.* (2007) follow this recommendation, using STRUCTURE and GENELAND SPATIAL to analyse wolf population structure in North America. They identified 4 population groupings that agreed between models, but also identified populations that differed between models, despite using the same optimum K value of 7. In these cases where some agreement is not found, chosen populations were justified by the authors using geographical location and other defining features, in this case eventually settling on 10 clusters. This study demonstrates the difficulty of combining and comparing the results from different methods when they identify different population groupings. The spurious identification of populations by Bayesian clustering techniques has been a cause of concern (Frantz *et al.* 2009) and

could bias results when faced with deviations from random mating not caused by genetic discontinuities, for example along an isolation by distance (IBD) gradient 'cline'.

An alternative approach is to use progressive partitioning in combination with Bayesian clustering. This method has been used in other areas of cluster analysis and resembles a hierarchical approach used by Coulon *et al.* (2008). It operates on the assumption that hierarchical clustering algorithms produce a nested series of partitions based on a criterion for generating clusters based on similarity (Jain *et al.* 1999). Progressive partitioning of the dataset restricts K to 2, extrapolating 2 clusters at each sub-division of populations. Clusters identified in the first round of analysis should be the most differentiated, with clusters derived from these having progressively lower levels of genetic differentiation between them. The progressive partitioning approach may allow for the identification of apparently spurious populations when looking for clusters at lower levels of genetic differentiation, although it has the disadvantage of explicitly ignoring the likelihood values for the different values of K.

Similar approaches have been carried out by Coulon *et al.* (2008) that used a hierarchical approach and by Carmichael *et al.* (2007) that combined the results of two Bayesian clustering techniques. Combining the assignments produced by the different Bayesian techniques can be problematic as they can estimate different optimum K values for the same dataset, therefore producing different populations which cannot be compared. Carmichael *et al.* (2007) tried to control for this by using only the estimate of K from one technique however, even when the same K value was used they found that the partitions identified could differ between models.

By using the progressive partitioning method all partitions produced by the Bayesian techniques can be displayed using GIS and compared to identify consistent patterns of sub-structuring at different degrees of differentiation. Clusters that find agreement between techniques can be used to build a picture of population structure and combined to test further genetic summary statistics.

3.2.2. The Eurasian otter – population history in the UK

The Eurasian otter in the UK declined significantly during the late 1950's and early 1960's (Coxon *et al.* 1999; Conroy & Chanin 2000; Mason & Macdonald 2004). By the mid 1970's the UK population was largely confined to strongholds in parts of Scotland, northern Ireland, mid and west Wales and southwest England (Jones & Jones 2004) with a small remnant population in East Anglia (Jessop & Cheyne 1992). The most likely cause, given the suddenness of the decline, was the introduction of the organochlorine group of insecticides (particularly dieldrin), and polychlorinated biphenyls (PCBs) (Conroy & Chanin 2000; Mason & Macdonald 2004).

Detailed monitoring programmes have shown that since the late 1970's there has been a slow expansion of the otter population in the UK (Conroy & Chanin 2000), which may be the result of reduced pollution. Population expansion and recolonisation is believed to be occurring in areas of central and southern England as a result of breeding and by dispersal, from the west (southwest England and the Welsh borders) and from the north (Scotland); (Coxon *et al.* 1999; Conroy & Chanin 2000). In addition to natural increases, a number of deliberate otter releases were made, to augment fragmented and declining wild populations, following recommendation by the Nature Conservancy Council after the first otter survey of England (Lenton *et al.* 1980).

Releases were carried out in central and southern England by The Otter Trust (OT), who bred otters and released 117 between 1983 and 1999. The OT used a founding stock made up of wild otters caught in live traps in East Anglia during the Ministry of Agriculture's Coypu Campaign, and also otters which were given to the trust (Wayre 1992) the origin of which are unknown. Releases were also made by the Vincent Wildlife Trust (VWT), who released groups of otters into North Yorkshire between 1990 and 1993 in the River Derwent catchment and the nearby River Esk (Strachen & Jefferies 1996). VWTs releases were of rescued otters rehabilitated but not bred in captivity, and included animals originating from north and east Scotland, Wales, Northern Ireland and southwest England (Rosie Green, Pers. Comm). Releases by both Trusts were deemed successful, with an increase in positive otter

sightings in release and surrounding locations within a short period of time (Strachen & Jefferies 1996; White *et al.* 2003).

The genetics of the UK otter population has previously been studied by Dallas *et al.* (2002) who found that there was no gene flow between the otter strongholds in Scotland, Wales and southwest England. Wales and the southwest of England were shown to have less microsatellite polymorphism than Scotland (Dallas *et al.* 2002). Little is known about either the population structure in southern and central England, or about the contributions made by dispersing individuals from otter strongholds or from introduced individuals.

Otters are difficult to study and observe in the wild; they are secretive, crepuscular and their protected status does not allow them to be disturbed. As a result monitoring is primarily indirect, using signs of presence such as footprints or spraint (a mixture of faeces and scent gland deposit). Such observations indicate otter presence, and spraint can be used in molecular studies to enable identification of individuals from DNA. Such analyses (eg. Chanin 2003; Dallas *et al.* 2003; Huang *et al.* 2005) provide a wealth of information on individuals and populations, but it is notoriously difficult, costly and time consuming to achieve reliable genotypes. DNA extracted from muscle tissue is much more reliable, in this study tissue samples from otters archived during a long-running post mortem study, from known locations in England and Wales were used. This provides a cost effective way of genotyping a representative sample of the UK otter population.

3:2.3. Expected populations based on population history

It was hypothesised that there would be five genetically distinct otter populations in the UK, and further sub-structuring found within these populations. It was expected that three genetically distinct population groupings would be found based around the otter strongholds, Wales and borders, southwest of England and Scottish borders, as reported by Dallas *et al.* (2002). Additionally, populations in East Anglia and North Yorkshire were expected to be found as these areas have received otter translocation and reinforcement.

3.2.4. Aims

The aim of this chapter is to identify the population structure of otters in the UK using Bayesian clustering. Further to this, the otter dataset is used to explore differences in the outputs from several established clustering models (STRUCTURE (Pritchard *et al.* 2000), GENELAND SPATIAL (Guillot *et al.* 2005a), and BAPS4 SPATIAL (Corander & Marttinen 2006); to compare these with the outputs from a novel approach using progressive partitioning, and with assumed populations based *a priori* on otter population history. Detection of 'true' population structure will allow the identification of the degree of gene flow between populations, the identification of recolonisation events from otter strongholds and the assessment of the success of otters reintroduced 20-30 years ago.

3.3. Methods

Individual multilocus genotypes were produced at 15 loci for 566 otter road casualties using methods described in Chapter 2. The majority of tissue samples were from Wales and England, with a smaller number from Scotland and Ireland.

Three approaches to define populations were used: (1) definitions based on software used to estimate *K* and assign individuals to populations, (2) a progressive partitioning method, which restricts *K* to 2 at each division, and (3) selection of populations based on known otter population history.

For (1) and (2) the multilocus genotypes for 566 individuals at 15 loci were analysed for population structure using the three programs STRUCTURE, BAPS4 SPATIAL, and GENELAND SPATIAL. In general, for all programs the author's guidelines were followed and default values taken where applicable. Specific parameters used are shown in Table 3.1.

3.3.1. Definition of populations based on Bayesian clustering techniques

The three Bayesian clustering techniques were used to estimate an optimum *K* and assignments were mapped for comparison using ArcMap v 9.2 (ESRI 2007) (shape files provided by the Environment Agency). Individuals were categorised within given populations if they had greater than 50% assignment to that population.

Table 3.1. The parameters used for the Bayesian clustering techniques in the current study.

Package	Algorithm	Iterations	Model	*K*	Estimation of K
STRUCTURE	Described by Pritchard *et al.* (2000); Falush *et al.* (2003); Pritchard & Wen (2003)	1,000,000 iterations, using CONDOR (CONDOR is a specialised workload management system for compute-intensive jobs). Burn-in 100,000	Admixture, model, assuming correlated allele frequencies	Set from 1 to 11, with 5 independent runs of each	Two approaches used; The highest estimated log probability of data Pr(X|K) estimates the most likely number of clusters (Pritchard *et al.* 2000). Evanno *et al.* (2005) which uses the second order rate of change of the likelihood function with respect to *K*.
BAPS4 Spatial	Described in Corander & Marttinen (2006). Corander *et al.* 2003, 2004)		Clustering at the individual level, using the spatial model Admixture analysis Minimum number of individuals in a cluster = 1. Default values were used for admixture priors. 100 iterations were used to estimate the admixture coefficients, 200 reference individuals from each population and 20 iterations to estimate the admixture coefficients for the reference individuals	Vector of values for the maximum number of clusters (*K*) with five replicates of *K* = 5, 10 and 15	After all the *K* values were processed, the stored results were merged based on the logML values with the best 10 partitions displayed The *K* value with the highest likelihood was chosen.
GENELAND SPATIAL	Guillot *et al.* (2005)	500,000 MCMC iterations to identify *K* *(using CONDOR)* 200,000 iterations once *K* was identified *(using CONDOR)* Thinning = 100	The Dirichlet (λ) distribution was used following Guillot *et al.* (2005) using spatial data (Spatial D-model) The amount of uncertainty to spatial coordinates was set at 0.3 (author recommendation), maximum rate of Poisson process 566 (number of individuals); maximum number of nuclei in the Poisson-Voronoi tessellation 1698 (3 x the number of individuals).	Priors on *K*-uniform between 1 and 11	The most probable number of clusters (*K*) was found using five replicates comparing the histograms

3.3.2. Definition of populations based on progressive partitioning with Bayesian clustering

Analyses were conducted with a fixed K of 2. At each progressive partition, the identified clusters were subjected to further analysis at $K = 2$ until the clusters no longer split (this occurs when either all individuals are assigned to one population, or when all individuals show ~50% assignment to each of the two populations). An individual was assigned to a cluster if it had greater than 0.5 assignment to that population (to allow all individuals to progress to the next stage of the analysis). At each stage five replicate runs (all of $K = 2$) were performed and compared. Typically runs were identical, but difficulties arose in cases where most individuals were clearly assigned to two distinct groups, but remaining individuals were given 50/50 assignment. In such cases all individuals with 0.5 assignment were arbitrarily assigned to the same population for further analysis. In some cases, this method resulted in three distinct populations being forced into two clusters. Fortunately in this study all Bayesian algorithms assigned the individuals deemed to be from the third cluster to the same cluster as each other (although the cluster they were assigned to could differ between runs), and separated them out on the following partition. At each stage the 5 runs were compared for consistency, one run representing agreement between the assignments of the majority of runs was used for further analysis discarding the inconsistent runs.

3.3.3. Definition of populations based on otter population history

Bayesian clustering techniques actively seek significantly different partitions in HWE. Therefore, traditional genetic summary statistics must be interpreted with caution because of this bias and for the effect of over representation of certain populations when comparing populations identified by Bayesian clustering methods. For comparison, the summary statistics from populations defined *a priori* based on otter population history were calculated. From the 566 otter samples, five subsamples of ~50 individuals were randomly chosen to represent five key areas for otter population and for use in deriving summary statistics. These included three 'stronghold' areas where populations remained following population declines: (1) Wales and Borders, (2) Southwest England, and (3) North England, and two areas where populations are known to have been reinforced: (4) East Anglia and central England and (5) North Yorkshire.

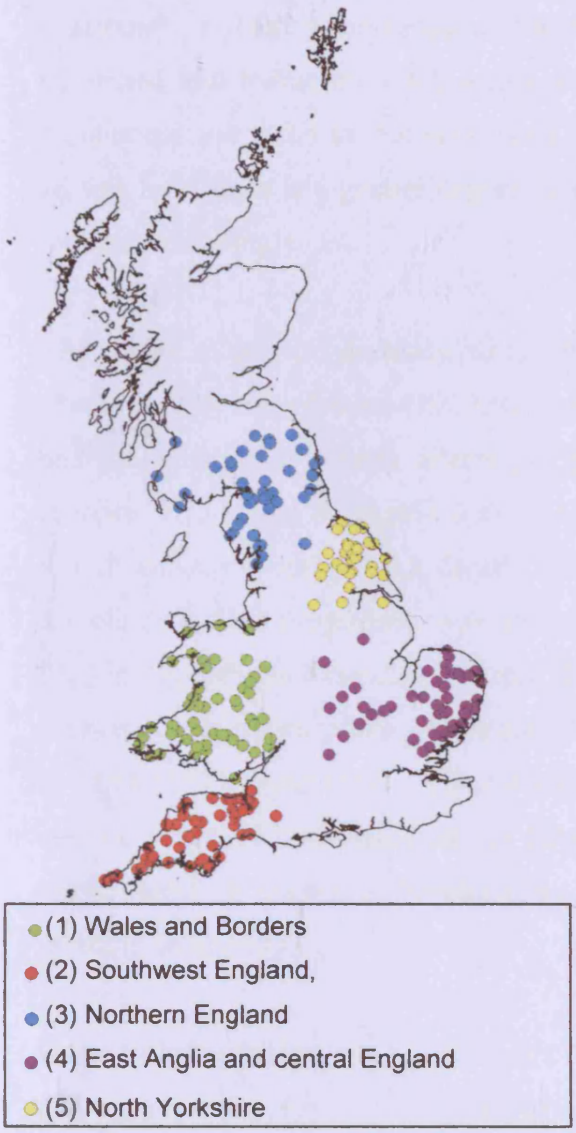


Figure 3.1. Locations of individual otters chosen *a priori* to make five subsamples of ~50 individuals randomly chosen from the total 566 otters used in Bayesian clustering analysis. Each subsample represents *a priori* defined otter populations based on otter population history (colour coded in the key above).

3.3.4. Genetic statistics

Genetic structure was quantified using standard summary population genetic statistics. These were applied separately to the *a priori* definitions and to the clusters identified by comparing and combining the outputs identified by the Bayesian clustering techniques (optimum *K* and progressive partitioning method). To combine

regional level analyses (regions described below) where individual assignment to a region was > 0.9, with agreement between all softwares. This value shows strong assignment, reflecting the strong partitioning at the regional level. Individuals were combined and included in sub-region analysis (sub-regions shown below) where assignment was > 0.5 to that sub-region with agreement by two or more softwares. At this level there is a greater degree of admixture so the assignment threshold was reduced accordingly.

The levels of genetic diversity within populations were estimated by calculating observed (Ho) and expected (He) heterozygosities, the average number of alleles (A) and the number of private alleles per population (Au) using the GDA software (version 1.1; Lewis & Zaykin 2001). Allelic richness (AR) was calculated using FSTAT software (version 2.9.3; Goudet 1995) adjusted for variation in subpopulation sample size. This programme was also used to estimate the inbreeding coefficient (F_{IS}) in populations separately and overall (F_{IS} is a statistic describing how well the genotype frequencies within populations fit with Hardy Weinberg expectation (Hartl & Clark 1997)). ARLEQUIN 3.1 was used to derive population pairwise comparisons and the statistical significance of F_{ST} values was tested with 10,000 permutations as implemented in ARLEQUIN 3.1 (Excoffier *et al.* 2005) and Bonferroni corrected for multiple comparisons.

Genotypic distribution for conformance with Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) was tested using GENEPOP 3.3 (Raymond & Rousset 1995) with all probability tests based on the Markov chain method (Guo & Thompson 1992) using 1,000 de-memorization steps, 100 batches and 1,000 iterations per batch.

3.3.5. Population history

The occurrence of any recent population bottlenecks (as predicted if the populations declined drastically during the period when pesticides were in heavy use) were inferred using BOTTLENECK 1.2 (Cornuet & Luikart 1996), assuming an infinite allele model (IAM), a step-wise mutation model (SMM), or a two-phase model of mutation (TPM, with 95% SMMs).

3.3.6. Immigration drift equilibrium

The software 2mod (Ciofi *et al.* 1999) can be used to estimate the relative likelihoods of a model of immigration-drift equilibrium versus drift since a certain time. The model of immigration-drift is assumed to be either an infinite island or continent-island model of gene flow, which both give rise to the same likelihood function (Rannala & Hartigan 1996). The calculation of the likelihoods for the pure drift case is as described by O'Ryan *et al.* (1998), and implemented in the program dlik1.1. The program estimates the relative likelihoods of the two models using an MCMC procedure as described in Ciofi *et al.* (1999) simulating the extent of the interaction between drift and gene flow using the parameter F (the probability that two genes share a common ancestor within a population (Dhuyvetter *et al.* 2005).

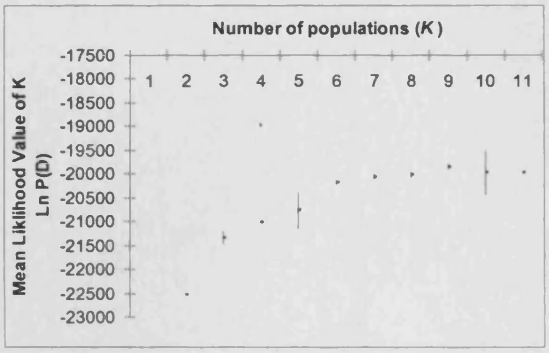
3.4. Results

3.4.1. Definition of populations based on Bayesian clustering techniques

The three Bayesian clustering techniques differed in their estimate of the number of optimum partitions (K). STRUCTURE showed optimum partitions with the highest likelihood value at $K = 9$, however it starts to plateaux at $K = 6$ (Figure 3.2a) but using the alternative method developed by Evanno *et al.* (2005) the largest rate of change ΔK shows $K = 4$ (Figure 3.2b), with smaller peaks at $K = 6$ and $K = 9$. GENELAND SPATIAL gave an optimum at $K = 6$ (Figure 3.3); which appeared to be heavily influenced by spatial information, with one individual (from the Shetland Isles) accounting for two of these populations. For BAPS4 SPATIAL the most likely number of clusters based on logML values varied between runs between $K = 8$ and 10, with the highest at $K = 9$.

The populations identified by each of the Bayesian clustering algorithms were mapped for comparison using ArcMap v 9.2 (ESRI 2007). Figures 3.4-3.7 display maps of the distribution of individuals to each of the optimum number of clusters derived by each of the Bayesian clustering techniques.

(a)



(b)

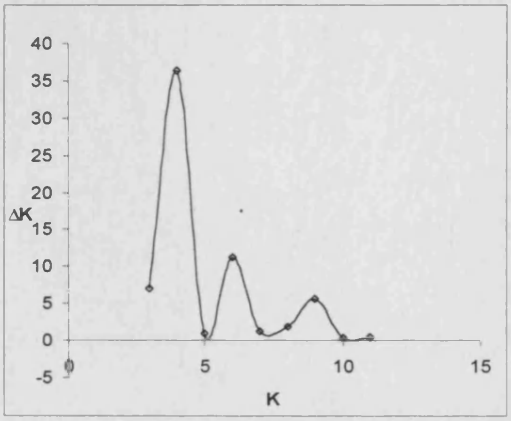


Figure 3.2 Estimation of *K* using STRUCTURE. (a) mean likelihood values for *K* 1-11, Ln P(D) (Pritchard *et al.* 2000). (b) Second order rate of change (ΔK) as a function of the number of populations *K*, based on Evanno *et al.* (2005).

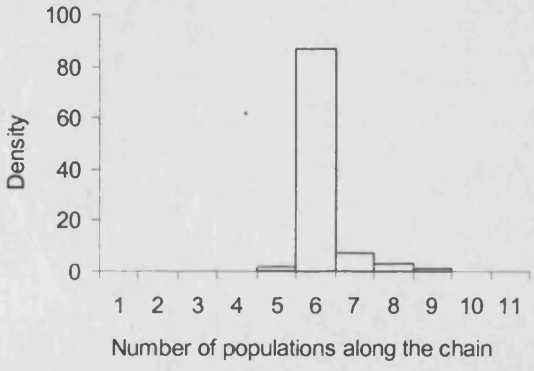


Figure 3.3 Estimation of *K* using GENELAND SPATIAL; Histogram showing the posterior density distribution of the number of clusters estimated from GENELAND analysis.

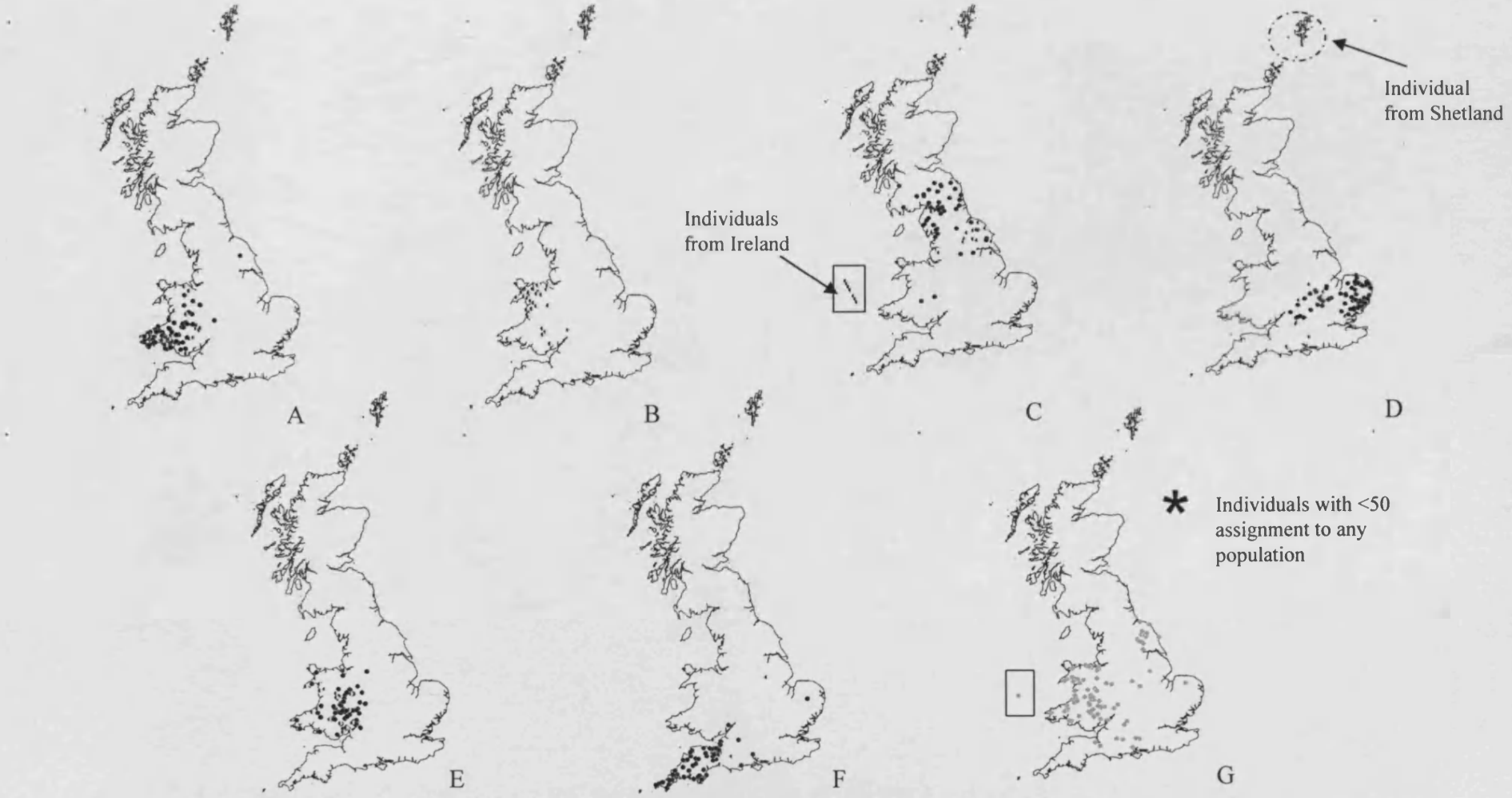


Figure 3.4. Distribution of individuals within each of the optimum number of clusters $K = 6$ derived by STRUCTURE.
Circles represent individuals; Black circles represent ≥ 75% assignment, ½ grey/½ black circles ≥50% assignment, grey circles <50% assignment to any population, mapped separately (map marked with an asterisk)

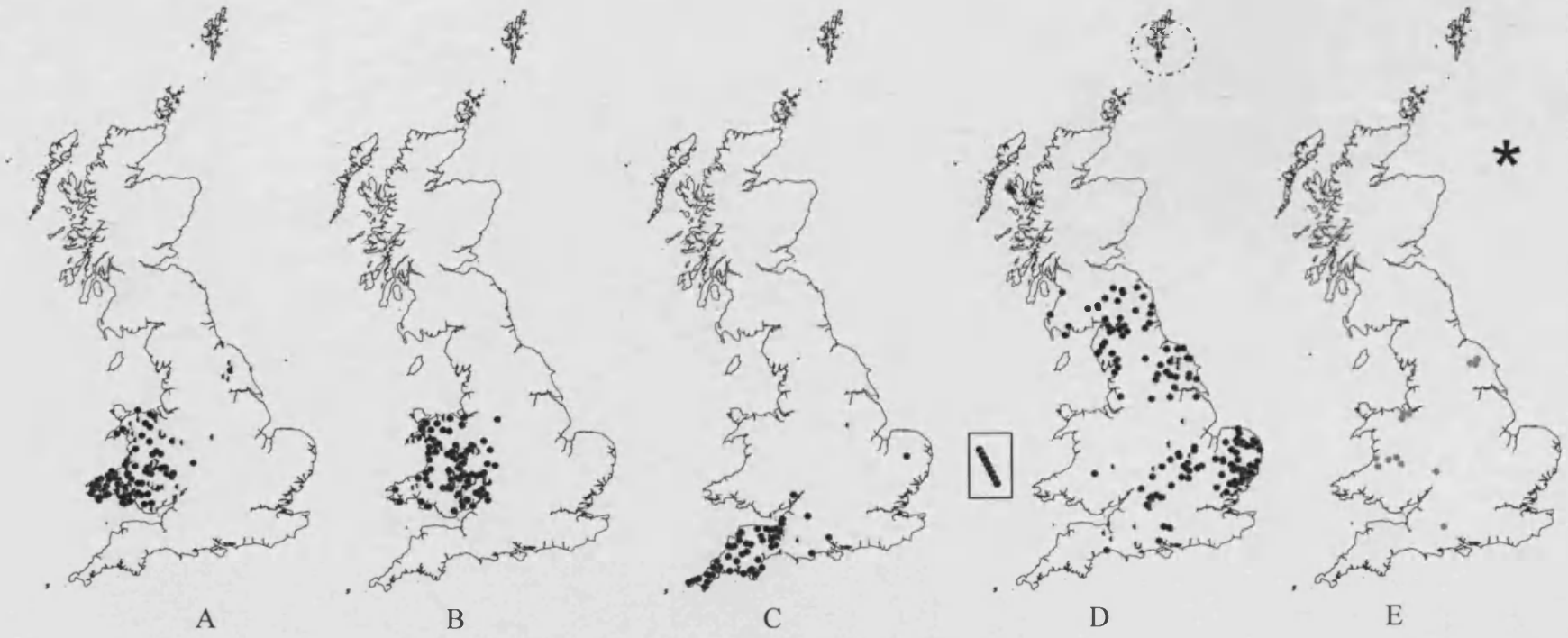


Figure 3.5. Distribution of individuals within each of the optimum number of clusters $K = 4$ derived by STRUCTURE.
Circles represent individuals; Black circles represent ≥ 75% assignment, ½ grey/½ black circles ≥50% assignment, grey circles <50% assignment to any population, mapped separately (map marked with an asterisk)

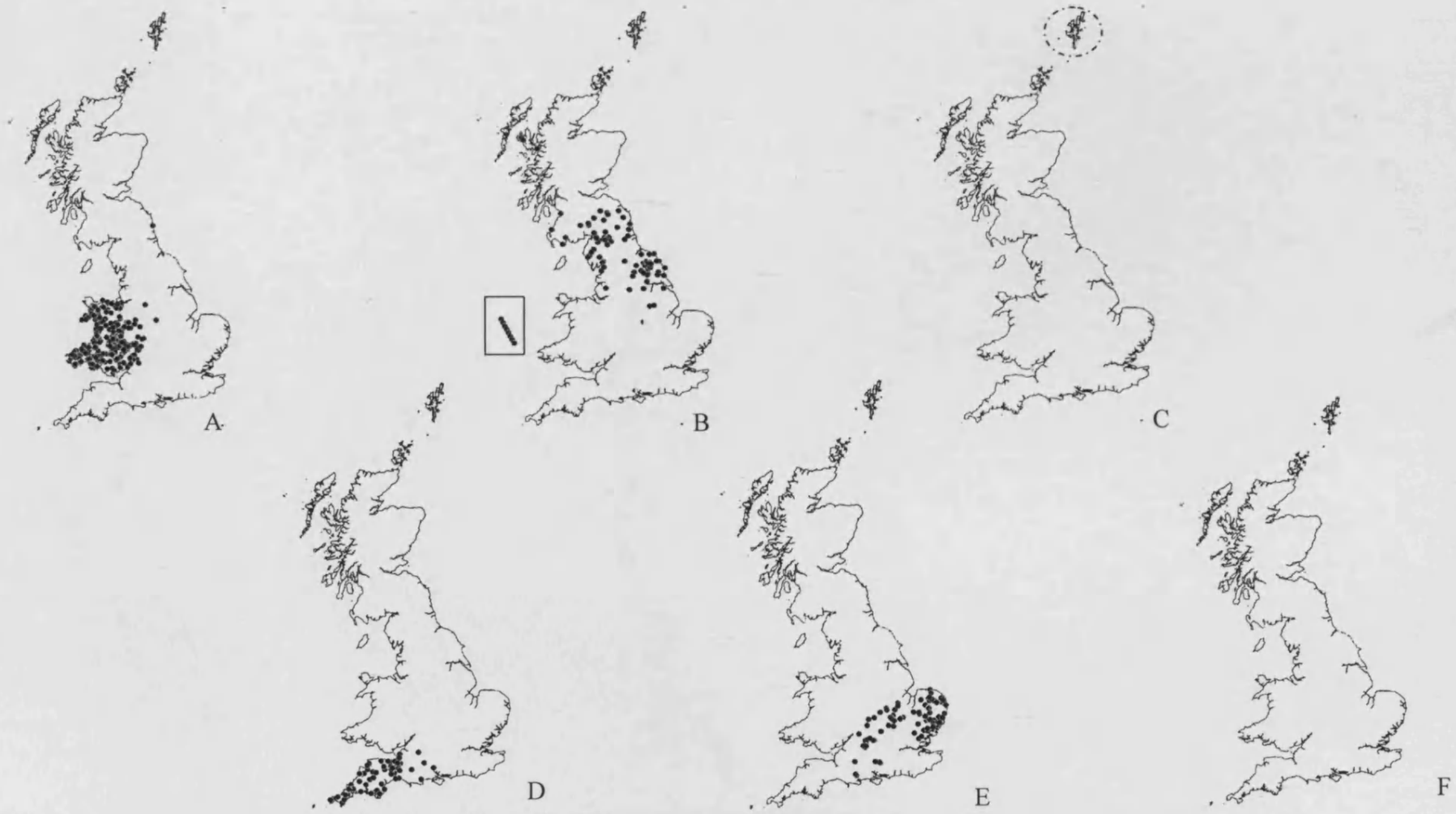


Figure 3.6. Distribution of individuals within each of the optimum number of clusters $K = 6$ derived by GENELAND SPATIAL.
Circles represent individuals; Black circles represent ≥ 75% assignment, ½ grey/½ black circles ≥50% assignment, grey circles

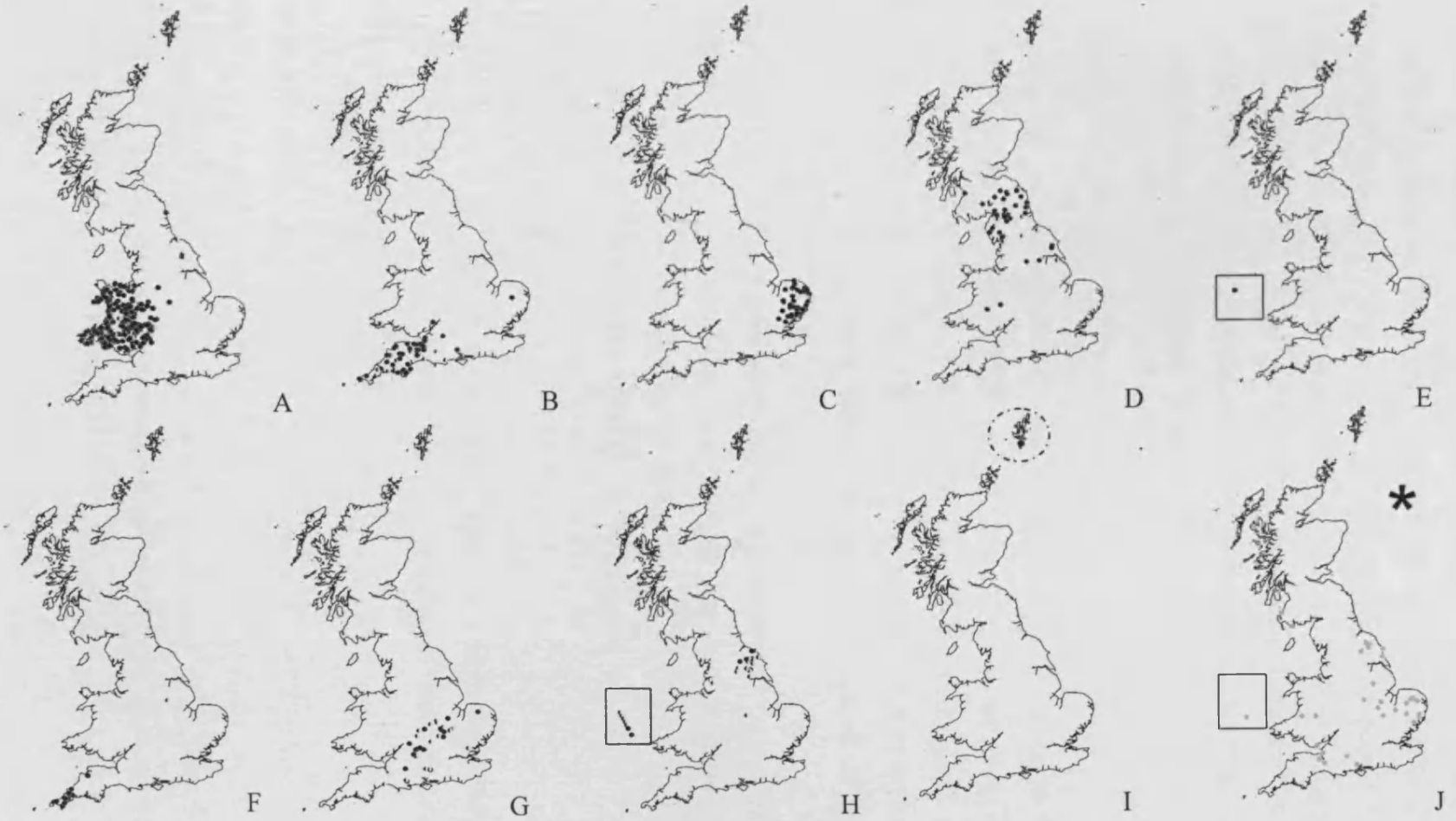


Figure 3.7. Distribution of individuals within each of the optimum number of clusters $K = 9$ derived by BAPS4 SPATIAL.
Circles represent individuals; Black circles represent ≥ 75% assignment, ½ grey/½ black circles ≥50% assignment, grey circles <50% assignment to any population, mapped separately (map marked with an asterisk)

GENELAND SPATIAL identified four main clusters, referred to hereafter as the 'Wales and Borders' (Figure 3.6a), 'North England' (Figure 3.6b), 'Southwest England' (Figure 3.6d) and 'Central England' (Figure 3.6c) clusters. Further subdivisions are identified by BAPS4 SPATIAL and by STRUCTURE (K = 4 and 6). BAPS4 SPATIAL and GENELAND SPATIAL identify a very similar Wales and Borders population (Figures 3.6a and 3.7a), which is further divided by STRUTURE at both $K = 6$ (Figures 3.4a, b, e) and $K = 4$ (Figures 3.5a and b). GENELAND SPATIAL and STRUCTURE identify a similar Southwest England population (Figures 3.6d and 3.5c), but this is divided by BAPS4 SPATIAL into two (Figures 3.7b and f), this division separating out individuals from Cornwall.

The remainder of the individuals were grouped together by STRUCTURE $K = 4$ (Figure 3.5d) however further sub division is found in this cluster; with GENELAND SPATIAL and STRUCTURE $K = 6$ identifying a similar 'North England' population (Figures 3.4c and 3.6b), which is divided into two by BAPS4 SPATIAL, the further division separating out individuals from Ireland and part of Yorkshire (Figure 3.7h).

Outside the four main clusters identified by GENELAND SPATIAL there are other notable results such as the assignment of the individual from Shetland into its own population by both GENELAND SPATIAL (Figures 3.6c and f) and BAPS4 SPATIAL (Figure 3.7i). For GENELAND SPATIAL it is assigned 50:50 to 2 additional populations.

Many individuals are not assigned by STRUCTURE (i.e. assignment < 50%) at $K = 6$ (Figure 3.4g) and $K = 4$ (Figure 3.5e). Assignment of these individuals was inconsistent between STRUCTURE runs.

GENELAND SPATIAL was conservative in its estimation of K when compared to STRUCTURE and BAPS4 SPATIAL. However, further analysis of each of the GENELAND SPATIAL populations separately assuming $K = 2$ identified further subdivisions that are similar to clusters identified by BAPS4 SPATIAL and STRUCTURE (Appendix 3.1-GENELAND SPATIAL $K = 9$).

3.4.2. Definition of populations based on progressive partitioning approach with Bayesian clustering techniques

Progressive partitioning was conducted on the otter dataset using each Bayesian Clustering technique, enforcing $K = 2$ at each stage. Figures 3.8-3.10 show the resulting groupings. In Figures 3.8-3.10 the clusters with the thick black boxed outline represent the regions. Clusters with a thin black boxed outline are clusters chosen to represent sub-regions. For BAPS4 SPATIAL the dashed arrow (----->) indicates steps not shown and the grey box contains the final partitions. Details of the selection criteria for regions and sub-regions are described below.

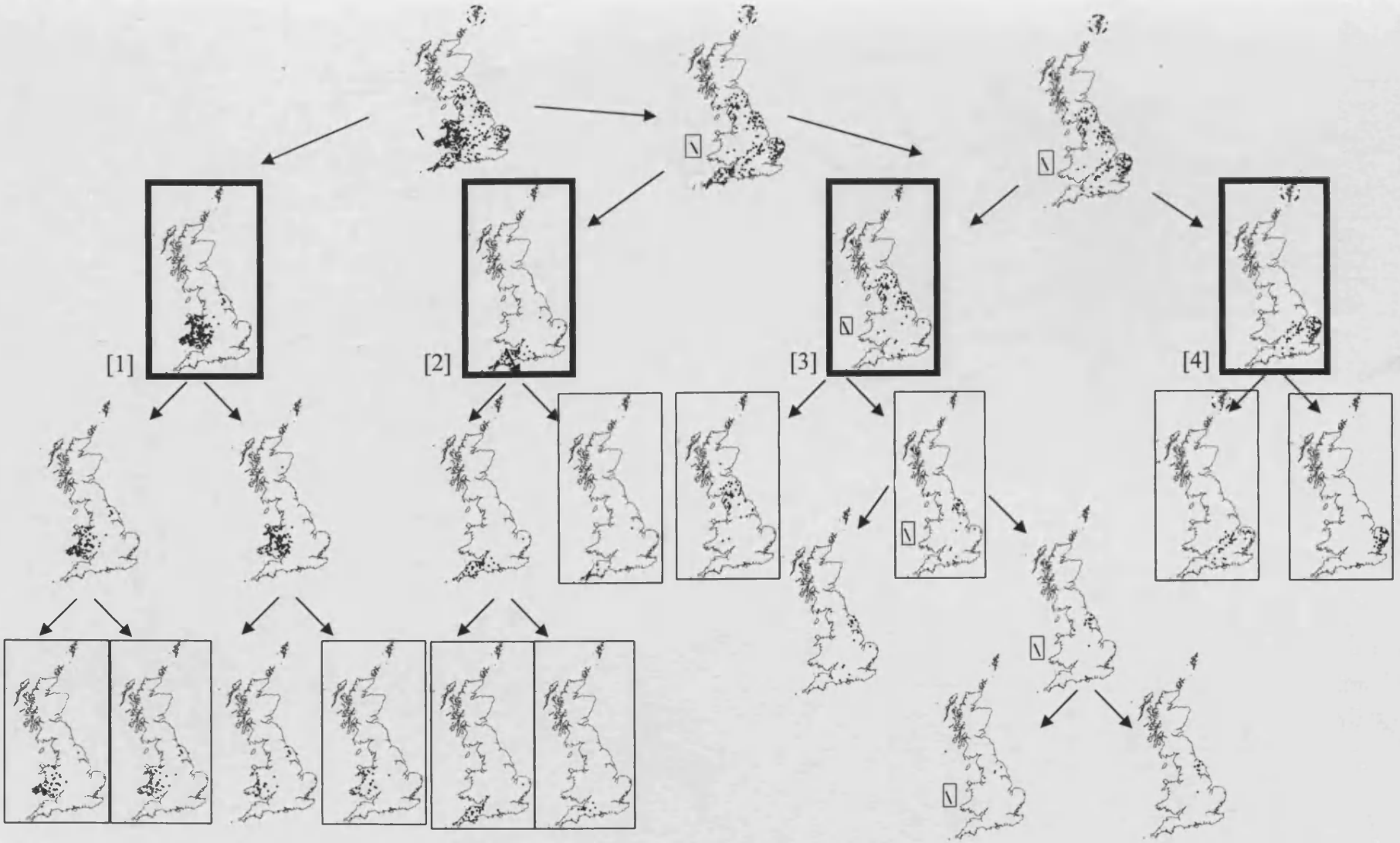


Figure 3.8. Distribution of individuals assigned to clusters identified by each step of the progressive partitioning approach (sequential steps of $K = 2$) using STRUCTURE. Circles represent individuals. Clusters with a thick boxed outline represent the regions. Clusters with a thin boxed outline are clusters chosen to represent sub-regions.

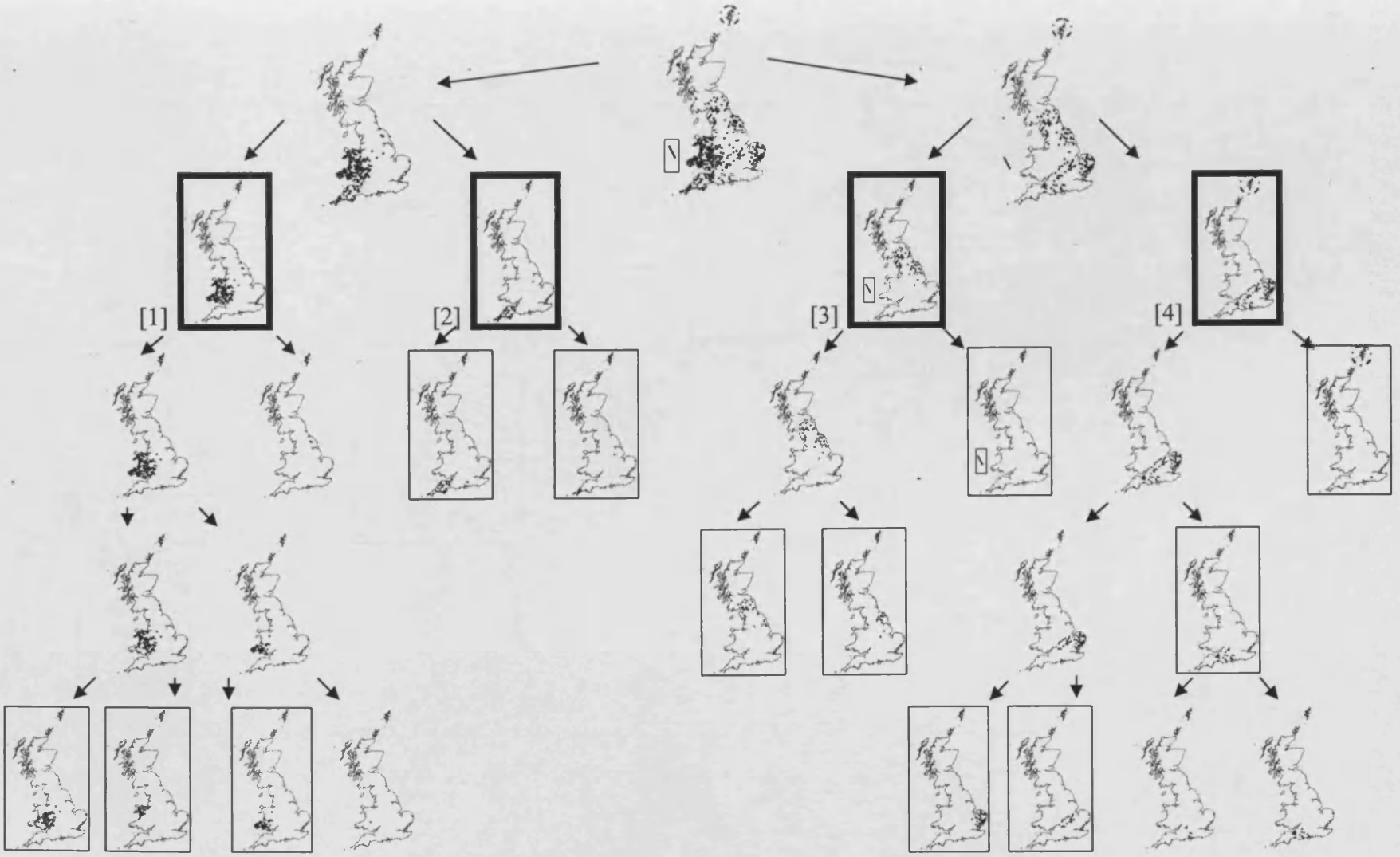


Figure 3.9. Distribution of individuals assigned to clusters identified by each step of the progressive partitioning approach (sequential steps of $K = 2$) using GENELAND SPATIAL. Circles represent individuals. Clusters with a thick boxed outline represent the regions. Clusters with a thin boxed outline are clusters chosen to represent sub-regions.

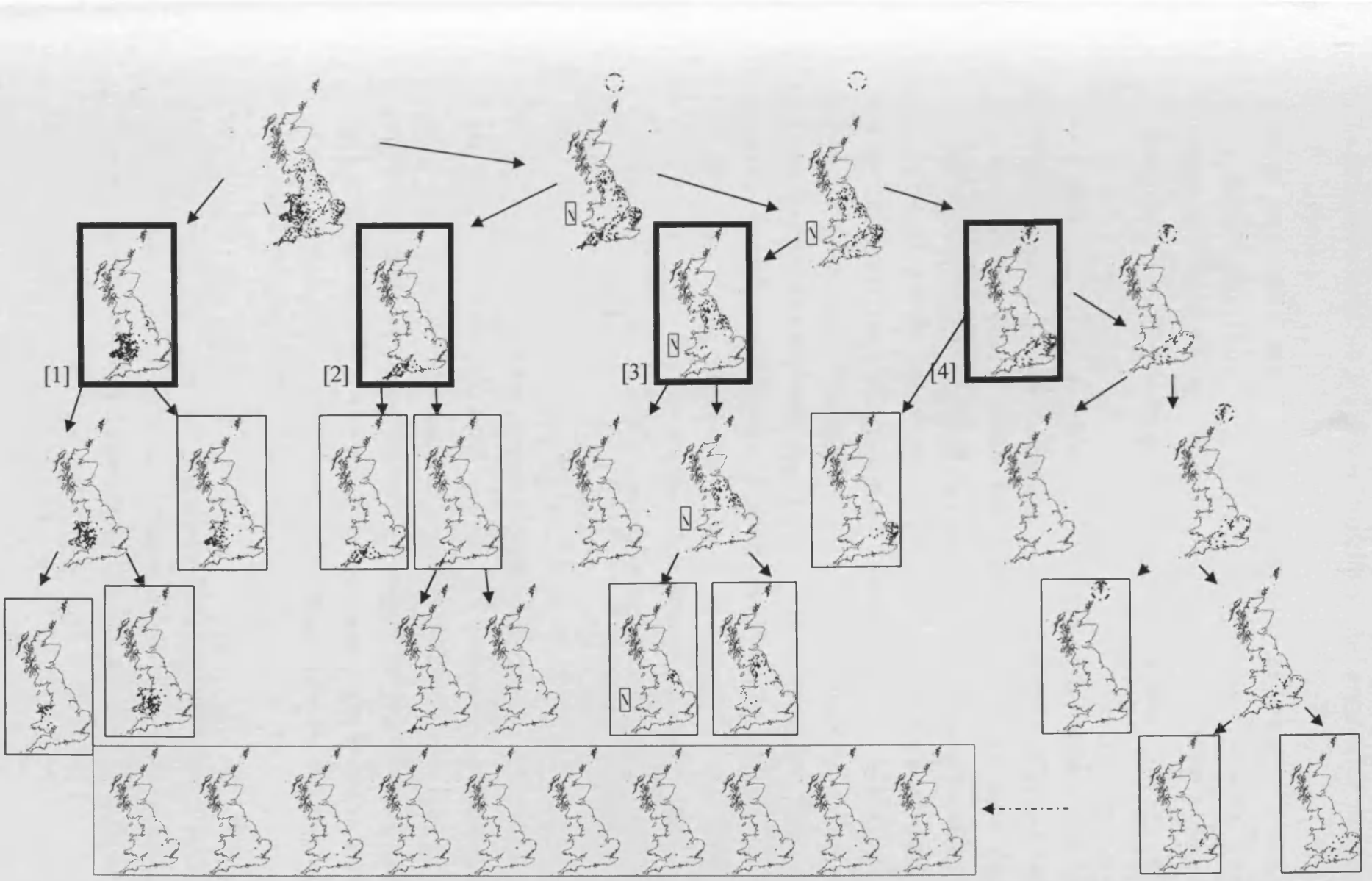


Figure 3.10. Distribution of individuals assigned to clusters identified by each step of the progressive partitioning approach (sequential steps of $K = 2$) using BAPS4 SPATIAL. Circles represent individuals. Clusters with a thick boxed outline represent the regions. Clusters with a thin boxed outline are clusters chosen to represent sub-regions.

3.4.2.1.Comparisons of population assignment for progressive partitioning approach

As with the optimal *K* method differences existed between the clusters identified by different softwares. There were however clusters that were consistent between methods in addition the use of progressive partitioning provided a method of identifying clusters at different degrees of genetic differentiation.

At a higher level of genetic differentiation there were four main clusters easily identified by all three Bayesian techniques (thick boxed clusters Figures 3.8-3.10), referred to here as regions: the Wales and Borders region [1], Southwest of England region [2] North England (including Irish samples) region [3], and a Central England region [4] (Figures 3.8-3.10). Further subdivisions varied between softwares. The Wales and Borders region [1] is divided into 4 overlapping clusters by STRUCTURE, 4 spatially distinct clusters by GENELAND SPATIAL, or 3 spatially distinct clusters by BAPS4 SPATIAL. The Southwest England region [2] is divided into 3 clusters by STRUCTURE and BAPS4 SPATIAL or 2 clusters by GENELAND SPATIAL. The North England region [3] is divided into 4 clusters by STRUCTURE, or 3 clusters by GENELAND SPATIAL and BAPS4 SPATIAL. The Central England region [4] is divided into 2 clusters by STRUCTURE, 5 clusters by GENELAND SPATIAL, or 13 clusters by BAPS4 SPATIAL.

To determine the most likely population substructure, clustering solutions given by the different methods (optimal *K* methods, and progressive partitioning) were compared. Sub-regions are suggested to be clusters that are shown by more than one method ideally with agreement between all techniques. Sub-structures deriving from the four regions are shown in Figure 3.12. Clusters that agreed between Bayesian clustering techniques for the progressive partitioning approaches and qualified for combination are surrounded by a thin black box in Figures 3.8-3.10.

In Wales and Borders region [1] a Southwest Wales sub-region (1a), a Northwest Wales sub-region (1b) and a Mid-Eastern Wales sub-region (1c) are identified by the progressive partitioning method of BAPS4 SPATIAL and GENELAND SPATIAL, as well as STRUCTURE optimal $K = 6$.

In the Southwest of England region [2] a sub-region on the tip of the Southwest Peninsula (2a) is overlapped by a larger sub-region (2b) identified by the progressive partitioning method of STRUCTURE, BAPS4 SPATIAL and GENELAND SPATIAL.

For the North England region [3] two sub-regions were identified; a North England/ Southern Scotland sub-region (3a) and a sub-region including the Irish samples which cluster with samples in North Yorkshire (3b). These sub-regions were identified by BAPS4 SPATIAL for both the optimal *K* method and progressive partitioning method. STRUCTURE progressive partitioning method also identifies these sub-regions but also finds further substructure within sub-region 3b. GENELAND SPATIAL progressive partitioning method identifies similar sub-regions however partitions out the Irish samples from the region early on. The separate Irish sub-region was not identified as a separate sub region in Figure 3.12, as not all Bayesian techniques agree and this could be a remnant of the strong spatial prior attributed by GENELAND.

There was much variation in the subdivision of the Central England region [4]. Three sub-regions were tentatively identified. There was strong agreement between the progressive partitioning method of all Bayesian techniques for an East Anglia sub-region (4a), and an Oxfordshire sub-region (4b), although the latter differed in its spatial extent between all techniques (the core area of individuals in this sub-region were the same between the methods). An additional sub-region the West Country sub-region (4c) was identified by GENELAND SPATIAL and situated at the western side of the Central England region [4] and adjacent to the Southwest England region [2]. Sub-region (4a) was not identified by BAPS4 SPATIAL; however, it was identified by STRUCTURE progressive partitioning method but through the partitioning of the Southwest England region [2].

Other clusters either represented few (1-5) individuals or further sub-structuring was represented by only one method and so these clusters were not defined as sub-regions. One exception was made for the individual from the Shetland Isles (5a), identified by both GENELAND SPATIAL and BAPS4 SPATIAL in the progressive partitioning method and crucially by both in the optimal *K* method. It is thought

likely that this individual does represent a subpopulation but insufficient sampling in this area precludes confirmation.

3.4.3. Genetic statistics

Population genetic statistics were estimated fro the groups of individuals representing on selected individuals from the regions and sub-regions identified above. Selection criteria for regional analysis (individual assignment >0.9, agreement between all softwares) were met by 454 individuals (out of 566, 80.2%); Selection criteria for sub-region analysis (assignment > 0.5, agreement by at least two of the softwares) were met by 332 individuals from 566 genotypes (58.7%).

3.4.4. Regions

According to the screening criteria (3.3.4) 266 individuals were assigned to the Wales and Borders region, 50 individuals to the Southwest England region, 67 to North England region and 71 individuals assigned to the Central England region.

3.4.4.1. Population genetic diversity: regions

Allelic diversity ranged from 3.73 alleles per locus for the Southwest England region to 5.6 alleles in the North England region (Table 3.2). Each region showed private alleles ranging from one in the Southwest England region to 17 in the Central England region (Table 3.2). Values of observed heterozygosity (H_O) were in the range of 0.46 to 0.65, and values of (H_E) were in the range 0.49-0.70 (Table 3.2). Otters from Southwest England and the Wales and borders region showed the lowest levels of genetic diversity, while individuals from the North England region and the Central England region showed the highest levels of genetic diversity.

All of the regions displayed significant heterozygote deficiencies (F_{IS} = 0.053 to 0.075) as compared to Hardy–Weinberg expectations (Table 3.2; $*P$ <0·05, $**P$ <0·01, $***P$ <0·001). The departure from Hardy–Weinberg expectations in these regions suggests a Wahlund effect (Wahlund 1928). When two population samples are combined and analysed for departures from Hardy–Weinberg expectations as a single unit, the number of homozygotes become artificially increased because of the hidden population structure, therefore there may be further structuring present within the regions.

Table 3.2. Average summary genetic statistics for each region – over 15 loci

Locus	N	P	A	Ap	He	Ho	F_{is}	Ar	Au
[1] Wales and Borders	266	1	4.4	4.4	0.52	0.49	0.058***	3.90	2
[2] Southwest England	50	1	3.73	3.73	0.49	0.46	0.075**	3.73	1
[3] North England	67	1	5.6	5.6	0.70	0.65	0.065**	5.44	9
[4] Central England	71	1	5.53	5.53	0.68	0.64	0.053*	5.37	17

Population has a significant deviation from HWE (* p<0.05, **p<0.01,***p<0.001)

N- Sample size; P- proportion of polymorphic loci; A- mean number of alleles per locus; Ap- mean number of alleles per polymorphic locus; Ho- Observed heterozygosity; He- expected heterozygosity; F_{is} inbreeding coefficient Ar- allelelic richness; Au, number of unique alleles

All populations were found to be highly significantly differentiated from each other on analysis of the F_{ST} values (Table 3.3), with the North England region having the lowest pairwise F_{ST} value with the Central England region of (0.1), the Wales and borders region and Southwest of England region showing the greatest degree of differentiation with an F_{ST} value of 0.28. The North England region had the lowest overall F_{ST} values when compared with all other populations.

Table 3.3. Pairwise (F_{ST}) values between regions identified by Bayesian clustering

	[1] Wales and Borders	[2] Southwest England	[3] North England	[4] Central England
[1] Wales and Borders	0			
[2] Southwest England	0.28085***	0		
[3] North England	0.18587***	0.19876***	0	
[4] Central England	0.22308***	0.23021***	0.09981***	0

*** F_{ST} values highly significant, P < 0.001

3.4.4.2. Bottlenecks results: regions

The three mutation model scenarios were run, and the Wilcoxon statistic calculated. The results were dependent on the mutation model used, the infinite allele model (IAM) model for all populations found a significant *p* value <0.05 for heterozygote excess; there were no significant p values found for any population under the two-phase model of mutation (TPM) or step-wise mutation model (SMM).

3.4.4.3. 2mod model for immigration drift equilibrium: regions

The North England [3] and the Central England [4] regions have low F value modes (0.09, 0.16) which indicate that the probability of genes being identical by descent was low and more likely to be influenced by the gene flow model (Figure 3.11). In

contrast, the Wales and Borders [1] and the Southwest England [2] regions had higher F values (0.35, 0.38) which suggests that these populations are more influenced by drift.

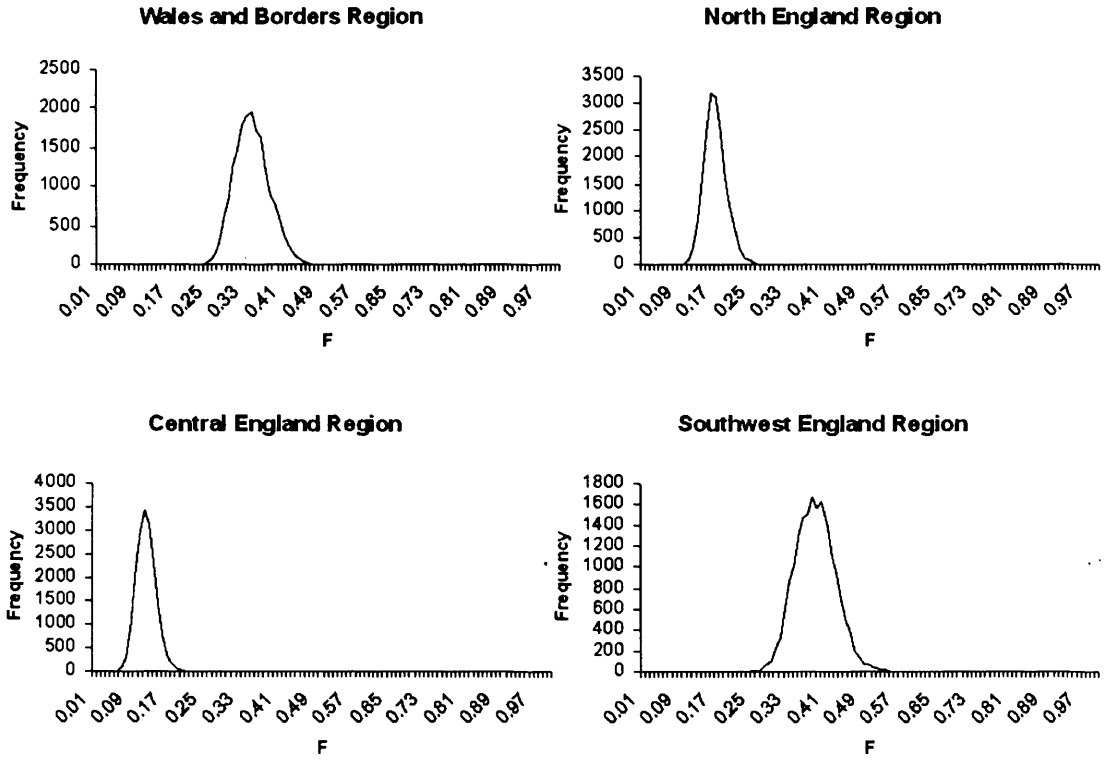


Figure 3.11. 2mod results for the Regional populations

3.4.5. Sub-Regions

According to the selection criteria (3.3.4) 332 individuals from 566 genotypes (58.7%) were assigned to 11 sub-regions and used for further analysis (Figure 3.12).

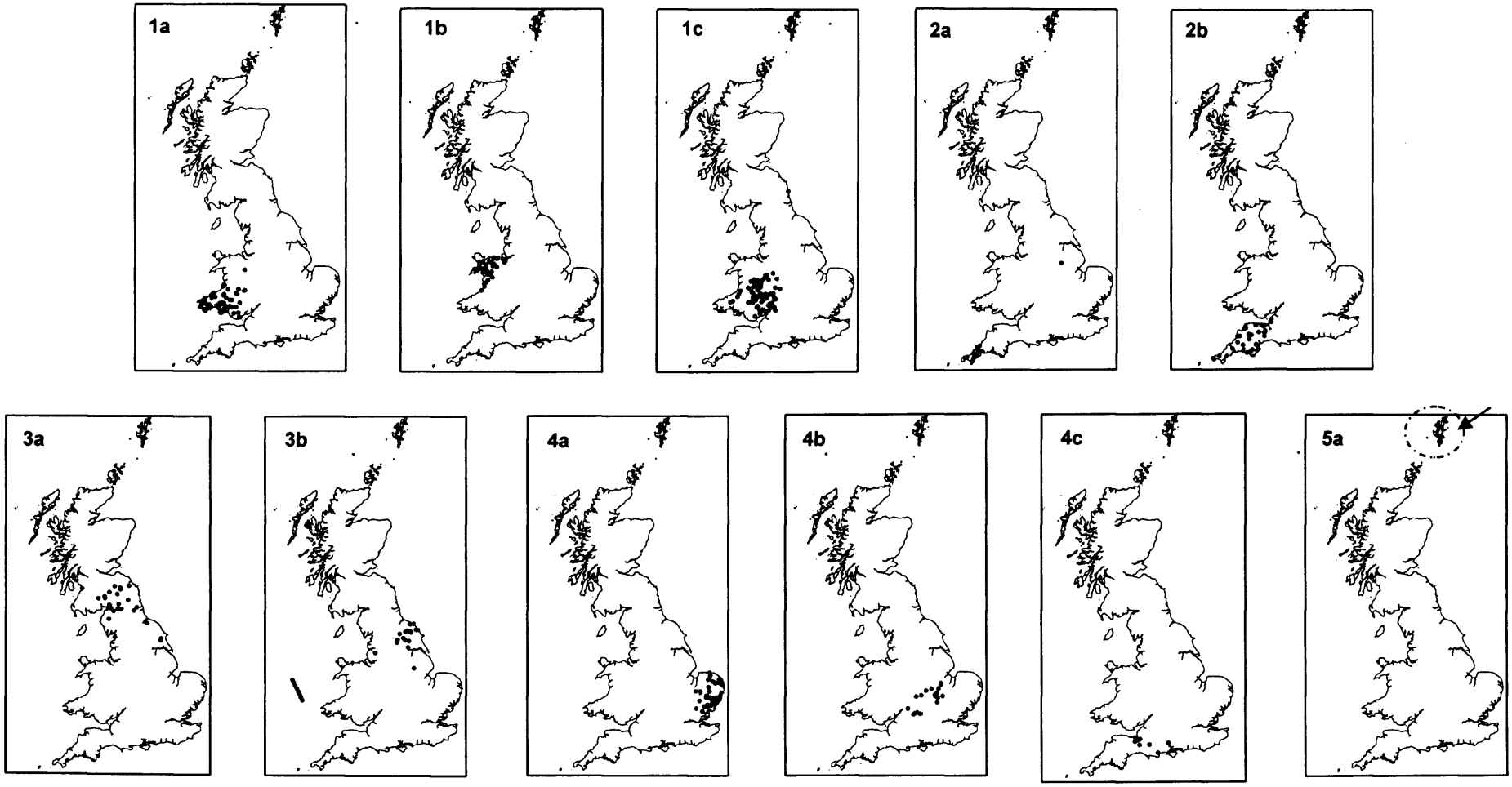


Figure 3.12. Location of individuals assigned to the 11 sub-regions by consensus between optimal *K* and progressive partitioning methods applied using three Bayesian clustering softwares.

3.4.5.1. Population genetic diversity: sub-regions

Allelic diversity ranged from 2.93 alleles per locus for the sub-region (2b) to 5.73 alleles in the sub-region (3b) (Table 3.4). Some sub-regions had no private alleles, while the sub-region (3b) showed the greatest number with five unique alleles, representing a third of all unique alleles. The sub-regions (4a-c) derived from the Central England region [4] between them possessed a further five unique alleles (Table 3.4). Values of observed heterozygosity (H_O) were in the range of 0.45 to 0.7, and values of expected Heterozygosity (H_E) were in the range 0.44-0.72. The levels of genetic diversity in the sub-regions reflected the levels of genetic diversity in the regions from which they were derived.

The Irish and North Yorkshire sub-region and the East Anglia sub-region displayed significant heterozygote deficiencies (F_{IS} = 0.053 to 0.075) as compared to Hardy–Weinberg expectations (*P <0·05, **P <0·01, ***P <0·001). This departure from Hardy–Weinberg expectations has occurred despite the fact that these sub-regions show the highest levels of genetic diversity of all the sub-regions. This suggests that there is further sub-structuring (Wahlund effect) within these two sub-regions.

Table 3.4. Average summary genetic statistics for all 11 sub-regions – over 15 loci.

Population	N	P	A	Ap	He	Ho	Fis	unique alleles
(1a) Southwest Wales	62	1	4.07	4.07	0.51	0.50	0.023	1
(1b) Northwest Wales	8	1	3.07	3.07	0.51	0.53	-0.027	0
(1c) Mid-Eastern Wales	97	1	3.8	3.8	0.47	0.47	0.001	0
(2a) Southwest Peninsula	13	1	2.93	2.93	0.44	0.45	-0.022	1
(2b) Southwest England	32	0.93	3.33	3.5	0.46	0.45	0.028	1
(3a) North of England/ Scottish Borders	26	1	4.8	4.8	0.68	0.65	0.045	2
(3b) Irish and North Yorkshire	34	1	5.73	5.73	0.72	0.7*	0.022*	5
(4a) East Anglia	39	1	4.53	4.53	0.61	0.59*	0.034*	2
(4b)Oxfordshire	21	1	5.33	5.33	0.67	0.7	-0.038	3
(4c) West Country	9	1	3.73	3.73	0.58	0.58	-0.002	0
(5a) Shetland	1	0.27	1.27	2	0.27	0.27	0	0
Mean	31.09	0.93	3.87	3.95	0.54	0.54	0.008	15

Population has a significant deviation from HWE (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

N- Sample size; P- proportion of polymorphic loci; A- mean number of alleles per locus; Ap- mean number of alleles per polymorphic locus; Ho- Observed heterozygosity; He- expected heterozygosity; F_{is} inbreeding coefficient Ar- allelelic richness; Au, number of unique alleles

All sub-regions were found to be highly significantly differentiated from each other on analysis of the F_{ST} values (Table 3.5), (apart from sub-region 5a –sample size of 1). The F_{ST} values between sub-regions derived from the same region were lower than between sub-regions derived from different regions, for example between sub-region 1a and 1c $F_{ST} = 0.12$, sub-regions 2a and 2b $F_{ST} = 0.17$, sub-region 1a and 2b $F_{ST} = 0.30$.

Table 3.5. Pairwise multi-loci (F_{ST}) values for 11 sub-regions identified by Bayesian Clustering

Within region comparison Between regions comparison

Sub-region	1a	1b	1c	2a	2b	3a	3b	4a	4b	4c	5a
1a	-										
1b	0.09***	-									
1c	0.12***	0.10***	-								
2a	0.32***	0.36***	0.36***	-							
2b	0.30***	0.32***	0.32***	0.17***	-						
3a	0.20***	0.25***	0.16***	0.26***	0.23***	-					
3b	0.17***	0.22***	0.14***	0.20***	0.21***	0.06***	-				
4a	0.24***	0.30***	0.22***	0.30***	0.29***	0.12***	0.14**	-			
4b	0.22***	0.29***	0.20***	0.26***	0.26***	0.09***	0.11**	0.12***	-		
4c	0.26***	0.29***	0.24***	0.18***	0.05***	0.13***	0.14**	0.20***	0.16***	-	
5a	0.43	0.48	0.41	0.48	0.44	0.21	0.26	0.32*	0.21*	0.31	-

*significant p-value <0.05, **significant p-value <0.05, ***highly significant p-value<0.001

3.4.5.2. Bottleneck: sub-regions

The three mutation model scenarios were run, and the Wilcoxon statistic calculated. The results were dependent on the mutation model used (Table 3.6). The IAM model found significant heterozygote excess in 6 of 10 sub-regions (Sub-region 5a was excluded from this analysis sample size of 1). Under the TPM model however, significant heterozygote excess was found only for the North England/ Scottish borders sub-region. No significant results were found for the SMM mutation model.

Table 3.6. Bottleneck results for the 11 sub-regions identified by Bayesian Clustering.

	IAM		TPM		SMM	
Sub-region	one tail for H excess	two tails	one tail for H excess	two tails	one tail for H excess	two tails
1a	0.00754*	0.01508*	0.7894	0.45428	0.91559	0.18762
1b	0.11465	0.22931	0.53296	0.97797	0.61923	0.80396
1c	0.00754*	0.01508*	0.82043	0.3894	0.87381	0.27686
2a	0.16513	0.33026	0.70026	0.63867	0.73776	0.5614
2b	0.01477	0.02954*	0.6651	0.71484	0.76843	0.50159
3a	0.00008**	0.00015**	0.04163*	0.08325	0.10388	0.20776
3b	0.00003**	0.00006**	0.06027	0.12054	0.1947	0.3894
4a	0.00021**	0.00043**	0.22714	0.45428	0.64014	0.76154
4b	0.00418*	0.00836*	0.68066	0.67877	0.82043	0.3894
4c	0.07571	0.15143	0.38077	0.76154	0.48898	0.97797

P values for the Wilcoxon statistic calculated

H = Heterozygote. BOTTLENECK 1.2 (Cornuet & Luikart 1996), assuming an infinite allele model (IAM), a step-wise mutation model (SMM), or a two-phase model of mutation (TPM, with 95% SMMs).

3.4.6. Populations based on otter population history

3.4.6.1. Population genetic diversity: *a priori* defined populations

The *a priori* chosen groups (Figure 3.1) reflect the genetic summary statistics of those identified by Bayesian clustering techniques. Otter strongholds of Wales and Borders and Southwest England have lower levels of heterozygosity than the North England population (Table 3.7). Areas of population reinforcement in Central England and North Yorkshire have high levels of genetic diversity, with heterozygosity levels equivalent to that of the North England stronghold population, the highest levels of allelic richness, and 15 unique alleles between them.

F_{ST} values show that all these populations are highly significantly differentiated from each other with Wales and Borders and Southwest of England populations having the highest F_{ST} values (Table 3.8). The greatest numbers of unique alleles (11) were found in the reinforcement areas of Central England whilst no unique alleles were found in the Wales and Borders population.

3.4.6.2. Bottleneck: *a priori* defined populations

All populations showed heterozygosity excess under the IAM with the North Yorkshire population also showing heterozygosity excess under the TPM model. The Southwest population shows a significant deficit in heterozygosity under TPM and SMM (Table 3.9).

Three of the five pre-defined populations displayed significant heterozygote deficiencies (F_{IS} = 0.053 to 0.075) as compared to Hardy–Weinberg expectations. The East Anglia and the North Yorkshire populations show a significant departure from Hardy–Weinberg expectations despite showing the highest levels of genetic diversity (Table 3.7). The third population showing a significant departure from HWE is the Southwest England population which has the lowest levels of genetic diversity. This suggests that there is further genetic sub-structuring (Wahlund effect) within these populations.

Table 3.7. Average summary genetic statistics over all loci for five *a priori* defined populations based on population history.

a priori defined population	*N*	*P*	*A*	*Ap*	*He*	*Ho*	F_{IS}	*Au (Private alleles)*	*Ar (Allelic Richness)*
Wales and Borders	50	1.00	4.20	4.20	0.54	0.53	0.01	0	4.13
Southwest England	50	1.00	4.47	4.47	0.53	0.49	0.072***	1	4.38
Central England	50	1.00	5.67	5.67	0.69	0.63	0.092**	11	5.58
North England	48	1.00	5.27	5.27	0.68	0.63	0.07	8	5.18
North Yorkshire	43	1.00	5.40	5.40	0.70	0.68	0.035**	4	5.4
Mean	48.2	1.00	5.00	5.00	0.63	0.59			

Table 3.8. Pairwise F_{ST} values for five populations identified *a priori* based on otter population history

a priori defined population	Wales and Borders	Southwest England	Central England	North England	North Yorkshire
Wales and Borders	0				
Southwest England	0.22929***	0			
Central England	0.17821***	0.20823***	0		
North England	0.17162***	0.19024***	0.10374***	0	
North Yorkshire	0.10364***	0.16574***	0.0832***	0.05067***	0

Table 3.9. Bottleneck results for five *a priori* defined populations based on population history.

	IAM		TPM		SMM	
a priori defined population	one tail for H excess	two tails	one tail for H excess	two tails	one tail for H excess	two tails
Wales and Borders	0.00134*	0.00269*	0.83487	0.35913	0.92429	0.16882
Southwest England	0.0365*	0.073	0.99097	0.02155*	0.99866	0.00336*
East Anglia	0.00011**	0.00021**	0.1514	0.3028	0.55481	0.93408
North England	0.00003**	0.00006**	0.35986	0.71973	0.57654	0.89038
North Yorkshire	0.00003**	0.00006**	0.04163*	0.08325	0.13843	0.27686

P values for the Wilcoxon statistic calculated

H = Heterozygote. BOTTLENECK 1.2 (Cornuet and Luikart 1996), assuming an infinite allele model (IAM), a step-wise mutation model (SMM), or a two-phase model of mutation (TPM, with 95% SMMs).

3.5. Discussion

3.5.1. Otter population structure

Bayesian clustering techniques identified that the sampled otters could be divided into four regional populations. These reflect the known population history, with historic strongholds of the North England, Southwest England, and Wales and Borders represented, as well as the Central England region where OT introductions were carried out. These regional populations showed significant departure from HWE and could be further subdivided into 10 sub-regions; an individual from Shetland (5a) was assigned to its own 11th population. Only two of these sub-regions showed significant departure from HWE, both in areas where populations had been reinforced with rehabilitated (3b, North Yorkshire) or captive bred (4a, East Anglia) otters.

All the regions are highly significantly differentiated from one another, with the Southwest England region [2] showing the greatest effect of isolation, followed by Wales and Borders [1]. The genetic statistics indicate that these populations are isolated and despite being neighbouring populations there is no gene flow between them. This agrees with the finding of Dallas *et al.* (2002) who suggested there was no gene flow between the otter strongholds. In addition the lack of gene flow between otter populations is supported by the findings of Stanton *et al.* (2009) which found differences in the haplotype diversity between the regions within the UK.

Of the four regions, the Wales and Borders [1] and Southwest England [2] regions posses the lowest levels of genetic diversity, with few unique alleles. Despite the lack of evidence of them having recently gone through a bottleneck, they show evidence of being affected by genetic drift (2mod) and they both show additional sub-structuring. Dallas *et al.* (2002) also identified a southwest peninsula sub-region (2a) within the Southwest region [2] but did not identify sub-structuring (1a,b,c) in the Wales and Borders region [1], (possibly due to their small sample size in this area). Latch *et al.* (2008) also found sub-structure in what appeared to be a single continuous population of the North American river otter (*Lutra canadensis*) population in southern Louisiana, with no obvious landscape features that could account for the identified genetic discontinuities. Further analysis should be

undertaken to identify why these sub-structures exist, for example, correlation with landscape features measuring historical and contemporary factors affecting gene flow.

The North England [3] and the Central England [4] regions have higher levels of genetic diversity, with the results from 2mod indicating that both populations are more likely to be under the influence of gene flow rather than drift. The North England region [3] is a known otter stronghold and is adjacent to a strong population of otters in Scotland. Scottish populations are reported to have greater genetic diversity than southern UK populations (Dallas *et al.* 2002), and it is possible that there has been and continues to be immigration into this population from other populations in Scotland. The North England region [3] also contains areas where rehabilitated otters were introduced by the Vincent Wildlife Trust (VWT). The Central England region [3] also has high genetic diversity and 2mod results reflect immigration into this population; this may be attributable to the reinforcement campaign by the OT. Both these regions also show further sub-structuring (3a,b and 4a,b,c).

It is notable that one of the sub-regions identified within the North England region suggests a similarity between samples from Ireland and those from North Yorkshire (3b); this is evidence of the success of VWTs rehabilitation program, which released otters into North Yorkshire from a number of source locations including Northern Ireland (Rosie Green, Pers.comm).

The individual from Shetland (5a) was classed into its own population, despite not possessing unique alleles, perhaps reflecting the effects of isolation and genetic drift on its genotype.

The Central England population [4] was concentrated in East Anglia and thought to be small, fragmented and unviable (Crawford *et al.* 1979, Lenton *et al.* 1980), but otter surveys (Strachen & Jefferies 1996; Crawford 2003) have shown that this population has since expanded considerably. Since otters can travel up to 40 km a day (Durbin 1998), and radio tracking studies have identified large home range sizes (38.8 $\pm$ 23.4 km Green *et al.* 1984) there was an expectation that a wave of otters

would be dispersing into Central England from otter strongholds in the west (southwest England and the Welsh borders) and from the north (Scotland) (Coxon *et al.* 1999; Conroy & Chanin 2000). However, the genetic data does not support this expectation, showing no evidence of contributions from adjacent populations; it seems probable that the increase in numbers results primarily from reinforcement with captive bred otters by the OT or natural expansion of the remnant otter population.

The Central England region is probably the most anthropogenically influenced population, having suffered the effects of persecution and pollution, and by the early 1980s its survival seemed doubtful (Strachen & Jefferies 1996). The OT carried out 117 releases between 1983 and 1999, from captive bred stock. Unfortunately, details of some of the source populations have not been revealed. Given the population history, a population bottleneck might be expected but is not supported by the genetic data, which shows high levels of genetic diversity compared with the otter strongholds. The 2mod results indicate that it is likely that there has been gene flow into the population. The high genetic diversity and high number of unique alleles (17 unique alleles compared to nine in North England, two in Wales and Borders and one in the Southwest England regions) suggests that the founding stock was bolstered from individuals not sourced from any of the surrounding populations. There is also further subdivision within this region (sub-regions 4a and 4b) which need further fine scale analysis to interpret the possible causes. Analysis of the breeding stock and captive breeding lineages of the reintroduced otters would add much information to the understanding of this population.

3.5.2. Comparisons of Bayesian clustering techniques

An important function of Bayesian Clustering techniques is to provide an estimate of the number of populations (K) in a dataset. This study found differences in the estimated value of K between techniques and one technique could provide different estimates of K in different runs. The difficulty of estimating K is represented by the warning given by the authors of STRUCTURE (Pritchard *et al.* 2000; Pritchard & Wen 2003), that the estimated log probability of data $\Pr(X|K)$ used to identify K is really only an indication of the number of clusters and an ad hoc guide. There are two recognised methods used to identify K from the results of STRUCTURE, both of which

gave different results $K = 6$ (following Pritchard *et al.* 2000) and $K = 4$ (following Evanno *et al.* 2005). BAPS4 SPATIAL differed in its estimate of the most likely number of partitions K between runs (8 - 10) this was a result of using a different string of maximum populations in the input parameters (not shown).

When using an estimated optimal K both BAPS4 SPATIAL and STRUCTURE could not assign some individuals to a single population (>50% assignment) and split the assignment of these individuals to three or more populations (Figures 3.4g, 3.5e and 3.7j). GENELAND SPATIAL identifies 4 distinct populations that tie in well with the population history and singles out a unique Shetland individual, however, after further analysis of these distinct populations further substructure can be identified (Appendix 3.1) resembling those produced by BAPS4 SPATIAL. This suggests that BAPS4 SPATIAL is identifying populations to a higher level of population structure than GENELAND SPATIAL.

Differing estimates of K is a typical complication when using multiple Bayesian clustering techniques (Carmichael *et al.* 2007; Coulon *et al.* 2008; Lecis *et al.* 2008). Therefore, the user is left with the dilemma of choosing one set of results over others.

Carmichael *et al.* (2007) chose to use one Bayesian clustering technique (STRUCTURE) to estimate the number of optimum partitions (K) and used this as the value for further Bayesian clustering techniques (GENELAND), thus allowing for the same number of clusters to be compared between methods. In this study optimal K values ranged from 4-9, however even when the same number of clusters were identified by the different techniques they partitioned the data differently and identified some differing clusters; a feature also found by Carmichael *et al.* (2007). The differences in the estimation of K indicate that some of the Bayesian clustering techniques may be missing minor underlying genetic clusters (Corander *et al.* 2008) and are identifying partitions at differing degrees of genetic differentiation. Therefore, it is difficult to not only compare the partitions identified by different techniques but even more difficult to try to find agreement to combine the results to make a more robust clustering solution.

Based on the results of GENELAND SPATIAL estimating $K = 6$ and that subsequent analysis divides these clusters further (Appendix 3.1), if Bayesian clustering techniques are used to identify partitions at lower than optimal estimates of K they should be identifying partitions with a greater degree of genetic differentiation. As the value of K decreases this would thus reflect the grouping of individuals into more differentiated clusters as found by Perez-Espona *et al.* (2008). In cases where individuals are assigned to less clusters than the estimated K they will still group with individuals that best fit the criteria of the algorithm and will be the most closely related. As with GENELAND, further substructure can be identified within the identified clusters, therefore the user would set a K value which has been recommended in some cases (Corander *et al.* 2008). However, the user is still left with the dilemma of choosing a K value, in this study clusters identified could differ between models even at very low K values $K = 2, 3, 4$ (data not shown)

The most structured way to demonstrate how populations are clustered is to use a progressive partitioning approach, using a stepwise clustering method of $K = 2$, with $K = 2$ performed on each output at each stage allowing the user to define clusters at different degrees of genetic differentiation. This progressive partitioning approach gives the user a structured format to identify clusters sequentially, starting at the partition with the greatest genetic differentiation and ending with the clusters with the lowest levels of genetic differentiation. The results of the progressive partitioning approach confirmed the hypotheses derived during the analysis of the partitions for the estimated optimal K values, that the Bayesian clustering techniques are identifying population units with varying degrees of genetic differentiation between them. When Bayesian clustering techniques are run to estimate the optimal K they are actually identifying population groupings with differing degrees of genetic differentiation between them, for example BAPS4 SPATIAL identifies sub-regional units but also regional units which are later shown to contain further sub-structure in the progressive partitioning method. This is a remnant of forcing these techniques to estimate K from complicated wild population structures made up of regions and sub-regions each with different degrees of gene flow between them. The progressive partitioning method allows the techniques to identify the major partitions first and then further substructure within those partitions, in addition the progressive partitioning approach also means that clusters identified later on in the process with

little genetic differentiation between them are not influenced by genotypes from other populations.

The results of the progressive partitioning method should however, be interpreted with caution as the programs were not designed with this approach in mind and it ignores the relative likelihood of the outcome, however in some cases it has already been recommended that the *K* value used be set by the user instead of letting the algorithm learn the value under a given upper bound *K* (Corander *et al.* 2008). The use of the estimated optimal *K* value for each Bayesian clustering technique alone restricts the ability to compare and combine the results. The results of progressive partitioning appear to be informative and give a higher agreement between programs (using non-spatial and spatial priors) than the optimal *K* approach allowing partitions to be compared and combined to give an insight into the population genetic structure identifying regions and sub-regions.

Clustering algorithms are a tool for identifying patterns and should not be interpreted as the endpoint. The use of multiple methods builds an accurate picture not possible with one method alone. Progressive partitioning helps interpret why the Bayesian clustering techniques identify different values of *K* and why even when the same *K* is identified the clusters found can differ. This is because the clusters identified are dependent on the algorithm of each Bayesian clustering technique. Progressive partitioning is a structured and robust approach removing the error associated with estimating *K* and allowing the interpretation of the partitions that are produced at multiple levels of genetic differentiation, thus allowing the combination and comparison of each Bayesian clustering technique.

Not all the clusters identified by progressive partitioning were accepted because they did not have agreement between more than one method, or were represented by [1-5] individuals. They should not be ignored however, as they may also provide much needed information, and help reveal the differences between softwares. For example, the Central England region is sub-divided repeatedly by BAPS4 SPATIAL during progressive partitioning, perhaps because BAPS4 is sensitive to factors such as, changes in allele frequencies due to introductions, high numbers of unique alleles and small sample size. GENELAND SPATIAL partitions data with a strong spatial prior

ideal for finding population boundaries but in so doing looses information about admixture events, the use of GENELAND SPATIAL alone in the progressive partitioning method would have resulted in the North Yorkshire and Irish sample cluster being missed.

Progressive partitioning is a powerful tool but needs to be interpreted with caution, as the more you split the populations the more chance you have of finding partitions along a cline (Frantz *et al.* 2009). However, the use of progressive partitioning may help control for this feature. The average individual population assignment to each partition at $K = 2$ can be plotted; this graphical display gives an idea of how much admixture there is between clusters. Since STRUCTURE has no priors and best demonstrates this effect and has been used to produce plots of individual assignment to a population at each partition $K = 2$ during the progressive partitioning process (Appendix 3.2). At greater degrees of genetic differentiation (early partitions) the partitions are more distinct, i.e. individuals are assigned to one population or another with little/ no admixture; at lower levels of genetic differentiation there is increased admixture between clusters. Therefore if there is no admixture present the researcher can be confident that there is little possibility of the clusters being identified along a cline and produced as a result of IBD, if there is admixture present this could be a sign of IBD or as a result of restricted gene flow between populations (Explained further in the Appendix text accompanying Appendix 3.2). How much factors, such as IBD and landscape restrictions to gene flow are affecting the sub-regions is a matter of debate and each sub-region should be investigated separately. The progressive partitioning method provides an extra tool to investigate these phenomena and have a greater understanding of the clusters produced using Bayesian clustering techniques.

Whilst advocating progressive partitioning, the standard approach using likelihood values to identify optimal clusters (K) provides an important comparison, as this utilises the whole data set at once. Again with the knowledge of the idiosyncrasies of each program important information can be identified. For example, STRUCTURE appears to be unable to identify individuals that come from un-sourced or under represented populations (a feature also found in Chapter 2). This is because it needs to create ancestral populations from the genotypes and allele frequencies available; it

needs a large sample of individuals to create these ancestral populations, and any individuals from non sourced populations will be assigned to the next best cluster. GENELAND and BAPS4 SPATIAL use different algorithms and can identify individuals from underrepresented populations. This is important when looking for migrants if the population of origin has not been sampled. BAPS4 SPATIAL however may be too sensitive in identifying these individuals as it continues to separate the Central England region into 13 clusters many of which are made up of 1 or 2 individuals. This does not occur in the known strongholds and this may be a feature of the changes in allele frequencies or unique alleles as result of reinforcement in this area.

Populations derived *a priori* from known population history were included to account for any bias in the identification of the clusters by the Bayesian clustering techniques. Identification of samples using Bayesian clustering techniques is based on identifying populations in HWE and can result in uneven sampling sizes and may bias the genetic summary statistics, to control for this evenly sampled populations based on population history alone were also analysed.

The regions identified by Bayesian clustering techniques reflected populations chosen *a priori,* thus allowing the comparison of the genetic statistics. Summary genetic statistics showed the same pattern from both the methods, although the pairwise F_{ST} values between regions derived from the Bayesian Clustering techniques were slightly inflated over those based on population history. These results suggest that the genetic summary statistics for clusters derived from Bayesian clustering techniques can be used to draw conclusions on the otter population structure.

3.5.3. Conclusions

The best practice when using Bayesian Clustering techniques is to use more than one method. Comparison of the methods should be made and agreement between programs is desirable to increase robustness of the results. Where there is conflict in the identification of populations the use of a progressive partitioning method to identify structuring and sub-structuring within the population is recommended. Progressive partitioning using Bayesian clustering techniques also provides a tool to investigate the strength of partitions and the plots of individual population

assignment provide one method to detect whether the clustering algorithms are likely to be identifying clusters along a cline (IBD effect).

Despite the possible bias of creating populations based on HWE, traditional summary statistics can still be used to make conclusions although care should be taken as F_{ST} values may be slightly inflated.

3.5.4. Conservation implications for the UK otter population

It appears that dispersal from otter strongholds is limited. For the viability of the otter population in the UK gene flow between sub-populations is desirable, but evidence suggests that this is limited despite the highly mobile nature of this species. Dispersal may be limited by a number of factors which may be extrinsic (e.g. environmental barriers, unsuitable habitat) or intrinsic (e.g. reluctance to disperse into areas with an unfamiliar prey base).

Whilst there appears to be some gene flow between sub-regions within regions, the regions are isolated from one another. If this continues isolated regions such as Wales and borders and Southwest England could continue to feel the effects of drift and lose genetic variability if gene flow is not established. Genetic variability should be monitored over time, and if deemed appropriate, gene-flow might be enhanced by provision of mitigation schemes, or by translocations. In order to establish what mitigation would be suitable, further studies should attempt to identify barriers to dispersal. With doubt over the origins of some introductions, genotypes from un-sourced Scottish and other European populations should be analysed to identify their origin.

Chapter 4

Landscape Features Affecting Gene Flow between Otters in the Wales and Borders Region

4.1. Abstract

In this Chapter exploratory analysis was used to assess the influence of several landscape features, natural and manmade, on otter dispersal in the Wales and Borders region. ARCVIEW GIS was used with the PATHMATRIX extension to create resistance-to-movement surfaces and correlated the genetic distance with the 'effective distance' between 216 individuals throughout the Wales and Borders region. Effective distances were created from a range of resistance values for each of the landscape variables. The results showed that otter dispersal in Wales is heavily influenced by slope and upland habitats. Whilst otters are not limited by geographical distance in Wales the sub-structuring identified in Chapter 3 was a result of Isolation by Effective Distance (IBED) with steep slopes and unproductive upland habitat in the mountainous areas in the centre of Wales (the Cambrian and Brecon Beacons mountain ranges) acting as a permeable barrier. In addition the large urban settlements in the south and southeast of Wales appear to be acting as a barrier, this finding is reinforced by the lack of positive sites of otters in these areas in a recent otter survey. This area of landscape genetics, which quantifies the effects of landscape features on dispersal, is in its infancy; recommendations are made on how to improve on this pioneering area of research.

4.2. Introduction

Four regional populations were identified in the UK dataset using Bayesian clustering techniques (Chapter 3). These regions are separated by large areas currently un-occupied by otters and have no contemporary gene flow between them. It is important to link fragmented populations to ensure the genetic integrity of the populations by allowing gene flow between them, since the isolation of populations can have detrimental demographic and genetic effects (Couvet 2002).

4.2.1. Otter movements

Surveys have shown that the otter population is expanding and is once again found in historically occupied areas (Strachen & Jefferies 1996; Jones & Jones 2004). Some of this recolonisation is a result of reintroductions and some is a result of the expansion of natural populations. It is important for the management of wild animals to monitor their movements and to understand their landscape usage to enable the identification of possible barriers (both contemporary and historical) that may limit dispersal (Rosenberg *et al.* 1995; Manel *et al.* 2003, 2005).

Due to the crepuscular and nocturnal nature of otters, they are a difficult species to monitor. The semi-aquatic lifestyle of otters requires that they live in proximity to water; as a result, they inhabit approximately linear habitats and territories that follow the water's edge. The energetic costs associated with feeding primarily in water can mean that otters spend three quarters or more of their time on land (Durbin 1998), so they also require suitable terrestrial breeding and resting sites, commonly referred to as holts or couches. These may be tunnels under waterside trees, or more open 'nests' in dense vegetation such as reed beds. Otter spraints provide a tool to identify otter presence and even abundance, however they cannot reveal much about otter demographics or movement patterns of individuals (Mason & Macdonald 2004). Other methods can be used, for example radio telemetry, however this requires the trapping of individuals, which may be problematic due to the low capture rate, small population sizes, potential for injuries caused by handling and is illegal without a licence due to the species' status as endangered (Mills *et al.* 2000). Genetic analysis allows the identity of individuals to be characterised, providing an abundance of information (Chanin 2003; Dallas *et al.* 2003; Huang *et al.* 2005).

The movements of otters is poorly understood, radio telemetry studies have identified large home range sizes (38.8 ± 23.4 km) (Green *et al.* 1984) to a maximum of 84km (Chanin 2003) and the ability to disperse 40 km in a day (Durbin 1993) along the river course. Otters have been shown to move overland (Kruuk 2006) but the role of landscape features as potential corridors or barriers to dispersal are unknown.

There is an apparently continuous distribution of otters within the regions identified in Chapter 3 with no obvious barriers to dispersal; each region however, does show genetic sub-structuring. The Wales and Borders otter region is an example of the complexity of gene flow within populations found in nature. Populations are frequently divided into subpopulations that are connected by differing degrees of gene flow (Perez-Espona *et al.* 2008). Natural populations occur in a landscape mosaic in which environmental features restrict or promote movement and dispersal of individuals and this influences the distribution of genetic variation within a population (Taylor *et al.* 1993; Storfer *et al.* 2007; Perez-Espona *et al.* 2008; Wang *et al.* 2009). Many studies have considered habitat to be a mosaic of suitable and unsuitable areas (Danielson & Hubbard 2000; Coulon *et al.* 2004), however, it is more likely that animals perceive landscapes as a gradient of varying quality and resistance to movement (McIntyre & Barret 1992; Manning *et al.* 2004; Cushman *et al.* 2006; McGarical *et al.* 2009). The identification of how genetic variation within and between populations relates to landscape features will allow the evaluation of how these landscape features affect the movement of organisms, informing conservation and management practices (Crandall *et al.* 2000; Banks *et al.* 2005).

Landscape genetics combines landscape ecology with molecular genetics; it enables the spatial mapping of genetic data such as genetic distances and individual population assignments (as derived from Bayesian clustering, for example) and the potential correlation with landscape or environmental features (Manel *et al.* 2003; Berthier *et al.* 2005). Being able to visualise how two populations are distributed may allow for the identification of cryptic genetic discontinuities (barriers to gene flow) across geographic features and can reveal incidences of secondary contact between previously isolated populations (Manel *et al.* 2003).

4.2.2. Landscape connectivity and effective distance

Landscape genetics incorporates ideas and techniques from landscape and molecular ecology. Geographical Information Systems (GIS) can be used to test informative hypotheses concerning the effect of landscape structure on the movement of organisms and how organisms perceive habitat connectivity (Holderegger & Wagner 2006). Habitat connectivity depends not only on landscape structure but also on the mobility of the organism (Adriaensen *et al.* 2003). In combination, these factors give rise to the concept of landscape connectivity, defined by Taylor *et al.* (1993) as 'the degree to which a landscape facilitates or impedes movement among resource patches'. It is this interaction that may strongly shape evolutionary processes by affecting dispersal and thereby effective movements (i.e. movement followed by successful reproduction), which drive gene flow across a landscape (Coulon *et al.* 2006). Landscape features that influence effective movement can be identified by studying gene flow in relation to landscape structure. Correlations between genetic and geographic distance matrices for individuals and populations using Mantel tests have been used extensively (Manel *et al.* 2003) for example Pogson *et al.* (2001), Diniz-Filho et al. (2008), Pico *et al.* (2008) and Allentoft *et al.* (2009). It is an important statistical tool for identifying 'isolation by distance' (IBD) (Wright 1943) described by Manel *et al.* (2003) "*when genetic differentiation between individuals (or populations) increases with their geographical distance (because gene flow declines at larger distances)*". Spatial autocorrelation methods are used to assess associations between the genetic similarity/difference among pairs of individuals and geographical distance, by testing whether the observed genotype of an individual is dependent on the genotype of a neighbouring individual (Manel *et al.* 2003). A spatial correlogram is used to evaluate the behaviour of autocorrelation as a function of distance (Manel *et al.* 2003). Spatial autocorrelation is able to determine the scale of the spatial pattern, however it cannot identify the location of genetic discontinuities (for example a mountain, river etc) (Manel *et al.* 2003).

Many landscape genetic studies use GIS-based data to incorporate landscape features in order to identify how the type of landscape influences genetic distance, in addition to the influence of linear geographic distance alone (e.g. Broquet *et al.* 2006; Cushman *et al.* 2006; Kozak *et al.* 2008; Perez-Espona *et al.* 2008; Spear *et al.*

2005). The cost to an organism to move across a landscape is termed the effective distance (Verbeylen *et al.* 2003) and can be used to reveal the effect of landscape features on microevolutionary processes in the context of isolation by distance (Ray 2005). To do this a cost is associated with a landscape feature, the magnitude of which depicts how much it impedes or facilitates movement of individuals of that species. This is compared with a measurement of gene flow. Measurements of genetic distance can be identified between populations using Wright's F_{ST} (1951; Perez-Espona *et al.* 2008), and by analyzing the pair-wise estimates of genetic distances between individuals (Rousset 2000) as in Cushman *et al.* (2006).

There are different terms which have been used to describe the cost to an organism to move across a landscape, such as 'landscape connectivity', 'effective distance', 'effective geographic distance' (EGD), 'functional distance', 'least cost path' (LCP) or the inverse 'landscape resistance or isolation' (Adriaensen *et al.* 2003). Recent studies agree on 'effective distance' (Adriaensen *et al.* 2003) to describe the transformed Euclidean distance which accounts for the effect of landscape and behaviour (Adriaensen *et al.* 2003). As in Adriaensen *et al.* (2003) the term 'effective distance' will be used here as the ecological translation for the calculated cost distance.

A least cost approach can be used to incorporate detailed geographical and behavioural information, to derive the effective distance using a cost grid based on assumed habitat value (Coulon *et al.* 2004; Vignieri 2005; Spear *et al.* 2005; Cushman *et al.* 2006; Epps *et al.* 2007). The least cost method has become popular as a result of least cost algorithms based on graph theory (Drielsma *et al.* 2007) becoming available in the toolboxes of the most recent GIS packages such as ArcView v 3.2 (ESRI 2000), ArcMap v 9.2 (ESRI 2007) and Idrisi (ClarkLabs, Worcester, MA, USA).

Cost grids using GIS packages such as ArcView v 3.2 (ESRI 2000) are reclassifications of environmental layers (rasters) based on a cost to movement/resistance attributed to that habitat. The next step is to use the least-cost algorithm, to determine the least cost path between locations (Broquet *et al.* 2006).

The least cost algorithm calculates the path with the lowest cost to movement (least cost path or effective distance) between the source and destination. This path will avoid some elements of the landscape more resistant to movement and preferentially travel through more permeable (lower cost) features, thus minimizing the sum of the 'costs' of every feature crossed on the way (Verbeylen *et al.* 2003; Broquet *et al.* 2006). Cost grids can be measured in the same way as is Euclidean distance (metres) so their effects on gene flow can be correlated with measures of genetic distance in the same way, using Mantel tests (Adriaensen *et al.* 2003).

4.2.3. Land-cover maps; attributing resistance-to-movement values

Digitised land cover maps are available for use in GIS, and can be used to construct cost grids. The land cover maps chosen should best represent the way the study species experiences the environment (Broquet *et al.* 2006).

Resistance values are based on the cost of movement/resistance attributed to the habitats that make up the cell in the cost grid. The resistance value attributed to a landscape feature provides a link between the landscape GIS information and the ecological-behavioural aspects of the study organism (Adriaensen *et al.* 2003).

Different approaches can be used when selecting resistance values. Typically an informed judgement is made based on data already available in the literature (see also Ray *et al.* 2002; Adriaensen *et al.* 2003). In some cases actual field data has been used to estimate resistance values, for example for the Iberian Lynx *(Lynx pardinus* see Ferreras 2001), where it was calculated as the inverse of habitat preference. Species may not however perceive landscapes according to our assumptions concerning connectivity and habitat quality (With *et al.* 1997; Cushman *et al.* 2006). To try and gauge whether appropriate resistance values have been selected, models can be run with different resistance sets (e.g. Adriaensen *et al.* 2003). In species where the effect of landscape features on the movement of species is unknown, allocation of resistance values can be quite arbitrary. Due to this unpredictability Perez-Espona *et al.* (2008) assessed a range of cost cell values (0.0001, 0.0003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 3, 10, 30, 100, 300, 1000, 3000, 10 000, 30 000) for each landscape feature between sampled populations. These values were chosen so that the logarithm of the cost values increased in uniform

steps and can be considered as approximations of the true cost values. Cells without the given landscape feature had a cost to movement of 1. Individual pair-wise genetic and the different effective distance matrices can be compared using Mantel and Partial Mantel tests.

4.2.4. Mantel tests

The Mantel test (Mantel 1967) measures the degree of association between two dissimilarity matrices (Cushman *et al.* 2006) and can compare any genetic distance and geographic distance matrices (Diniz-Filho *et al.* 2008). Where there is a strong spatial IBD effect this can confound the relationship with other landscape features; to overcome this a partial regression on three distance matrices can be tested (Legendre & Fortin 1989).

4.2.5. Partial Mantel tests

Partial Mantel tests can be used to estimate correlations between distance matrices while controlling for the influence of other factors (Smouse *et al.* 1986). For example, Carmichael *et al.* (2001) used Partial Mantel tests to estimate correlations between physical barriers and genetic distance between populations while controlling for the influence of physical distance. Cushman *et al.* (2006) used Partial Mantel tests as a part of a causal modelling framework (Legendre 1993) to assess the support for seven organizational models containing 110 resistance hypotheses. A significant Partial Mantel correlation between the genetic matrix and a cost matrix after removing the effects of geographic distance indicates that a specific landscape-resistance hypothesis is correlated to the genetic structure of the population (Cushman *et al.* 2006).

The objectives of this study are to identify the influence of habitat and landscape features on gene flow and dispersal by correlating the genetic relatedness among individuals with landscape connectivity (measured as the 'effective distance') by the creation of resistance-to-movement surfaces (Manel *et al.* 2003; Coulon *et al.* 2004; Broquet *et al.* 2006; Cushman *et al.* 2006; Storfer *et al.* 2007). Resistance-to-movement surfaces were created from maps of landscape features identified as influencing fish abundance, and features that may directly facilitate or restrict otter movement.

4.2.6. Otter habitat

The land surrounding wetlands is known as the riparian zone. Riparian vegetation can vary from grassland to woodland and is very important for the functional ecology of streams and rivers. The riparian zone is important for fish species, a major source of prey for otters in the UK, as it influences channel morphology and bank stability, providing shade and cover, maintaining water quality, and can provide large wood debris and organic matter (leaf litter) which is fed on by aquatic invertebrates essential for the productivity of the stream food web (Durbin 1998; Richardson 2004). There is increasing evidence that prey abundance is a limiting factor for otters and is a more important determinant of otter habitat than low human disturbance and riparian cover (Sjoasen 1997; Ruiz-Olmo *et al.* 2001). Rivers, like other ecosystems, are not only affected on a fine spatial scale by the immediate habitat but river ecosystems are influenced by climate, topography, geology, with the land use in the surrounding areas strongly influencing the local riparian habitat and biological diversity of streams and rivers (Snelder 2002; Allen 2004). Roth *et al.* (1996) found that measures of land use surrounding riparian vegetation at larger spatial scales were superior predictors of stream ecological integrity than were more local measures.

Generalizations can therefore be made between the effect of the surrounding terrestrial habitat and topography on the productivity of the aquatic habitat and thus suitability to otters. The following habitats and topological features are described because of their significant coverage in the study area and their recognised effect on fish communities and otter ecology.

4.2.7. Broadleaf woodland habitat

Broadleaf trees and shrubs in the riparian zone have been shown to enhance aquatic biodiversity compared to other vegetation types (Broadmeadow & Nisbit 2004), by providing large woody debris (LWD) to rivers which is beneficial to fish and aquatic invertebrates increasing the complexity of the river/stream. They also contribute organic matter to streams (leaf litter) which aquatic invertebrates feed on and are integral to the productivity of stream food webs (Richardson 2004). Some studies suggest that otters have a preference for areas with woody bankside vegetation (Jenkins & Burrows 1980; Mason & Macdonald 1986), providing places to hide, rest

and denning sites. Therefore it is expected that this type of habitat will facilitate otter movement.

4.2.8. Anthropogenic habitat

The impacts of roads in the ecological landscape include habitat loss, fragmentation, and degradation. Mortality due to road traffic accidents is considered one of the biggest threats to otters in the UK (Chanin 2006). Other anthropogenic effects include urbanisation, which generally has a substantial detrimental effect on river ecology (Booth & Jackson 1997; Allen 2004), causing a change in habitat, reduced fish production, invertebrate and algal assemblages, increased pollution (Allen 2004) increased surface run off, and modified banks (Paul & Meyer 2001; Allen 2004). Resting sites will be scarce and disturbance high for otters in urban areas and it is expected that these detrimental effects will increase with greater percentage cover of urban areas.

4.2.9. Upland habitat

Predominantly found in Northern or upland areas, these are high altitude environments which have low temperatures and low nutrient availability (Ward 1998), therefore upland areas are expected to be less productive than lower altitude environments. Upland freshwater habitats in Wales are also recovering from acidification caused by acidic precipitation largely through fossil fuel combustion (Kowalik *et al.* 2007). Acid deposition has had a wide range of impacts on the soil and vegetation (Holden *et al.* 2007), and can lead to the deterioration of the ecology of adjacent streams (Holden *et al.* 2007). The upland moorlands of Wales have also experienced a long history of pastoral management (Yeo & Blackstock 2002), with almost 44 million sheep in the British Isles in 1993 including 11 million in Wales. In Wales upward of 88% of the sheep population graze on upland and hill areas where farming is difficult (Sansom 1999). Increased numbers of sheep have led to overgrazing and increased surface runoff; this can cause weakening of river banks by increasing flow and increasing the erosive power of floods (Sansom 1999). This has resulted in accelerated bank erosion, causing wider and shallower river channels, redistribution of cobbles and gravel and causing a decline in habitat heterogeneity and fish populations, particularly salmonids (Environment Agency 1998). Intensive grazing also weakens the ability of the native vegetation to resist acidification and

may lead to the formation of acid grassland. As a result of the low nutrient status and the detrimental effects of management of these areas on river ecology, it is expected that this habitat will have a high cost to movement for otters.

4.2.10. Slope

Otter movements have been linked to slope in previous studies. White *et al.* (2003) found an association between stream gradient and sprainting activity of otters whilst Janssens *et al.* (2008) found that the slope of the water divide between catchments could act to impede the colonisation ability of otters. They both suggested that otters take the route of least effort, and would avoid steeper slopes. Slope also has a negative effect on some aspects of river ecology, for example increased flow reduces channel morphology and fish abundance in that area. In the anthropological literature it has been hypothesised that the relationship between an increase in slope and movement effort is exponential (Chapman 2003), placing a more realistic emphasis on flatter areas, and greater avoidance of steeper slopes (Chapman 2003), it is expected that slope will have some cost to movement for otters.

The creation of resistance-to-movement surfaces, effective distance matrices and the correlation with genetic distance to identify landscape features that impede or facilitate gene flow is a new and emerging field in landscape genetics. Few studies have used such approaches (Ray 2005; Cushman *et al.* 2006; Perez-Espona *et al.* 2008; Wang *et al.* 2009).

4.2.11. Aims

The aim of this chapter is to explore this emerging area of landscape genetics, and combine various approaches in the literature to identify landscape features that influence otter genetic structure and hence gene flow and dispersal. In Chapter 3 Bayesian clustering algorithms were used to identify sub-structuring within the UK otter population. Four regions were identified and further sub-structuring was also found within these regions. In the Wales and Borders region three sub-regions were identified and this region was selected to identify if the Bayesian clustering techniques were actually detecting population clusters that are shaped by landscape features or just detecting spurious clusters along a cline, IBD effect. Correlations between landscape features and gene flow are tested at three spatial scales: within

sub-regions, between sub-region pairings and within the whole region. Otters are expected to move predominantly along water ways and avoid steep slopes (Jassens *et al.* 2007). Anthropogenic features such as roads and urban areas as well as the unproductive upland habitat are expected to restrict otter movement, while the areas of broadleaf woodland are expected to enhance otter dispersal. It is also predicted that there will be no isolation by distance effect found within sub-regions and a weak correlation if any of the landscape features with genetic distance within these sub-regions.

4.3. Methods

4.3.1. Study area and genetic sampling

The study focused on the Wales and Borders region identified by (Chapter 3). Three spatial scales were chosen to identify movement within and between populations. 214 individuals used in Chapter 2 were chosen to represent an unbiased sample of the Wales and Borders region (Chapter 2; figure 2.2). Sub-regions were identified by hierarchical Bayesian clustering (Chapter 3; figure 3.11) and comprised 91 individuals for the Southwest Wales sub-region, 24 individuals from the Northwest Wales sub-region and 115 individuals from the Mid-Eastern Wales sub-region. In total seven population groupings were identified, (1) Sub-region, Southwest Wales (2) Sub-region, Northwest Wales, (3) Sub-region, Mid-Eastern Wales, (4) Combined Sub-regions 1 and 2, (5) Combined Sub-regions 2 and 3, (6) Combined Sub-regions 1 and 3. (7) Wales and Borders region.

Wales is located in central-west Great Britain. Much of Wales's diverse landscape is mountainous, particularly in the north and central regions. The study area also includes the bordering counties in England (Hereford, Gloucester, Chepstow, Shrewsbury) which are much less mountainous. The total area used for this analysis was 223km by 232km (51,736km^2).

4.3.2. Genetic distance matrix

GENALEX 6 (Peakall & Smouse 2006) was used to create a pair-wise genetic distance matrix calculated for codominant data following the method of Smouse & Peakall (1999). The complete microsatellite dataset (216 individuals and 15 loci) was used with the linear genetic function selected as recommended by the authors when creating a genetic distance matrix for use with Mantel tests. Genetic distance matrices were produced for the seven population groupings.

4.3.3. Geographic distance matrix

A linear geographic distance matrix was created with GENALEX using the individual XY coordinates. A geographic distance matrix was also constructed using a least-cost distance matrix from a resistance-to-movement surface where all cells were given a cost of 1, this represents the Euclidean (straight line) distance between

individuals. The two matrices both measure the straight line distance between individuals, and were compared in order to test the performance of PATHMATRIX; distances were highly correlated ($r = 0.95, p < 0.001$).

When testing the null hypothesis of isolation by distance, the linear relationship between pair-wise genetic distances and the logarithm of geographic distances is characterized (Rousset 1997, 2000), however the author of PATHMATRIX concedes that the benefit of using the logarithmic scale for a cost distance matrix is unknown (Ray 2005). The arithmetic measures of cost distance were used for this analysis to compare cost distance and genetic distance results and the arithmetic measure of Euclidean distance was also used in keeping with this method. Preliminary analyses showed that correlations with genetic distance were similar for both log and arithmetic distance measures (not shown) while these measures were also highly correlated with each other. PATHMATRIX was used to produce an effective distance matrix for each resistance-to-movement surface, for all population groupings (1-7).

4.3.4. Isolation by distance analysis

Mantel tests of matrix correspondence have been widely used in population genetics to examine microevolutionary processes, such as isolation-by-distance (IBD) (Telles & Diniz-Filho *et al.* 2005). Simple Mantel tests (Mantel 1967; Smouse *et al.* 1986) were used to identify correlations between genetic distance and geographical distance using the software MantelTester (Bonnet & Peer 2002) which uses the Program zt (Bonnet & Van de Peer 2002). The program zt uses the Pearson's correlation coefficient as a measure of the correlation between the matrices.

4.3.5. Spatial autocorrelation

Spatial autocorrelation analysis was conducted using the software GenAlEx version 6 (Peakall & Smouse 2006) which uses pairwise geographic and genetic distance matrices to calculate an autocorrelation coefficient r at various geographical size classes (Peakall *et al.* 1995; Smouse & Peakall 1999). The autocorrelation coefficient provides a measure of the genetic similarity between pairs of individuals whose geographic separation falls within the specified distance class. The 'global' spatial autocorrelation method of Smouse & Peakall (1999) employs a multivariate

approach to simultaneously assess the spatial signal generated by multiple genetic loci.

A simple Mantel test was performed for all effective distance matrices produced from each resistance-to-movement surfaces with the genetic distance matrix. This was conducted for all of the population groupings (1-7).

Where *p* values were significant for the simple Mantel test between a genetic distance matrix and an effective distance matrix, a Partial Mantel test (Mantel 1967; Smouse *et al.* 1986) was performed with the geographic distance matrix partialled out. Where *p* values for Partial Mantel tests remained significant (<0.05) there is a significant correlation between the landscape feature and the genetic distance, the *r* values produced represent the genetic differentiation explained by that landscape feature. Here the rationale adopted from Perez-Espona *et al.* (2008) is used, that the cost value with the highest *r* value best reflects this relationship between the landscape feature and gene flow.

Funk (2005), Cushman *et al.* (2006), Frantz *et al.* (2006) all use Mantel and Partial Mantel tests to find correlations between genetic distances and landscape features. Partial Mantel tests are controversial due to potential underestimation of type I error (Raufaste & Rousset 2001; Rousset 2002). Castellano & Balletto (2002) however, argue that this concern has been overstated (Epps *et al.* 2007). As ecologists collect more data, the probability of finding some spurious results that are significant by chance is quite high (Moran 2003), and it has become standard in ecology to use a Bonferroni-type correction to reduce the probability of type I error (Verhoeven *et al.* 2005). This approach has however been subject to criticism as being overly conservative and increases the risk of type II error (Moran 2003; Nakagawa 2004). The application of the Bonferroni correction to multiple groups of data can be inconsistent and manipulated. In this analysis there are 5 landscape features with up to 11 resistance values tested for several population groupings. A Bonferroni correction could be applied per landscape feature, per resistance value, or per population. If applied over the whole dataset (over 300 experimental runs) a very small $p <$ value would be required to show a significant result. The appropriate threshold to declare a test statistic's *p* value significant becomes complex when more

than one test is performed (Verhoeven *et al.* 2005). For this reason and the arguments made in Moran (2003), Garcia (2004) and Nakagawa (2004), a result will be deemed significant based on the typically used $p < 0.05$ cut-off and the interpretations will be based on this, while being aware of the risk of type I errors.

4.3.6. Effect of landscape features on otter populations

Resistance-to-movement surfaces for anthropogenic and natural landscape features were produced using ArcView v 3.2 (ESRI 2000). Landscape features were identified from the Countryside Information System (CIS) version 8 geographical database application (http://www.cis-web.org.uk/). A land cover map representing land use in the UK has been constructed from a computer classification of satellite images (LCM2000). Cover for each landscape feature is measured as a percentage of each $1km^2$ calculated from the Ordnance Survey's 1998 digital 1:250,000 Strategi dataset. Slope was measured per km^2 calculated from the Ordnance Survey's digital 1:50,000 Panorama dataset (OS95 (CIS v6)).

Landscape features were selected based on evidence from the literature on their putative effect on river ecology, fish communities and otter ecology, and include terrestrial habitats (rivers and open water, broadleaf woodland, anthropogenic factors, upland) and topographical feature, slope.

4.3.7. Land cover resistance-to-movement surfaces

Landscape features were classified as facilitating or resisting otter dispersal, based on evidence of otter preference or through the effect on limiting or enhancing their prey availability. A summary of landscape features, resistance model used, source location and classification criteria are in Table 4.1.

Table 4.1. Land features used to create resistance-to-movement surfaces, facilitation/ resistance, model(s) used to create values for resistance surfaces, source of landscape feature data and CIS description of landscape feature.

Landscape feature	Resistance	Model used	CIS dataset	CIS classification
Broadleaf woodland	Facilitate	1.Linear 2.Arbritrary 3.Categorical	Broadleaf forest dataset from Land Cover Map 2000 (CIS v6)	Stands of native broadleaved trees (such as oak, ash and beech), non-native broadleaved trees (such as sycamore and horse chestnut), and yew trees, where the percentage cover of these trees in the stand exceeds 20% of the total cover of the trees present.
Rivers	Facilitate	1.Linear 2.Arbritrary 3.Categorical	Rivers and open water datasets from the (Rivers Ordnance Survey: Geographic Data 98 (CIS v6))	Rivers and open water
Anthropogenic factors	Restrict	4.Linear 5.Non-linear ^2 7.Arbritrary	A Roads, B Roads, Minor Roads, Motorway and built up datasets from Ordnance Survey: Geographic Data 98 (CIS v6))	A roads, B roads, minor roads, motorway and built up areas and gardens
Upland	Restrict	4.Linear 5.Non-linear ^2 6. Non-linear ^3 7.Arbritrary	Upland dataset from Land Cover Map 2000 (CIS v6)	An amalgamation of four Broad Habitats (Dwarf Shrub Heath, Bog, Montane and Inland Rock) to produce one category with characteristics predominantly found in northern or upland areas.
Slope	Restrict	4.Linear 5.Non-linear ^2 6. Non-linear ^3	Slope dataset from the Ordinance Survey (OS: Altitude and Slope Data:1995)	Percent Slope

A value was provided for the percentage coverage of each 1km square cell of the study area of the landscape feature in each dataset.

** where datasets were combined, the percentage coverage of the landscape features of each dataset were added together to form a combined percentage coverage of each cell.*

Cells containing landscape features of interest were given resistance values of <1 where they are thought to facilitate otter movement, or >1 where they are thought to impede otter movement (methods modified from Perez-Espona *et al.* 2008); cells which did not contain the landscape feature were assigned resistance values of 1.

4.3.8. Resistance models

4.3.8.1. Model 1: Linear, facilitate movement

This model assumes a linear function where resistance decreases with increasing percentage of landscape feature in the cell. To test this Equation 4.1 was used to reclassify the percentage scores of the landscape feature in each cell to produce a resistance-to-movement surface.

Equation 4.1.

Resistance = 1 - ((the percentage of landscape feature in each cell)/ 100)

4.3.8.2. Model 2: Arbitrary values, facilitate movement

This model tests the assumption that otters only need a small amount of low resistance habitat to allow movement through a cell. Each cell was given the same arbitrary resistance value if it contains any amount of the landscape feature (0.1% - 100%).

Nine arbitrary resistance values that were <1 (0.0001, 0.003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3) were used to create resistance-to-movement surfaces that facilitate movement as in Perez-Espona *et al.* (2008).

4.3.8.3. Model 3: Categorical, facilitate movement

This model tests the assumption that as the area of the landscape feature that facilitates movement increases the resistance-to-movement decreases, however there may not be a linear function present. For example, small amounts of landscape feature may have a proportionally greater effect on facilitating movement through a cell than areas with larger values. This third model places greater weight on smaller values in a cell to facilitate otter movement while still allowing larger areas of habitat to have a greater facilitating effect. The resistance values reclassify the percentage landscape feature as categorical functions (Table 4.2) to create a resistance-to-movement surface.

Table 4.2: reclassification of landscape features that facilitate otter movement

Percentage landscape feature	Resistance
0 - <0.1	**1**
0.1≤ - <1	**0.75**
1≤ -<10	**0.5**
10≤ -100	**0.25**

4.3.8.4. Model 4: Linear, restricts movement

This model assumes a linear function where resistance increases with increasing percentage of landscape feature in the cell. To test this, Equation 4.2 was used to reclassify the percentage scores of the landscape feature in each cell to produce a resistance-to-movement surface.

Equation 4.2.

Resistance = 1 + (% of landscape feature in each cell)

4.3.8.5. Model 5 & 6: Non-linear, restricts movement

These models assume a non-linear function where resistance increases non linearly with increasing percentage of landscape feature in the cell. To test this Equations 4.3 & 4.4 were used to reclassify the percentage scores of the landscape feature in each cell to produce resistance-to-movement surfaces for models 5 (landscape feature squared) & 6 (landscape feature cubed), respectively.

Equation 4.3.

Resistance = 1 + ((% of landscape feature in each cell)$^{\wedge 2}$)

Equation 4.4.

Resistance = 1 + ((% of landscape feature in each cell)$^{\wedge 3}$)

4.3.8.6. Model 7: Arbitrary-linear, restricts movement

Here, a range of 9 arbitrary resistance values > 1 are compared. Maximum resistance (e.g. where 100% of the 1km^2 cell is taken up by the given landscape feature) is set to 3, 10, 30, 100, 300, 1000, 3000, 10000 or 30000, and lower resistances (where <100% percentage of the 1km^2 is taken up by the given landscape feature) are scaled linearly from 1 to the given maximum (see figure 4.1). To test this Equation 4.5 was

used to reclassify the percentage scores of the landscape feature in each cell to produce resistance-to-movement surfaces for each of the arbitrary values.

Equation 4.5.

1 + [((% cover of landscape feature) X (maximum cell cost -1))/ 100]

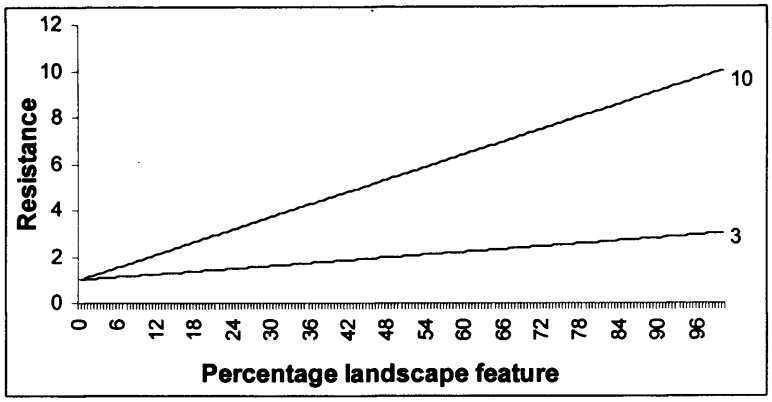


Figure 4.1. Arbitrary resistance values were used to develop hypotheses regarding resistance due to landscape features that were likely to impede movement. [*3*] = maximum cell cost (100% landscape feature cover) is 3; [*10*] = maximum cell cost (100% landscape feature cover) is 10.

4.3.9. Cost Grids: Resistance-to-movement surfaces

The resistance-to-movement surfaces were created (as above) for each of the landscape features. Resistance-to-movement values were calculated for each 1km^2, from which raster layers were created in ArcView v 3.2 (ESRI 2000). These were converted into grid layers using the convert to GRID function in ArcView v 3.2. Each cell in the resulting cost grid provides a value of the cost/resistance to the individual of moving through that cell. Shapefiles mapping all otter locations were also created in ArcView v 3.2. Seven shape-files were created for each of the population groupings 1-7. This structure enabled comparisons between genetic and effective distance at three spatial scales (Region, Sub-region, and paired Sub-regions)

4.3.10. Calculating the effective distances (least cost paths)

PATHMATRIX Ver. 1.0 (*Environmental Science Research Institute, Redlands, USA*) extension (Ray 2005) is based on the cost distance algorithm implemented in the ArcView module Spatial Analyst. PATHMATRIX uses the selected cost grid and location shape-files to define the least-cost path between each pair of individuals to create an effective distance matrix of the least cost distance (lcd).

4.4. Results

4.4.1. Isolation by distance analysis

Mantel tests found no evidence of a significant correlation between Euclidean geographical distance and genetic distance within any of the sub-regions (Table 4.3). This result could be expected because these samples were previously identified by Bayesian clustering as single panmictic units. There was, however, a significant correlation, and thus an isolation by distance effect, found at the other spatial scales, both at the regional level and when the sub-regions were paired (Table 4.3).

Table 4.3. Mantel test for geographic (Euclidean) distance and genetic distance matrices for the 7 population groupings.

Population grouping	*r*	*P*
1) Southwest Wales sub-region	0.01	0.4208
2) Northwest Wales sub-region	-0.04	0.3472
3) Mid-Eastern sub-region	0.01	0.3463
4) Southwest (1) and Northwest (2) Wales combined population	**0.12*****	**0.0005**
5) Mid-Eastern (3) and Northwest (2) Wales combined population	**0.07***	**0.0107**
6) Southwest (1) and Mid-Eastern (3) Wales combined population	**0.09*****	**0.0001**
7) Wales and Borders region	**0.08*****	**0.0002**

r = correlation coefficient for Mantel test

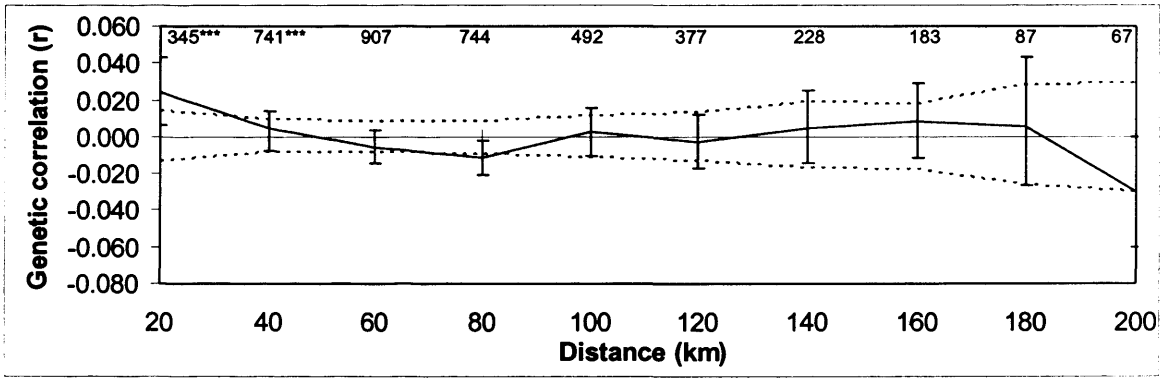
4.4.2. Spatial autocorrelation analysis

Spatial autocorrelation analysis further resolves the scale of the genetic spatial structure for each population grouping. The correlograms (figures 4.2a-g) show the genetic correlation *r* as a function of distance (kilometres) for pairwise comparisons between individual otters, computed using GENALEX (Peakall & Smouse 2006); autocorrelations were produced for a distance class size of 20 km.

The patterns of spatial structuring were similar for all population groupings; within sub-regions there were significant positive *r* values at distance classes 20 and 40km (the Northwest sub-region showed no significant *r* values possibly as a result of the low sample size). At a larger spatial scale when sub-regions were combined and for

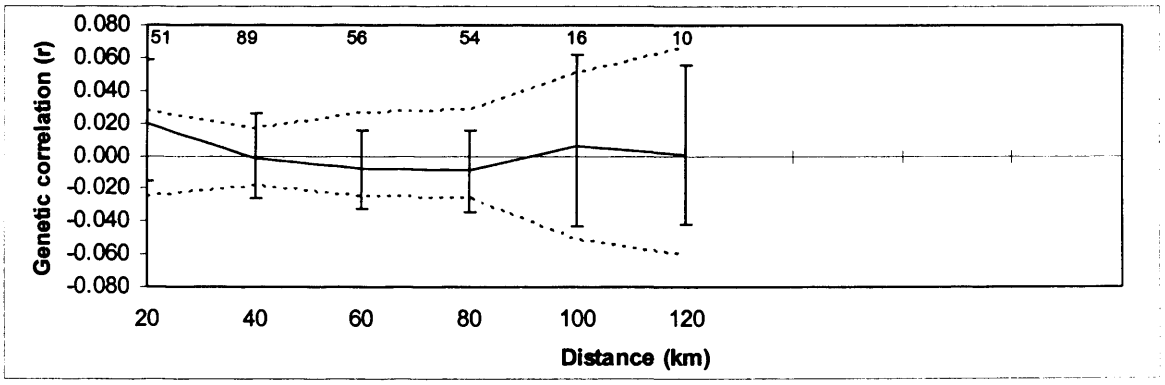
the Wales and Borders region as a whole, there were significant positive *r* values at the distance classes 20, 40 and 60km. The x-intercept provides an estimate of the extent of non-random (positive) genetic structure or neighbourhood size, this value is dependent on the distance size classes and the number of samples in the distance class (Epperson 1990). The first x-intercept ranged from 38.3 to 59km within sub-regions to 70.5-84.2km when sub-regions were combined (figures 4.2a-g).

Figure 4.2a-g: Correlograms for each of the population groupings 1-7. The 95% confidence interval (dashed line) and the bootstrapped 95% confidence error bars are also shown. The numbers of pairwise comparisons within each distance class is shown above the plotted values. Stars indicate statistically significant positive spatial autocorrelation values (**p* <0.05; ***p* <0.01; ****p* <0.001).



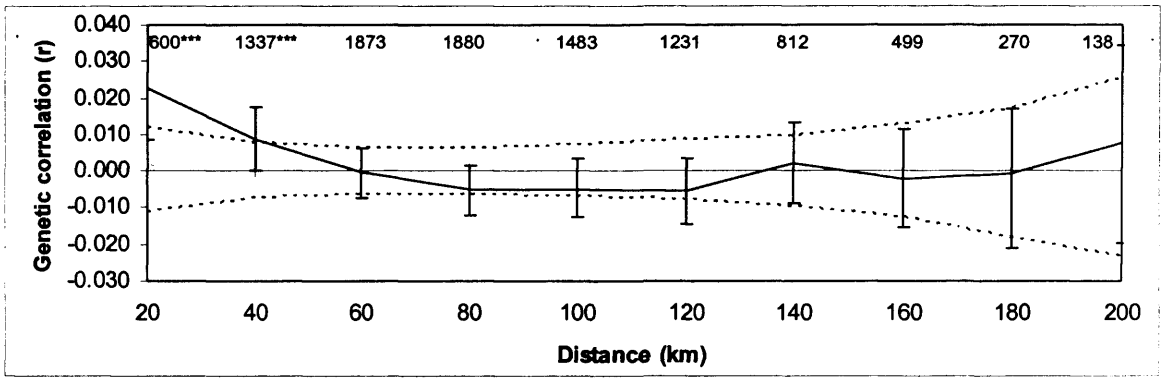
x-intercept 48.9 km

Figure 4.2a. Correlogram for population grouping (1) Southwest Wales sub-region



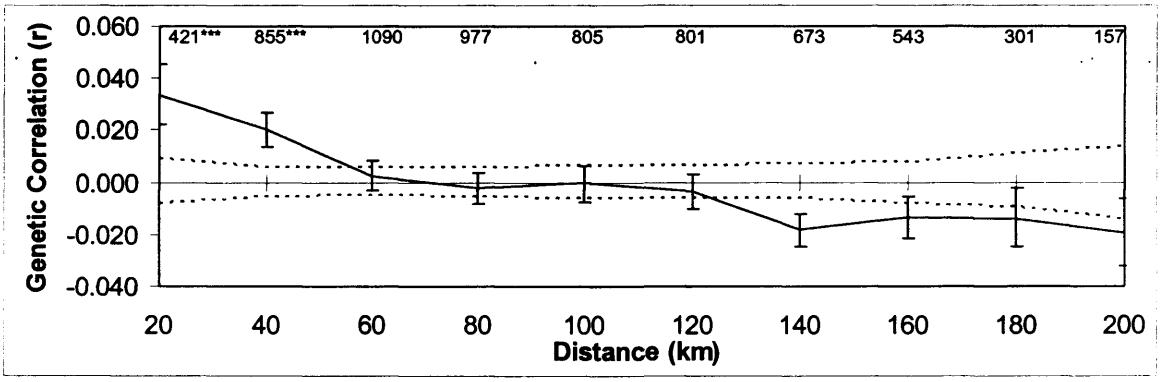
x-intercept 38.3 km

Figure 4.2b. Correlogram for population grouping (2) Northwest Wales sub-region



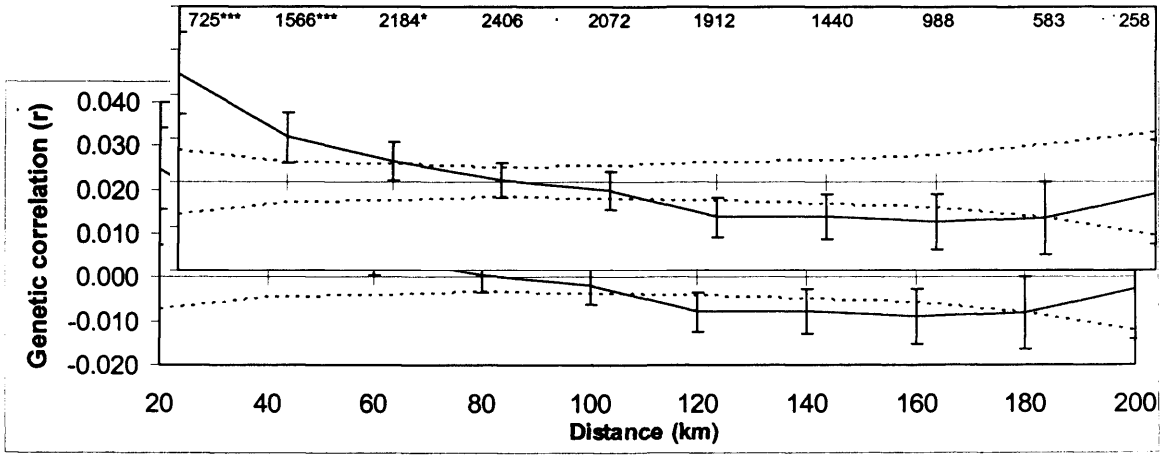
x-intercept 59.0 km

Figure 4.2c. Correlogram for population grouping (3) Mid-Eastern Wales sub-region



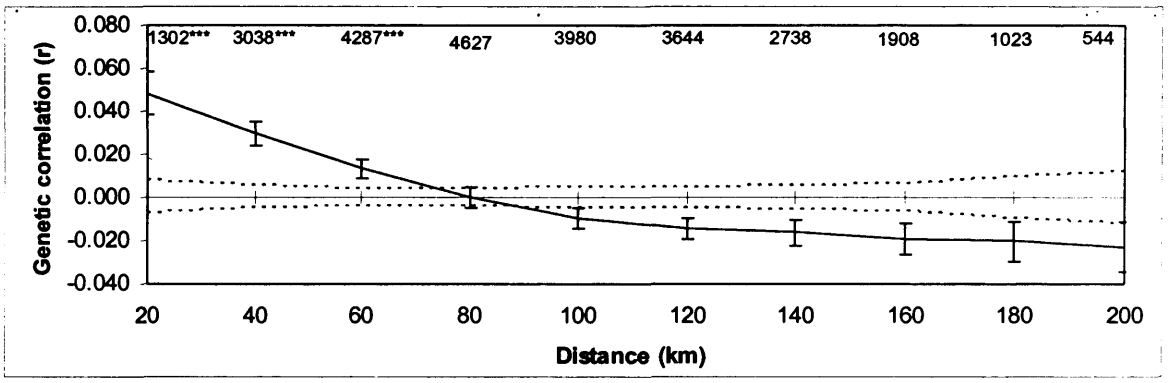
x-intercept 70.5km

Figure 4.2d. Correlogram for population grouping (4) Southwest and Northwest Wales population



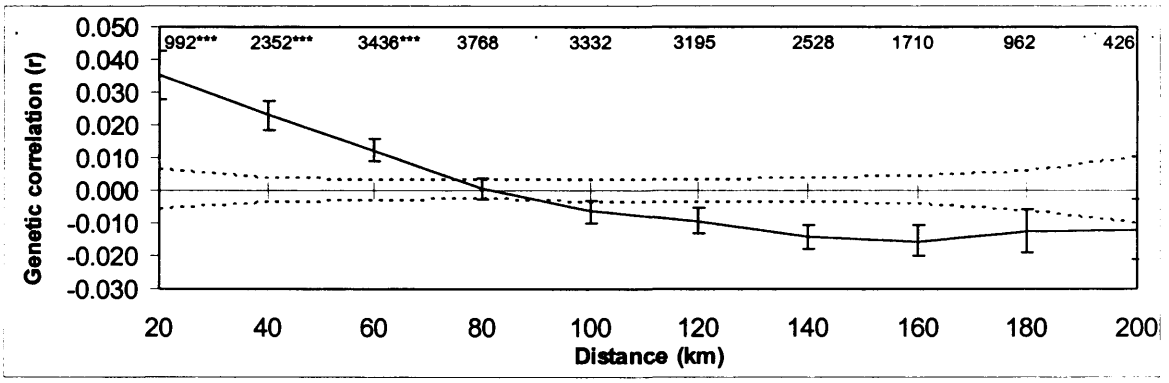
x-intercept 84.2km

Figure 4.2e. Correlogram for population grouping (5) Mid-Eastern and Northwest Wales population



x-intercept 74.9km

Figure 4.2f. Correlogram for population grouping (6) Southwest and Mid-Eastern Wales population



x-intercept 82.2 km

Figure 4.2g. Correlogram for population grouping (7) Wales and Borders region

4.4.3. Effect of landscape features on population structure

Mantel tests were used to identify correlations between genetic distance and the effective distance matrices produced by resistance surfaces of landscape features. Taking into account the fact those individuals that are geographically close can be expected to be genetically relatively similar, Partial Mantel tests were also conducted to control for the this effect of geographic distance (Tables 4.4-4.8). Within each table the correlation coefficients (r) for simple Mantel tests (effective distances with genetic distances) and Partial Mantel tests (geographic distance partialled out) are displayed for each model (that produced the resistance-to-movement surface from which the effective distances were derived), for each population grouping. Asterisks indicate statistically significant correlation values ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$). r values underlined indicate that not only are the genetic and effective distance matrices significant when geographic distance is controlled for ($p < 0.05$), but geographic distance is no longer significantly correlated with genetic distance when the effective distance was controlled for ($p > 0.05$). Habitat maps with the effective distances plotted were produced for the highest significantly correlated r values for each population grouping and landscape feature (Appendix 4.2- 4.15).

4.4.4. Landscape features that facilitate otter movement

4.4.4.1. Broadleaf woodland and rivers

Effective distance matrices produced by resistance-to-movement surfaces for broadleaf woodland and rivers, showed a similar pattern, with significant correlation with genetic distance for four population groupings (Tables 4.4 and 4.5). These population groupings however, also showed a significant correlation with geographic distance, when this was controlled for (Partial Mantel test) only two population groupings showed significant correlations with genetic distance. These population groupings (4 & 5) included the pairing of the Northwest population with another sub-region. This implies that both broadleaf woodland and rivers facilitate the movement of otters within and between these pairs of sub-regions, however they appear not to have the same influence on otter movement elsewhere, particularly within individual regions.

4.4.5. Landscape features that impede otter movement

4.4.5.1. Anthropogenic factors

Effective distance matrices produced by resistance-to-movement surfaces of Anthropogenic factors did not show any significant correlation with genetic distance for any of the population groupings (Table 4.6).

4.4.5.2. Upland habitat

Effective distance matrices produced by resistance-to-movement surfaces of upland habitat showed significant correlation with genetic distance for 5 of 7 population groupings (Table 4.7). It continued to show a significant correlation in 5 of 7 population groupings even when the geographic distance was controlled for. A significant correlation was found at all spatial scales above sub-regional and also for one sub-region [Mid-Eastern (3)]. The models with the highest r value varied between population groupings with model 5 (slope squared) explaining the most variation for population grouping 2 (Mid-Eastern Wales sub-region), while model 7 (X30) explained the most variation for population grouping 4 (Southwest and Northwest Wales). This result suggests that upland habitat is important at restricting otter movement, even within sub-regions where there is no correlation with geographic distance.

4.4.5.3. Slope

Effective distance matrices produced by resistance-to-movement surfaces of slope showed significant correlation with genetic distance for 5 of the 7 population groupings (Table 4.8). When the geographic distance was controlled for, four of the seven population groupings continued to show a significant correlation with genetic distance. For these population groupings linear slope and slope squared showed significant correlation with genetic distance when the effect of geographic distance was accounted for. Especially prominent is that it has a significant correlation within subpopulation 1 (Southwest Wales sub-region) where there is no correlation with geographic distance. These results suggest that slope is important for the restriction of otter movement with the effect on otter movement detectable even within sub-regions.

Table 4.4. The effect of rivers: Mantel tests and Partial Mantel tests.*r* values for effective distances (created from resistance-to-movement models for rivers) and genetic distance matrices for the 7 population groupings. Asterisks indicate statistically significant correlation values (**p* <0.05; ***p* <0.01; ****p* <0.001). *r* values underlined indicate that not only are the genetic and effective distance matrices significant when geographic distance is controlled for (*p* <0.05), but geographic distance is no longer significantly correlated with genetic distance when the effective distance was controlled for (*p* > 0.05).

	Resistance values that Facilitate movement									
	Model 1	**Model 2: Arbitrary cell cost**								**Model 3**
	Linear (%)	**0.0001**	**0.0003**	**0.001**	**0.003**	**0.01**	**0.03**	**0.1**	**0.3**	**Categorical**
1) Southwest Wales sub-region										
Simple Mantel test (genetic distance and effective distance)	0.012	0.014	0.016	0.015	0.016	0.017	0.016	0.018	0.017	0.015
Partial Mantel test (geographic distance partialed out)										
2) Northwest Wales sub-region										
Simple Mantel test (genetic distance and effective distance)	0.016	0.143	0.143	0.143	0.144	0.143	0.097	0.041	0.022	0.000
Partial Mantel test (geographic distance partialed out)										
3) Mid-Eastern sub-region										
Simple Mantel test (genetic distance and effective distance)	0.012	0.050	0.050	0.050	0.048	0.037	0.021	0.012	0.010	0.006
Partial Mantel test (geographic distance partialed out)										
4) Southwest (1) and Northwest (2) Wales combined population										
Simple Mantel test (genetic distance and effective distance)	**0.127**	0.053	0.079	0.079	0.062	**0.116***	**0.116***	**0.141*****	**0.140*****	**0.138*****
Partial Mantel test (geographic distance partialed out)	0.071					0.053	0.053	**0.092***	**0.107***	**0.095***
5)Mid-Eastern (3) and Northwest (2) Wales combined population										
Simple Mantel test (genetic distance and effective distance)	**0.066***	**0.099***	**0.100***	**0.108***	**0.103***	**0.106***	**0.106***	**0.067***	**0.064***	**0.057***
Partial Mantel test (geographic distance partialed out)	**-0.060****	**0.091***	**0.091***	**0.090***	**0.090***	0.078	0.078	-0.013	-0.039	**-0.091****
6)Southwest (1) and Mid-Eastern (3) Wales combined population										
Simple Mantel test (genetic distance and effective distance)	**0.090*****	0.033	0.035	0.039	0.049	**0.071***	**0.081****	**0.081****	**0.083****	**0.084****
Partial Mantel test (geographic distance partialed out)	-0.01239					0.014	-0.005	-0.031	-0.035	-0.03548
7) Wales and Borders region										
Simple Mantel test (genetic distance and effective distance)	**0.074****	0.040	0.050	0.041	0.050	**0.067***	**0.070***	**0.067****	**0.070****	**0.069****
Partial Mantel test (geographic distance partialed out)	-0.018					0.017	-0.006	-0.026	-0.025	-0.028

Table 4.5. The effect of Broadleaf woodland: Mantel tests and Partial Mantel tests *r* values for effective distances (created from resistance-to-movement models for Broadleaf woodland) and genetic distance matrices for the 7 population groupings. Asterisks indicate statistically significant correlation values (*p <0.05; **p <0.01; ***p <0.001). *r* values underlined indicate that not only are the genetic and effective distance matrices significant when geographic distance is controlled for (p <0.05), but geographic distance is no longer significantly correlated with genetic distance when the effective distance was controlled for (p > 0.05).

	Resistance values that Facilitate movement									
	Model 1	**Model 2: Arbitrary cell cost**								**Model 3**
	Linear (%)	**0.0001**	**0.0003**	**0.001**	**0.003**	**0.01**	**0.03**	**0.1**	**0.3**	**Categorical**
1) Southwest Wales sub-region										
Simple Mantel test (genetic distance and effective distance)	0.013	-0.016	-0.006	-0.013	-0.006	0.009	0.016	0.017	0.015	0.013
Partial Mantel test (geographic distance partialed out)										
2) Northwest Wales sub-region										
Simple Mantel test (genetic distance and effective distance)	-0.027	0.005	0.023	0.011	0.023	0.058	0.034	-0.006	-0.024	-0.024
Partial Mantel test (geographic distance partialed out)										
3) Mid-Eastern sub-region										
Simple Mantel test (genetic distance and effective distance)	0.010	0.071	0.070	0.071	0.070	0.058	0.041	0.030	0.023	0.018
Partial Mantel test (geographic distance partialed out)										
4) Southwest (1) and Northwest (2) Wales combined population										
Simple Mantel test (genetic distance and effective distance)	**0.131*****	-0.033	0.044	-0.007	0.044	**0.137****	**0.149****	**0.142*****	**0.136*****	**0.146*****
Partial Mantel test (geographic distance partialed out)	**0.091***					0.073	**0.099***	**0.116***	**0.117****	**0.132****
5)Mid-Eastern (3) and Northwest (2) Wales combined population										
Simple Mantel test (genetic distance and effective distance)	**0.068***	0.054	0.084	0.065	0.084	**0.117****	**0.115****	**0.097****	**0.085****	**0.084****
Partial Mantel test (geographic distance partialed out)	-0.035					**0.092***	**0.112***	**0.112****	**0.084***	0.059
6)Southwest (1) and Mid-Eastern (3) Wales combined population										
Simple Mantel test (genetic distance and effective distance)	**0.090*****	0.009	0.035	0.018	**0.082****	**0.063***	**0.082****	**0.091*****	**0.091*****	**0.087*****
Partial Mantel test (geographic distance partialed out)	-0.008				-0.004	-0.004	-0.004	0.011	0.007	-0.006
7) Wales and Borders region										
Simple Mantel test (genetic distance and effective distance)	**0.076****	-0.048	-0.001	-0.032	-0.001	0.051	**0.070****	**0.078****	**0.079****	**0.081*****
Partial Mantel test (geographic distance partialed out)	-0.011						-0.009	0.001	-0.001	0.008

Table 4.6. The effect of anthropogenic factors: Mantel tests and Partial Mantel tests *r* values for effective distances (created from resistance-to-movement models for anthropogenic factors) and genetic distance matrices for the 7 population groupings. Asterisks indicate statistically significant correlation values (*p <0.05; **p <0.01; ***p <0.001). *r* values underlined indicate that not only are the genetic and effective distance matrices significant when geographic distance is controlled for (p <0.05), but geographic distance is no longer significantly correlated with genetic distance when the effective distance was controlled for (p > 0.05).

	Resistance values that Restrict movement										
	Model 4	**Model 5**	**Model 7: Arbitrary cell cost**								
	Linear (%)	**Non-linear (%^2)**	**(%x3)**	**(%x10)**	**(%x30)**	**(%x100)**	**(%x300)**	**(%x1000)**	**(%x3000)**	**(%x10000)**	**(%x30000)**
1) Southwest Wales sub-region											
Simple Mantel test (genetic distance and effective distance)	0.011	-0.012	0.007	-0.002	-0.019	-0.042	-0.057	-0.062	-0.064	-0.064	-0.065
Partial Mantel test (geographic distance partialed out)											
2) Northwest Wales sub-region											
Simple Mantel test (genetic distance and effective distance)	-0.020	0.114	-0.028	-0.026	-0.024	-0.020	-0.008	0.000	-0.004	-0.008	-0.009
Partial Mantel test (geographic distance partialed out)											
3) Mid-Eastern sub-region											
Simple Mantel test (genetic distance and effective distance)	-0.006	-0.003	0.013	0.012	0.007	-0.006	-0.020	-0.031	-0.033	-0.034	-0.034
Partial Mantel test (geographic distance partialed out)											
4) Southwest (1) and Northwest (2) Wales combined population											
Simple Mantel test (genetic distance and effective distance)	**0.085***	-0.023	**0.127*****	**0.125*****	**0.115****	**0.086***	0.025	-0.040	-0.069	-0.082	-0.085
Partial Mantel test (geographic distance partialed out)	-0.061		0.069	0.043	-0.005	-0.060					
5)Mid-Eastern (3) and Northwest (2) Wales combined population											
Simple Mantel test (genetic distance and effective distance)	0.032	0.012	**0.068***	**0.064***	0.054	0.032	0.005	-0.018	-0.027	-0.029	-0.030
Partial Mantel test (geographic distance partialed out)			-0.041	-0.067							
6)Southwest (1) and Mid-Eastern (3) Wales combined population											
Simple Mantel test (genetic distance and effective distance)	**0.071****	0.007	**0.09*****	**0.088*****	**0.084*****	**0.072****	0.048	0.025	0.014	0.010	0.009
Partial Mantel test (geographic distance partialed out)	-0.037		-0.012	-0.017	-0.026	-0.037					
7) Wales and Borders region											
Simple Mantel test (genetic distance and effective distance)	0.048	0.002	**0.074****	**0.072****	**0.066****	0.048	0.016	-0.016	-0.027	-0.032	-0.033
Partial Mantel test (geographic distance partialed out)			-0.018	-0.024	-0.034						

Table 4.7. The effect of Upland habitat: Mantel tests and Partial Mantel tests *r* values for effective distances (created from resistance-to-movement models for upland habitat) and genetic distance matrices for the 7 population groupings. Asterisks indicate statistically significant correlation values (*p <0.05; **p <0.01; ***p <0.001). *r* values underlined indicate that not only are the genetic and effective distance matrices significant when geographic distance is controlled for (p <0.05), but geographic distance is no longer significantly correlated with genetic distance when the effective distance was controlled for (p > 0.05).

	Resistance values that Restrict movement											
	Model 4	Model 5	Model 6	Model 7: Arbitrary cell cost								
	Linear (%)	Non-linear (%^2)	Non-linear (%^3)	(%x3)	(%x10)	(%x30)	(%x 100)	(%x 300)	(%x 1000)	(%x 3000)	(%x 10000)	(%x 30000)
1) Southwest Wales sub-region												
Simple Mantel test (genetic distance and effective distance)	0.051	0.086	0.074	0.015	0.020	0.030	0.051	0.078	0.094	0.097	0.092	0.090
Partial Mantel test (geographic distance partialed out)												
2) Northwest Wales sub-region												
Simple Mantel test (genetic distance and effective distance)	-0.073	-0.032	-0.029	-0.029	-0.041	-0.056	-0.072	-0.076	-0.044	-0.025	-0.009	0.000
Partial Mantel test (geographic distance partialed out)												
3) Mid-Eastern sub-region												
Simple Mantel test (genetic distance and effective distance)	0.036	**0.141***	0.168	0.015	0.019	0.024	0.036	0.058	**0.091***	**0.098***	0.093	0.090
Partial Mantel test (geographic distance partialed out)		**0.140***							**0.095***	0.097		
4) Southwest (1) and Northwest (2) Wales combined population												
Simple Mantel test (genetic distance and effective distance)	**0.140*****	0.094	0.069	**0.129*****	**0.132*****	**0.136*****	**0.140*****	**0.132***	**0.123***	**0.112***	0.100	0.096
Partial Mantel test (geographic distance partialed out)	0.081			**0.094***	**0.104***	**0.099***	0.081	0.062	0.071	0.082		
5)Mid-Eastern (3) and Northwest (2) Wales combined population												
Simple Mantel test (genetic distance and effective distance)	**0.111*****	**0.109***	**0.099***	**0.075****	**0.082****	**0.092****	**0.111*****	**0.130*****	**0.144*****	**0.126****	**0.105***	**0.097***
Partial Mantel test (geographic distance partialed out)	**0.116****	**0.098***	**0.098***	0.027	**0.078****	**0.101****	**0.116****	**0.117****	**0.125****	**0.113***	**0.101***	**0.096***
6)Southwest (1) and Mid-Eastern (3) Wales combined population												
Simple Mantel test (genetic distance and effective distance)	**0.107*****	**0.121****	**0.116****	**0.092*****	**0.095*****	**0.098*****	**0.107*****	**0.118*****	**0.072***	**0.072***	0.063	0.061
Partial Mantel test (geographic distance partialed out)	**0.062***	**0.109****	**0.122****	0.010	0.035	0.037	**0.062***	**0.076***	**0.074***	**0.074***		
7) Wales and Borders region												
Simple Mantel test (genetic distance and effective distance)	0.095	**0.091***	**0.080***	**0.077****	**0.080*****	**0.084*****	**0.094*****	**0.105*****	**0.100****	**0.076***	0.060	0.055
Partial Mantel test (geographic distance partialed out)		**0.080***	**0.081***	-0.010	0.000	0.015	0.043	**0.063***	**0.072***	0.064		

Table 4.8. The effect of slope (percent): Mantel tests and Partial Mantel tests *r* values for effective distances (created from resistance-to-movement models for slope) and genetic distance matrices for the 7 population groupings. Asterisks indicate statistically significant correlation values (**p* <0.05; ***p* <0.01; ****p* <0.001). *r* values underlined indicate that not only are the genetic and effective distance matrices significant when geographic distance is controlled for (*p* <0.05), but geographic distance is no longer significantly correlated with genetic distance when the effective distance was controlled for (*p* > 0.05).

	Resistance values that Restrict movement		
	Model 4	**Model 5**	**Model 6**
	Linear (%)	**Non-linear (%^2)**	**Non-linear (%^3)**
1) Southwest Wales sub-region			
Simple Mantel test (genetic distance and effective distance)	0.051	**0.098***	**0.101***
Partial Mantel test (geographic distance partialed out)		**0.103***	**0.102***
2) Northwest Wales sub-region			
Simple Mantel test (genetic distance and effective distance)	-0.081	-0.011	0.031
Partial Mantel test (geographic distance partialed out)			
3) Mid-Eastern sub-region			
Simple Mantel test (genetic distance and effective distance)	0.023	0.046	0.048
Partial Mantel test (geographic distance partialed out)			
4) Southwest (1) and Northwest (2) Wales combined population			
Simple Mantel test (genetic distance and effective distance)	**0.165*****	**0.195*****	**0.200*****
Partial Mantel test (geographic distance partialed out)	**0.128****	**0.171****	**0.166*****
5)Mid-Eastern (3) and Northwest (2) Wales combined population			
Simple Mantel test (genetic distance and effective distance)	**0.101*****	**0.124****	**0.118****
Partial Mantel test (geographic distance partialed out)	**0.071***	**0.112***	**0.102***
6)Southwest (1) and Mid-Eastern (3) Wales combined population			
Simple Mantel test (genetic distance and effective distance)	**0.075***	0.041	**0.101*****
Partial Mantel test (geographic distance partialed out)	0.058		0.048
7) Wales and Borders region			
Simple Mantel test (genetic distance and effective distance)	**0.125*****	**0.137*****	**0.089***
Partial Mantel test (geographic distance partialed out)	**0.093****	**0.125*****	**0.084***

4.5. Discussion

GIS techniques were combined with landscape genetic methodologies and highlighted several landscape features that may influence gene flow within the highly mobile otter population in the Wales and Borders region. The analysis was conducted on three spatial scales, sub-regional level - pair-wise combination of the sub-regions and - regional level (complete Wales and Borders region).

Spatial autocorrelation analysis was used to show genetic structure in relation to distance for all population groupings, the x-intercept has been interpreted as a reflection of the size of the area occupied by related individuals (Epperson 1990). For the otter populations the x-intercept occurred at distances greater than 48km up to 84.2km, with significant spatial genetic structure detectable at 40-60km. This area is comparatively large when compared to the closely related but smaller American mink (*Neovison vison*) an invasive species in the UK which was found to have significant spatial genetic structure ranging from 1-5km in a study area in Scotland, with an x intercept of 20km (Zalewski *et al.* 2009) although it is smaller when compared to an expanding Italian wolf (*Canis lupus)* population which was found to have a significant spatial genetic structure of c230km (Fabbri *et al.* 2007). This is likely to be a reflection of the relatively large dispersal capabilities of the otter.

Despite these large dispersal capabilities of otters sub-structuring exists on a comparatively small spatial scale within the Wales and Borders region. The average distances between individuals within the same sub-region is 41-70km (Appendix 4.1) and the average distances between individuals of different sub-regions is 84-87km (Appendix 4.1). Compared to genetic neighbourhood sizes of up to 60km, otters in Wales have the capability to move between sub-regions in any direction. Despite this there was an Isolation-by-distance (IBD) effect in all population groupings made up of combined sub-regions. Manel *et al.* (2003) describe isolation by distance; "*when genetic differentiation between individuals (or populations) increases with their geographical distance (because gene flow declines at larger distances)*".

As otters are capable of moving between sub-regions and are not restricted by the straight line geographical distance, the IBD effect is a product of two spatially

separate clusters. The description of IBD by Hardy & Vekemans (1999) probably best describes it in the context of population genetics *"the process by which geographically restricted gene flow generates a genetic structure, because random genetic drift is occurring locally"*. The restriction of gene flow is a product of landscape features that are facilitating or restricting gene flow between sub-regions. The landscape features act to increase or decrease the 'effective' distance and this phenomenon may actually be better explained as isolation by effective distance (IBED) which would take into account both the geographic distance and the landscape features. Here IBED is described as *"the process by which gene flow is restricted geographically by landscape features, resulting in locally generated genetic structure due to random genetic drift"*.

To identify the effect of different landscape features resistance-to-movement surfaces were created and effective distance matrices were produced between individuals to test the correlation of landscape features with genetic distance. In many studies landscape features have been used to explain additional variation in genetic distance between populations (Perez-Espona *et al.* 2008; Wang *et al.* 2009) and between individuals (Broquet *et al.* 2006; Cushman *et al.* 2006; Zalewski *et al.* 2009).

4.5.1. Effects of landscape features on the population structure of otters

Freshwater rivers are an essential habitat for otters, however, there was only a significant correlation between rivers and genetic distance for two of seven population groupings, these both included the Northwest Wales sub-region in combination with one of the other sub-regions. Whilst rivers may be an important corridor for dispersal into the Northwest sub-region (located in and around the Snowdonia mountain range) the absence of correlation with rivers for other population groupings may indicate that the otters are capable of traversing a variety of non riparian habitat types during dispersal. It is important to note however, that the network of rivers in our study area is dense and may not be a limiting factor with a high correlation between Euclidean and effective distances for this landscape feature they would not be significant here. This has also been found in other species which have a high dependency on riparian habitat such as the American mink in Scotland (Zalewski *et al.* 2009). It might also reflect the resolution of the maps and the assignment of cost values. Adriaensen *et al.* (2003) recognized the importance of

using high resolution for linear features to create cost grids and the linear nature of rivers assigned as a percentage of a $1km^2$ cell may increase the risk of losing the significant effect.

A similar phenomenon occurs with broadleaf forests as with rivers, showing a significant correlation with population groupings that include the Northwest sub-region. Broadleaf woodland has been recognised as being important for otters by providing areas for denning and resting (Mason & Macdonald 1986) whilst also benefitting river morphology and ecology, thus availability of prey (Broadmeadow & Nisbit 2004; Richardson 2004). Here Broadleaf forests in combination with rivers may provide corridors for dispersal into the mountainous areas of North Wales where there are steep slopes and large areas covered by upland habitat. The results however, do not show a significant effect in facilitating the movement of otters within sub-regions, between the Mid-Eastern and the Southwest Wales sub-regions (6) or throughout the Wales and Borders region (7) as a whole. Analysis of the least cost map (Appendix 4.13) shows that the broadleaf forest was represented in most cells within the distributions of each sub-region. The effective distances show a significant correlation with geographic distances within the sub-regions (not shown); as a result its effect may not be distinguishable from that of geographic distance alone. Its distribution in these areas may not be a limiting factor in otter dispersal. However, these maps also show that this habitat has limited distribution between sub-regions in Southwest (1) and Mid-Eastern (3) Wales, but appear not to have a significant effect of facilitating gene flow between these sub-regions identified in population grouping 7.

4.5.2. Landscape features that restrict gene flow

Urban areas and roads have been identified as impeding dispersal in many species (Forman & Alexander 1998), including otters (Janssens *et al.* 2008) with roads considered the most important cause of death of otters in the UK (Chanin 2006). Anthropogenic factors did not, however, correlate to genetic distance for any of the population groupings.

Many anthropogenic features are recent attributes of the landscape and may not have had enough time to influence population genetic structure through the effects of

genetic drift and mutation (Frantz *et al.* 2009). Recent anthropogenic features have however, been found to be a major barrier to dispersal in species, such as the Iberian lynx (Ferreras *et al.* 2004) and the desert bighorn sheep (*Ovis canadensis nelsoni*) (Epps *et al.* 2005). Thus suggesting anthropogenic barriers can constitute a severe threat to the persistence of naturally fragmented populations

One major motorway exists in Wales, the M4, which is located in the urbanised areas of south Wales; in addition there are numerous duel carriageways (high speed four lane roads) that could be potential barriers. Whilst roads are a major cause of death in many species, in a study of the ecological effects of roads, Forman & Alexander (1998) found that except for a small number of rare species, road kills have minimal effect on population size. Despite the high mortality rate on roads in the UK (Philcox 1999) otters are a highly mobile species to which most roads are not an impenetrable barrier, but do pose a huge risk for individuals.

The sub-regions identified encompass many urban areas and analysis of the maps of effective distances produced for the urban area (resistance-to-movement surface) (Appendix 4.7) show that many urban areas throughout Wales are localised and therefore may not be acting as total barriers. Gula *et al.* (2009) also found that anthropogenic infrastructure did not restrict wolf dispersal. Otters are highly mobile, with neighbourhood sizes of greater than 80km identified from spatial autocorrelation analysis, they are capable of moving through unsuitable habitats, and are therefore capable of circumventing or passing through small towns and villages without significantly affecting population structure on the sub-regional scale. One exception is the large urban development in south/southeast Wales, which appears to divide the Southwest sub-region and the Mid-Eastern sub-region, this is not detected in the effective distance analysis maybe as a result of the lack of samples in and around this area. The lack of samples from this area despite higher density of roads and levels of road use is attributable to the low density and absence of otters; these rivers were historically heavily polluted and it has taken time for fish and subsequently otter numbers to increase (Jones & Jones 2004). Otter numbers are slowly starting to recover in the urban areas of the south (Jones & Jones 2004), as samples become more available further fine scale analysis may be able to identify important routes of dispersal into them and identify the source sub-region.

Using arbitrary cost values for upland areas significant correlations were identified between genetic distance and effective distances for the Mid-Eastern sub-region and all populations combining sub-regions, however the magnitude of the resistance varied between population groupings. Analysis of the least cost maps (Appendix 4.8-4.12) show that upland habitat appears to be located in areas dividing sub-regions.

The upland areas of Wales do not provide an optimal habitat for otters as food is limited. This is caused by a range of factors including altitude (Ruiz-Olmo *et al.* 1998), habitat damage by pastoral agriculture and acidification. Upland areas of Wales have a long history of pastoral management (Yeo & Blackstock 2002), with more than 11 million sheep grazing on upland and hill areas (Sansom 1999). The effects of overgrazing is detrimental to fish populations causing increased surface runoff; causing weakening of river banks by increasing flow and increasing the erosive power of floods (Sansom 1999). This has resulted in accelerated bank erosion, causing wider and shallower river channels, redistribution of cobbles and gravel and causing a decline in habitat heterogeneity and fish population particularly of salmonids (Environment Agency 1998). Intensive grazing also weakens the ability of native vegetation to resist acidification and may lead to the formation of acid grassland. Acidifying pollutants have and continue to be emitted to the atmosphere, leading to acidification of soils and freshwaters, and a loss of biota at all trophic levels (Monteith & Evans 2000). In addition upland areas of Wales have a solid geology consisting of granites and acid igneous rocks and there is little or no buffering capacity, rendering streams and lakes potentially susceptible to acidification (Monteith & Evans 2000), further reducing the carrying capacity of the already unproductive rivers (Mason & Macdonald 1989). This is supported by the absence of otter signs in the Brecon Beacon and Cambrian Mountains in the 2002 otter survey of Wales (Jones & Jones 2004).

The movement between sub-regions may therefore be restricted but not totally inhibited by these unproductive upland areas. A significant effect is also found within the Mid-Eastern sub-region (3), whilst geographic distance is not influencing the population structure (no IBD effect), the upland habitat is affecting the spatial structure (creating an IBED effect) and therefore dispersal within the sub region.

Slope was identified as having a significant effect on gene flow for three of the four population groupings that combined the sub-regions (the Southwest and Mid-Eastern sub-regions combined population grouping (6) had a *p* value of 0.06). In addition slope was also found to have an IBED effect correlating with genetic distance within the Southwest Wales sub-region (1) this may reflect the rugged terrain in south Wales, with many valleys sculpted by glaciers. No significant correlation was found within the other sub-regions, this may be a result of the low sample size (n = 24) in the Northwest sub-region (2) despite being a mountainous area. In the Mid-Eastern sub-region the majority of samples are located on flatter terrain, towards the east of Wales and the English border counties.

Slope has been identified to effect movement of wildlife in many species; Perez-Espona *et al.* (2008) identified that red deer (*Cervus elaphus*) preferred to move along valleys rather than across mountains, Kie *et al.* (2005) found that North American elk (*Cervus elaphus*) were less likely to move between drainages when there were steep slopes. In this study whilst a linear increase in slope was found to be significant, an increase in slope squared explained the most genetic variation for most population groupings, this reflects a more realistic emphasis on flatter areas, and greater avoidance of steeper slopes (Chapman 2003). Previous studies have also found slope to affect otter movement; White *et al.* (2003) found an association between stream gradient and sprainting activity of otters whilst Janssens *et al.* (2008) found that the slope of the water divide between catchments could act to impede the colonisation ability of otters. Slope therefore plays an important role in the movement of otters, forcing otters to take the route of least effort thus avoiding the steeper slopes.

4.5.3. Limitations of the analysis

Landscape genetics is an emerging field (Manel *et al.* 2003) and the analyses used in this chapter should be treated as an exploratory tool to identify the effects of landscape features on gene flow. The chapter took inspiration from the methodologies used in pioneering landscape genetic papers (such as Broquet *et al.* 2006; Cushman *et al.* 2006; Perez-Espona *et al.* 2008) to develop the field further and investigate the complex interaction between animals and their environment. The

effect of using multiple tests, significance of *p* values, and the use of Bonferroni corrections have been addressed above but should not be ignored; anomalies and further interpretation of the results and methodologies are addressed below.

By using multiple arbitrary resistant values for multiple spatial scales it was identified that if a landscape feature significantly correlated with gene flow in one population grouping, it did not necessarily correlate in others. In instances when it did show a significant correlation in both, the arbitrary resistance value that explained the most variation may differ between the population groupings.

Nakagawa & Cuthill (2007) speculate on the appropriateness of using the effective size (*r* value) calculated from two variables if influential covariates are not controlled for, and even suggest that the biological interpretation of the effect size statistic can sometimes be completely wrong if covariates are not considered (Nakagawa & Cuthill 2007). Whilst the effect of geographic distance was controlled for the landscape features were not controlled for against each other. There is the risk of intrinsic correlation between some or all these landscape features, which might confound the effect of a particular landscape feature on population genetic structure. For example, landscape features such as slope, altitude and upland habitat by their nature will be highly correlated, whilst upland habitat and broadleaf woodland may be negatively correlated. The interactions between the landscape features may account for the inconsistencies in the *r* values found between population groupings. Therefore the effect of the landscape feature may be influenced by the other landscape features surrounding it. For example where steep slopes occur the beneficial effect of a river may decrease due to increasing water flow and a reduction in channel morphology (Richards *et al.* 1996; Montgomery & Buffington 1997), having a detrimental effect on the overall river ecology and thus prey abundance (Lopez 2004).

Perez-Espona *et al.* (2008) conducted a similar study to this and identified rivers and inland lochs as landscape features that facilitate gene flow in red deer in the Scottish Highlands, and also found that red deer avoided steep mountain slopes. The correlation with rivers may in this case be a result of its negative correlation with mountain slopes, and on flatter terrain the association with rivers may be lost. This is

an example of the difficulty in identifying the effects of landscape features individually as all landscape features act in synergy. The question remains - are rivers facilitating red deer gene flow or are they principally avoiding steep slopes - both? Regardless of the answer, landscape managers can use this knowledge to aid the population in this area, but it does highlight the importance of taking into account the combined effect of landscape features, and means that great care must be taken in applying the results from one study population in a different area where such correlations may differ.

A landscape feature was only accepted as having a significant effect on otter dispersal when the effect of geographic distance was partialled out. This may not be appropriate in all cases, for otters it was expected that rivers would play a significant role in facilitating gene flow, however, due to the abundance of rivers in Wales the effect of rivers in some cases could not be differentiated from that of Euclidean distance. Therefore rivers may have no influence on otter movement or rivers are so abundant that their effect cannot be identified. As rivers are an important landscape feature for otter dispersal the interpretation of this result must be taken with caution. The opposite is true of landscape features that restrict otter movement: they do not show an intrinsic correlation with geographic distance and therefore if they have true effect on dispersal they are more likely to remain significant even when geographic distance is accounted for.

In cases where the effective distance remained significantly correlated with genetic distance when the geographic distance was accounted for, further investigation was conducted to identify if the geographic distance remained significantly correlated with the genetic distance when the effective distance was accounted for. In some cases previously significant IBD effects were no longer significant (underlined partial test scores in Tables 4.4-4.8), suggesting that there is no significant correlation between genetic and geographic distances. This confirms the hypothesis that population genetic structure of the Wales and Borders population is not shaped by geographic distance and an IBD effect, but rather by the effective distance dependent on landscape features creating an IBED effect.

It was found that the landscape features had different effects for the different population groupings, indicating that the scale of the analysis is important when identifying how landscape features effect animal movements. In this analysis three spatial scales based on population groupings were used. The use of different population groupings demonstrated the importance of looking at different scales. Landscape features that are significant on a regional scale may not be important between regions, and landscape features that effect dispersal between sub-regions may not be significant when looking at a larger regional scale.

Interestingly at the sub-regional level, when there was no IBD effect present, IBED could sometimes be found, and the influence of the landscape features on population structure could still be identified on this scale.

Many habitat types were not included in this analysis (e.g. improved grassland, semi-natural grassland or coniferous woodlands) but they may also be important in facilitating or impeding a dispersing otter. The variability in the optimum resistance attributed to the landscape features may reflect the additional facilitation or resistance by other landscape features present.

4.5.4. Future work

For this study landscape features were identified with the capacity to effect otter movement, however, the complex structures of wildlife populations are a result of the combination of all landscape features. Further study will be required to identify how the combined influence of these landscape features effect otter movement. Least cost path studies that combine landscape features into single resistance hypotheses are extremely time consuming tasks. Wang *et al.* (2009) studied the movements of the California tiger salamander, *Ambystoma californiense;* they combined three habitat types and tested 24,843 least cost path analyses over an area 10km². Cushman *et al.* (2006) used a causal hypothesis and tested 108 landscape resistant surfaces representing the factorial combination of four landscape features to study movement of black bears (*Ursus americanus*) over a 3,000 km² range. The current study area is much bigger than these studies up to (51,736km²); five landscape features have been tested for 7 different population groupings. Further work would be to include all possible landscape features and combine these landscape features to build up an

accurate picture of how landscape features interact to effect the dispersal and gene flow in otters in the Wales and Borders region.

The use of a newly designed simulation approach, CDPOP (Landguth & Cushman 2009) should also be investigated. CDPOP is a simulation approach which predicts the influences of landscape structure on the emergence of spatial patterns in population genetic data as functions of individual-based movement, breeding and dispersal. It also allows the quantification of how landscape resistance affects gene flow patterns, using simulations with different resistance grids (Landguth & Cushman 2009).

4.6. Conclusions

Whilst there are no absolute barriers to dispersal evident, using the sub-structuring identified by Bayesian clustering algorithms allowed the identification of landscape features that were important for otter dispersal. Gene flow between the sub-regions is restricted by both recent anthropogenic and historical landscape features. The industrialised urban areas of the southeast appear to act as a barrier at part of the border between the Southwest and Mid-Eastern Wales sub-regions, while slope appears to effect otter movements through much of the region. In the mountainous areas in the centre of Wales (the Cambrian and Brecon Beacons mountain ranges) the combination of slope with unproductive upland habitat appears to act as a permeable barrier creating an IBED effect, creating genetic sub-division within the Wales and Borders region. The Mountainous areas probably has a low carrying capacity for otters, unable to support bitches raising young, but these areas may allow the passage of transient individuals which are channelled by suboptimal terrain, thus allowing limited gene-flow between the sub-regions.

Chapter 5

General Discussion

5.1. Genetic structure of the UK otter population.

The UK otter populations identified by Bayesian clustering analysis, were as expected, isolated in known otter strongholds; North of England, Wales and Borders, Southwest England, and also Central England where there have been many reintroductions.

The four regional otter populations showed different degrees of genetic variability and all contained evidence of further substructure. There was no gene flow evident between the regional populations, as found by Dallas *et al.* (2002). When compared to levels of genetic diversity in European otter studies found by Randi *et al.* (2003), the Wales and Borders and the Southwest England regions had lower levels of genetic diversity and showed signs of genetic drift as a result of isolation. The North and Central England regions had higher than average levels of genetic diversity and showed signs of gene flow into the population.

The population history of the regions is evident in the genetic data; the Wales and Borders and Southwest regions are isolated and have received little or no gene flow into them, whilst the North England region borders a healthy and genetically diverse Scottish otter population (Dallas *et al.* 2002) not sampled here, and has also been the recipient of otters reintroduced by the Vincent Wildlife Trust (VWT). The Central England region, once a struggling fragmented population on the brink of extinction (Jessop & Cheyne 1992), has been the subject of an intense reintroduction program by the Otter Trust (OT), which augmented the population with otters bred in captivity from locally captured stock (Harrison 1988) and rescued otters from unknown origins.

Conservation management of otter populations is not as straight forward as just increasing the gene flow between the regions. The Welsh population, despite lacking in microsatellite diversity, has particularly high frequencies of rare and novel European mtDNA haplotypes (Stanton *et al.* 2009; see Appendix 6); gene flow into the Wales and Borders population may erode this haplotype with more common types. With questions remaining over the origin of some of the captive bred otters used in reintroductions, the consequences of linking the stronghold regions with a

region containing possibly non-native genetic information are not known. Therefore it is important for management of the whole UK otter population to understand the origin of the Central England region and determine the success of the reintroduced otters.

5.1.1. Contributions of expansions from otter strongholds and reintroductions of captive bred otters to the otter population in Central England

The high degree of genetic variation and number of unique alleles in the Central England region cannot be explained by the hypothesis that a small fragmented population was able to expand from a few founders. Neither does it fit the theory of a small number of founders caught in the wild to start a captive breeding programme for reintroductions. Captive breeding, to bolster effective population size and genetic variation of existing populations can result in detrimental genetic effects in the captive populations, such as: inbreeding depression, the loss of genetic variability via genetic drift, or domestication selection (Snyder *et al.* 1996; Storfer 1999). As a result these genetic problems may be introduced (via gene flow) into natural populations when captive animals are released (Storfer 1999). The OT reintroductions do not appear to have had these negative effects and they have been successful in turning a non viable fragmented population into a genetically diverse expanding population. Addressing the questions that have arisen with respect to the origin of released individuals may help explain the current situation.

The OT maintains that all otters released descend from animals taken from the wild in Britain, or received as orphaned otter cubs taken to help add genetic variation for breeding purposes (Harrison 1988). There are unsubstantiated reports that the OT may have cross bred sub species of *Lutra lutra barang* with European *Lutra lutra* when it housed both species in 1974-76 (Robin 1987), although this idea is disputed by the OT (Harrison 1988).

The OT was reported by Mucci (2008) to have bred two different blood lines: the OT *a* line (known origin of the animals) and *b* line (unknown origin of the individuals). Mucci (2008) state that animals released in the United Kingdom descend from the OT *b* line; with the probable non-European origin of these animals supported by mitochondrial and microsatellite data, with otters from the captive sample from East

Anglia forming a unique cluster away from other European otter populations in their study.

This study found a large number of unique alleles in the Central England region that were not present in the surrounding populations. These were perhaps contributed by OT captive bred otters of unknown origins, but it is also possible that reintroductions have been made by additional 3rd parties or that they are native unique alleles not found in adjoining regions.

Previous investigations however, into the source of captive bred otters by Robin (1987) were met with non-cooperation by the OT. If the success of the Central England population is a result of the reintroductions, and these reintroductions happen to be non native individuals, they will have non-native genotypes. The resulting effects potentially include introduction of genes that are poorly adapted to the local environment, disruption of local patterns of gene interaction, detrimental effects on the ability of a population to respond to future change, or might even result in the inadvertent introduction of a new sister species into the community. Any genetic changes which alter a species' ecological properties could be felt throughout the ecological community having a cascading effect. The conservation implications of this to the otter population are unknown and the area of interface between this expanding otter population and the native stronghold regions should be monitored. Comparison of genetics with that of European samples and historical samples from the area will be able to further inform management decisions.

Future work must also identify why there has been limited expansion away from the otter strongholds into adjacent areas. The identification of landscape features that are potentially preventing dispersal; are they natural or man made? For example hypotheses to be tested include the M5 motorway is a large continuous barrier (Figure 5.1), with rivers that flow east to west under it, therefore dispersal from the Wales and Borders region is restricted by this feature. The identification of the reasons for subdivision occurring within the regions should also be investigated.

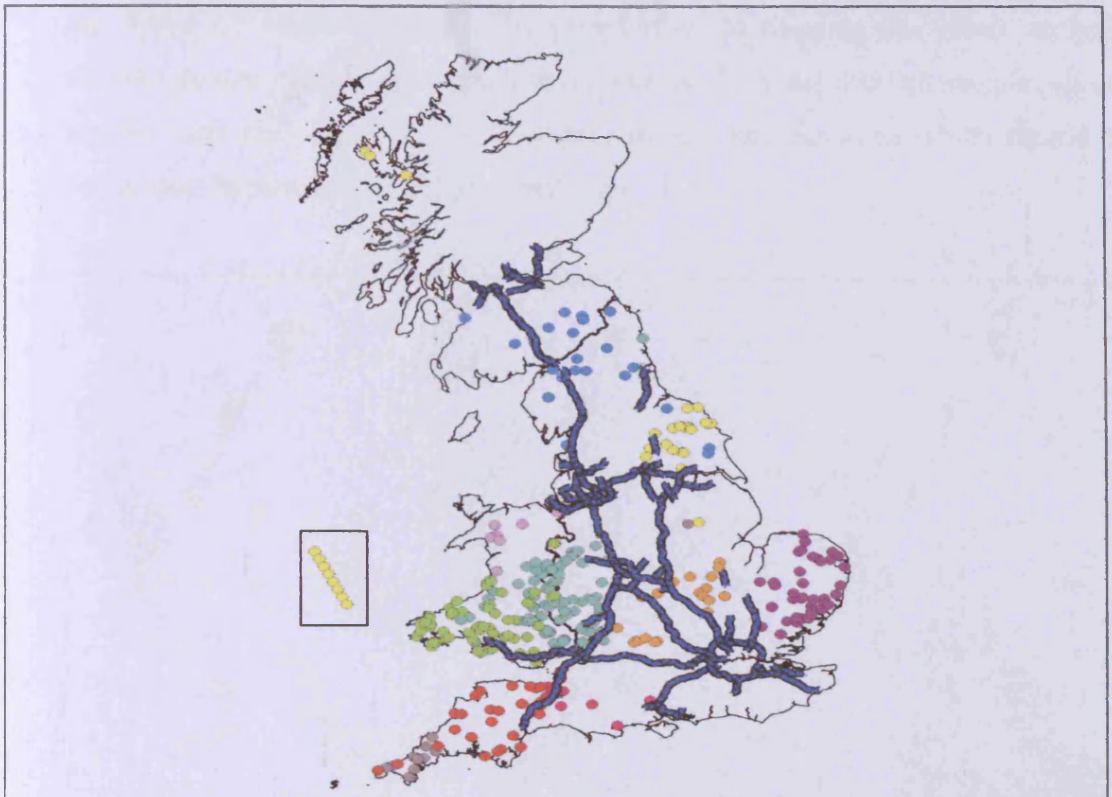


Figure 5.1. Map of the UK highlighting the motorway system blue roads with the 11 sub-regions identified by Progressive and optimal Bayesian clustering methods. Samples from Ireland contained in box on left (not representative of their exact geographic location).

5.1.2. Population substructure within the regions: possible causes, conservation implications and future work

Further population structure was found within regions, with 11 sub-regions identified by Bayesian clustering methods. The Irish and North Yorkshire sub-region (which also includes samples from Western Scotland) shows an unexpected grouping of samples which may be attributable to otter translocations. The Vincent Wildlife Trust provided otters for the reinforcement of a struggling otter population in the Derwent catchments in North East England, UK, between 1990 and 1993. These releases were assumed to be successful according to the monitoring of otter spraints by White *et al.* (2003) over 9 years. Figure 5.2 shows the origin (a) and release sites (b) of otters rehabilitated by the Vincent Wildlife Trust (data courtesy of Rosie Green), which can be compared with the results of Bayesian analysis (c). Individuals found in North Yorkshire and analysed here (Figure 5.2c) are shown to represent all source populations given for rehabilitated otters (Figure 5.2a) This indicates the success of the translocations and explains the grouping of Irish and Western Scottish samples

with those of North Yorkshire in progressive partitioning Bayesian analysis. Interestingly this tentatively suggests that given the Irish and Scottish samples cluster together there may be gene flow between Ireland and Scotland which should be investigated further using a larger sample size.

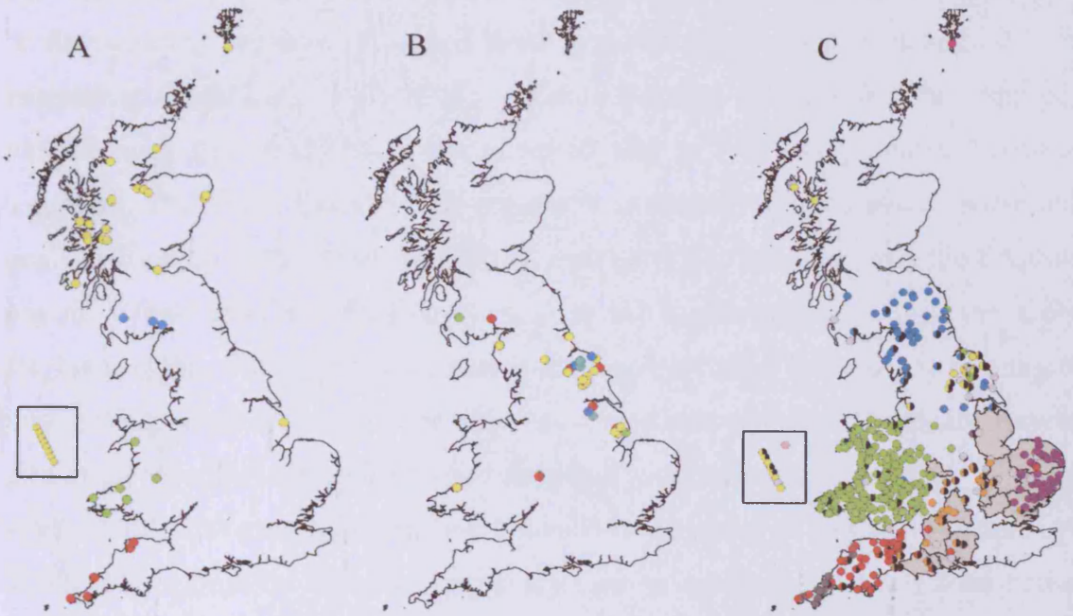


Figure 5.2: Maps showing source locations (a) and areas of reintroduction for rehabilitated otters (b) by the VWT and a map showing individual assignments by the software BAPS4 SPATIAL with areas highlighted that received OT released individuals.

Map C shows the location of individuals, using the colour of their assigned population identified by using the software BAPS4 SPATIAL; A coloured circle was used to indicate individuals with >0.75 assignment to a population. Where two population combined account for >0.75 the individual is plotted half of each population in a half circle. The source location of translocated individuals were identified and plotted with the colour of the population found by Bayesian clustering analysis in their source location (Map A). Using the same colours assigned to them in Map A these individuals were then plotted in their translocated sites (Map B). Records provided courtesy of Rosie green VWT show that 13 released otters originated from north and west Scotland, 3 from Wales, 3 from Northern Ireland and 2 from Southwest England.
(Map C). Highlighted areas indicate counties that received Otter Trust (OT) released individuals.

The Southwest England region also shows further substructure, splitting to form three sub-regions: the Southwest England sub-region, the Southwest peninsula sub-region and the West Country sub-region. Dallas *et al.* (2002) also found an isolated population in the western peninsula of Southwest England which is also home of the OT's Tamar Otter Sanctuary in Launceston Cornwall. This Otter Sanctuary breeds otters in captivity and is situated in the centre of this sub-region. There are no unique

alleles in this sub-region however, if the founding stock were derived from the Southwest region, the levels of genetic diversity are representative of an effect found in supportive breeding where despite no exogenous genes being introduced into the population (Ryman & Laikre 1991), progeny of a few founders routinely released reduces the genetically effective population size and genetic diversity of the population (Ryman & Laikre 1991), if this is occurring the sub-region will be at risk of the potential negative effects of small population size such as genetic drift and inbreeding (Frankham *et al.* 2002). A more detailed analysis will be required to identify why this subdivision has occurred and to look for potential barriers to dispersal. The West Country sub-region is identified in progressive partitioning analysis by both STRUCTURE and GENELAND SPATIAL, splitting from the Southwest region. BAPS4 SPATIAL identifies a similar sub-region splitting from the Central England region. This may reflect a secondary contact zone between the two regions, with mixing of alleles causing the differing allocations of these individuals. Bayesian clustering algorithms have been used to identify secondary contact zones by Durand *et al.* (2009) and even hybridization between two species of lemurs (Pastorini *et al.* 2009). This would be an interesting study case as a potential contact zone between native and possibly non-native introduced otters.

The Central England region split into the East Anglia sub-region and the Oxfordshire sub-region. Figure 5.2c shows the counties where reintroductions have been carried out by the OT. There are no obvious barriers to dispersal between these areas, and neither the history of releases in this area or the genetic data examined provide any obvious explanation for this subdivision. The genetic information in each area will be dependent on the specific introduced animals, but the OT procedure tried to group unrelated animals for release (Jessop & Cheyne 1992) and released into both areas occupied by sub-regions over the same time scale (1983 to 1999), so presumably used the same founding stock, however, the first otter carcasses recovered from these areas were found in 1995/96 during the time of the reintroductions and still showed the same sub-structuring. Progressive partitioning Bayesian clustering indicates little gene flow between the two sub-regions, suggesting isolation, and therefore susceptibility to processes such as drift, mutation or selection. The high genetic diversity shown by the sub-regions may reflect admixture from multiple origins of the founding stock, increasing the allelic diversity (Lefèvre *et al.* 2004). The effect of

the combined evolutionary forces of drift and selection on an introduced population depends on the initial genetic diversity and the environmental conditions (Lefèvre *et al.* 2004). Genetic variation can decrease with time since founding, as a result of drift (Aho *et al.* 2006). Continued genetic monitoring of these populations over time will provide further answers to the apparent success of the reintroductions, formation of sub-regions (drift, selection), origins of the captive bred otters and to help identify the contribution of the native otter population.

An individual from Shetland was also identified as a separate 'population' by two of the Bayesian clustering algorithms in both optimal estimates of *K*, and through progressive partitioning using Bayesian clustering. Surprisingly given its location and distance away from the other clusters it did not have any unique alleles, it did show low levels of genetic diversity (*He* = 0.27) (n = 1), however Dallas *et al.* (2002) also identified the Shetland population as having very low genetic diversity (*He* = 0.26).

The Wales and Borders region showed genetic substructure not detected by the analysis by Dallas *et al.* (2002). Three sub-regions were identified but there are no obvious barriers to movement between them (Chapter 3). The Northwest Wales sub-region (1b) although not identified in Chapter 2 was detected by progressive partitioning using Bayesian clustering and for STRUCTURE *K* = 6 and 9 in Chapter 3. The two sub-regions Southwest (1a) and Mid-Eastern (1c) may represent expansion from population refuges that formed in Southwest Wales and Mid Wales during the sharp population decline in the late 1950-60s attributable to the use of organochlorine insecticides. Alternatively they may be historical population units with limited gene flow between them. This was investigated further in Chapter 4 by exploring the effect of geographical distance and landscape features on gene flow and thus otter movement in the Wales and Borders region.

5.1.3. Further analysis of genetic sub-structuring in the Wales and Borders region.

Despite average distances up to 77.5km between individuals within sub-regions, there was no isolation by distance (IBD) effect reflecting the panmictic units identified by the Bayesian clustering algorithms. When sub-regions were combined

IBD was identified even though the average geographic distances between individuals only increased to 87.5km. Otters are highly mobile species and have been shown to have home range sizes of 38.8 (± 23.4 km) (Green *et al.* 1984) and the ability to disperse 40 km in a day (Durbin 1993). It seems unlikely that the sub-structuring in the Wales and Borders region is due to isolation by geographic distance alone but is heavily influenced by the landscape features that increase the effective distance between individuals from the different sub-regions. Therefore an isolation by effective distance (IBED) effect is created (Figure 5.3) causing sub-structuring within the region. Landscape features have been recognised as facilitating and impeding effective dispersal within natural populations, creating a landscape mosaic which influences the distribution of genetic variation within a population (Taylor *et al.* 1993; Storfer *et al.* 2007; Perez-Espona *et al.* 2008; Wang *et al.* 2009).

Despite the dramatic increase in sites with positive evidence of otters throughout Wales (Jones & Jones 2004), the mobile nature of the otter and short generation time of ~3 years (Pertoldi *et al.* 2001; Randi *et al.* 2003), the population substructure remains evident. The significant correlation with landscape variables suggests that the Welsh population is structured by the landscape. Whilst not acting as complete barriers to dispersal the vast areas of naturally low productivity and steep sloped upland habitat in the Cambrian and Brecon Beacon mountain ranges could be acting to increase resistance to movement. This resistance would have acted on populations irrespective of population declines limiting the amount of gene flow between the sub-regions and thereby leading to genetic differentiation. As the population in the Wales and Borders increases the gene flow between sub-regions may increase due to the increasing pressures to disperse; this might to some degree erode the strength of the differentiation. Alternatively the mountain ranges may form a natural territory boundary and as the Welsh otter population density increases the population subdivision could be enforced with the effect of "home-range pile" up as found by Strasburg (2006). When studying the influence of a highway on mobile carnivores in America, Strasburg (2006) found that home ranges bordering the highway were smaller and showed much greater overlap than those farther away. Strasburg (2006) hypothesised that as a result migrants, especially young migrants entering into these areas would find it difficult to establish and defend a territory and find mates. Similarly as the otter population increases the territories backing up to the

mountainous upland areas may prevent transient animals (also likely to be weaker from dispersal through unproductive areas) from entering into and establishing territories within adjacent sub-regions.

In terms of management, the Welsh otter population appears healthy and is expanding, and anthropological influences do not appear to be contributing to the population structure. However, the low lying route to the south of the mountain ranges which might connect the two southern sub-regions is blocked by highly urbanised areas. Both the Cambrian and Brecon Beacon mountain ranges and the urbanised areas of south Wales have been identified as areas where there are low numbers of otters by the otter survey by Jones & Jones (2004). Therefore in Wales conservation schemes may aspire to protect otter strongholds, and possible habitat corridors around/through upland areas. Management should also focus on mitigating otter dispersal through the urban areas of the south, thus creating a route for natural dispersal. The genetic integrity of the Mid-Eastern and Southwest sub-regions should be monitored and possible translocations between them to aid gene flow may be necessary in the future should one sub-region be suffering effects of isolation.

5.1.4. Future work

The effective management of the otter population depends on continued monitoring and understanding of the dispersal of otters, the identification of landscape features that may impede or facilitate movement, and whether mitigation can be provided against features with high resistance. Hypotheses that could be tested include the M5 motorway is acting as a barrier preventing dispersal from the Wales and Borders region. The Tamar Otter Sanctuary, Launceston Cornwall is associated with the sub-region identified in the Southwest peninsula of England. There should be further investigation into the long term effects of the reintroduced otters into North Yorkshire on the North England region, and into the identification of the source location of the reintroduced individuals into East Anglia, if they are not UK species, the implications for conservation management not only for this region but also for the other UK regions should be sought.

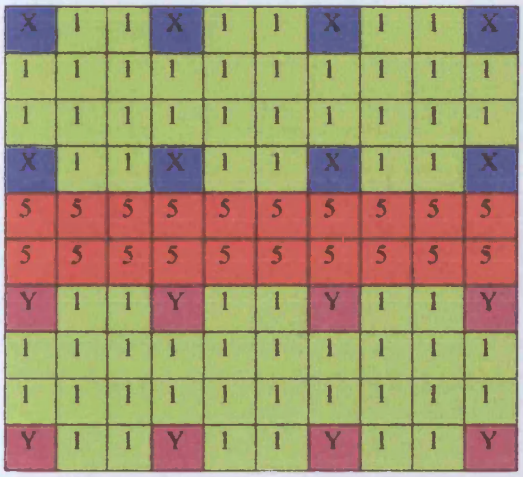
Figure 5.3. Isolation by distance (IBD) or Isolation by effective distance (IBED)?

Figure 5.3 shows individuals continuously distributed across a cost grid. The cost of movement across a cell is displayed within the cell. A cost of 1 unit is attributed as the basic cost of moving in a straight line across that cell (Euclidean geographic distance).

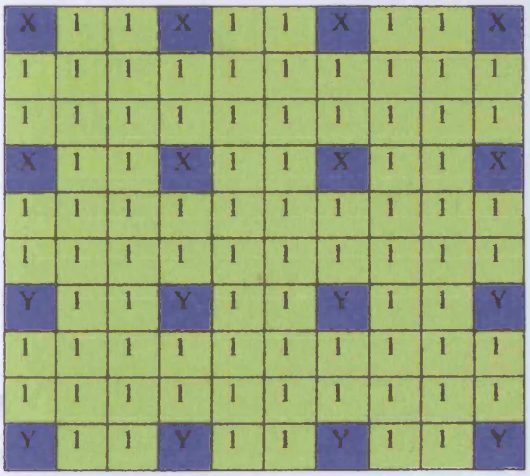
Figure 5.3a represents a single randomly mating panmictic unit made up of both group X and Y with no IBD effect present. In Figure 5.3b a landscape feature increases the cost of moving across the cell where it is present to 5 units. The sum of this cost of movement between individuals taking into account the cost of landscape features is the effective distance. If this species disperses at cost distances greater than 7-10 units very infrequently, then despite being geographically close the individuals of groups X and Y have limited contact and could become genetically differentiated.

Here despite the individuals having exactly the same distribution as Figure 5.3a an IBD effect would be detected due to the fact that individuals of the same cluster are on average spatially closer together than those from the other cluster. However the IBD effect is not a product of the Euclidean geographic distance between individuals. Figure 5.3c represents the spatial distribution for the Euclidean distance to have the same clustering effect as Figure 5.3b, Figure 5.3c would represent a true IBD effect, whereas the scenario in Figure 5.3b is best described as isolation by effective distance (IBED) taking into account the effect of the landscape feature, on this otherwise geographically continuously distributed population.

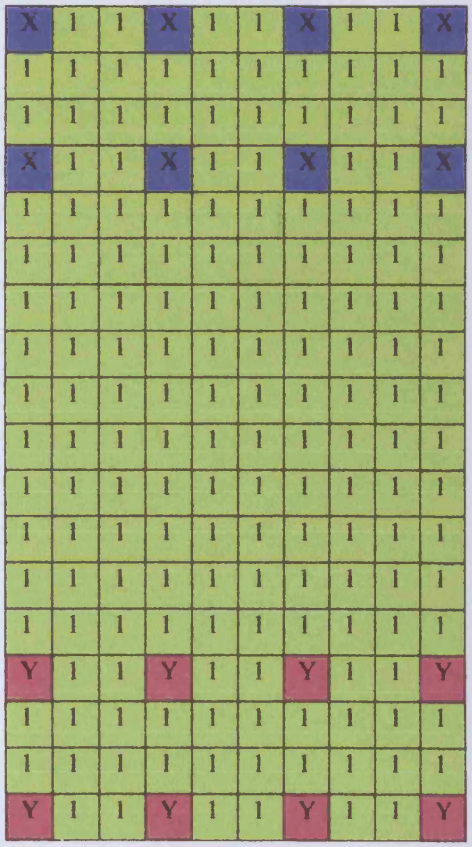
a.



b



c.



5.2. Advances in Landscape genetic techniques -

5.2.1. Bayesian clustering methods

5.2.1.1. Bayesian clustering outputs should be compared and combined

Clustering techniques are tools of exploration (Dubes & Jain 1976) and their interpretation can be subjective, based on the experience and judgement of the user (Dubes & Jain 1976, Jain *et al.* 1999). It is therefore not suitable to use one approach to give a clustering solution (Jain *et al.* 1999) and several clustering programs should be used and compared (Dubes & Jain 1976; Jain *et al.* 1999); when possible the results should be combined (Topchy *et al.* 2003). As the availability of Bayesian clustering techniques has increased landscape genetic studies have begun to use more than one method and compare and combine the results of the algorithms.

5.2.1.2. Difficulties in using more than one Bayesian clustering technique

Bayesian clustering techniques main use has been to estimate the number of genetic partitions or number of populations (K) in a dataset. A challenge identified in this study and others when using these programs is that they can differ in their estimation of the number of populations (K) and can also differ in how they assign individuals to those populations. Differences will occur because the models make assumptions such as the populations are in HWE with no immigration or emigration. Wild animal populations are dynamic and complex and do not always conform to the assumptions of the models. This discussion makes a detailed examination of Bayesian Clustering, using examples from the literature that infer population clusters and provide some guidance to how these models can be used to provide robust reliable solutions.

5.2.1.3. Differing estimates of the number of populations *(K)*.

The difficulty in identifying K is exemplified by the debate about its estimation by STRUCTURE (the most widely used Bayesian clustering method), for which there are two known techniques for estimating K, which can give different values (Pritchard *et al.* 2000; Evanno *et al.* 2005). When the estimation of K differs researchers choose the results of the model that best fits the known biological history of the species and are unable to combine the results of the different algorithms.

There are examples in the literature of studies that try to compare and combine the outputs of multiple Bayesian clustering algorithms, for example Carmichael *et al.* (2007) try to combine the results of STRUCTURE and GENELAND SPATIAL by creating a complicated clustering procedure. To avoid the problem of varying estimates of K they use only STRUCTURE to estimate K (using the method described by Pritchard *et al.* 2000) and use this as an input K value for both models. The estimation of K is however an important step, with a large influence on the resultant assignments, and Carmichael *et al.* (2007) identify this as a limitation in their study and recommend confirming the estimate of K with other Bayesian clustering methods.

5.2.1.4. Same *K* value different populations

A problem that occurs in the literature and in this study is that even when the K value is the same the populations identified by the Bayesian clustering assignments may be different. Therefore, the author is again left with the decision to choose the results of a particular program and question why one model finds further genetic structure within a cluster while others do not. Carmichael *et al.* (2007) were faced with this dilemma; when $K = 7$ they found agreement for a majority of clusters, but a single cluster identified in one software was divided into multiple clusters by the other software and vice versa. Carmichael *et al.* (2007) combined these outputs to assume a new K value of 10 from which they carried out all further study. However, to do this they used a series of assumptions based on the models and even geographical location of samples, assigning individuals from the same location to the dominant cluster at that sampling site. Final population assignments along with assignments from GENELAND and STRUCTURE were supplied in supplementary material and whilst a majority of samples agreed in their assignment they show many samples differed in their assignment between methods and were assigned based on these arbitrary assumptions to different populations. This compromise was followed to enable combination of multiple models and to try and increase the validity of their results, but is a very *ad hoc* method. They recommend the use of another model to estimate K, but they do not recommend a solution to different estimates of K, although the typical response would be to choose the software that is the most biologically plausible.

Coulon *et al.* (2008) use a hierarchical approach to study population genetic structure of Florida scrub jays (*Aphelocoma coerulescens*). They used the rationale given by Evanno *et al.* (2005) that "*this method detects the uppermost level of population structure when several hierarchical levels exist*". Coulon *et al.* (2008) identified *K* for the whole dataset and repeated the analysis on each of the *K* groups inferred. Coulon *et al.* (2008) found that this method infers genetic groups that were often spatially overlapping and for some, defined at a surprisingly fine spatial scale, they do not however, comment further on the results. In addition to their experimental hierarchical analysis they compared the more regular approach of estimating the optimal number of clusters (*K*) for STRUCTURE and GENELAND and comparing the assignments. To take into account that in Bayesian clustering analysis replicate runs may give slightly different solutions (Jakobsson & Rosenberg 2007), they used consensus analysis for 100 runs, using CLUMPP (Jakobsson & Rosenberg 2007) and their own method CONSANA (Coulon *et al.* 2008) using functions of the program R: Ihaka & Gentleman (1996). The CLUMPP approach gives the average assignment value over all runs to the individual, while accounting for label switching between runs. Their CONSANA approach is more complicated in that it not only tries to find the population an individual has the highest inferred ancestry for, but tries to cluster paired individuals that are assigned to the same genetic group in more than X% of the runs. Despite their complicated approach they eventually use the CONSANA approach with GENELAND because it is the most biologically plausible.

Bayesian clustering analysis is an area that is developing, but as recognized by many authors the interpretation of the results is difficult (Coulon *et al.* 2008, Carmichael *et al.* 2007) and most authors still choose their population samples based on biological plausibility. In addition these procedures are complex and use a number of time consuming steps to produce numerous replicates.

5.2.1.5. Investigation into the idiosyncrasies of Bayesian clustering techniques

In Chapters 2 and 3, the aim was to better understand the idiosyncrasies of Bayesian clustering methodologies using a large, wild animal georeferenced data set to identify a procedure that would allow the comparison and then combination of the results of different Bayesian Clustering softwares.

In Chapter 2 softwares differed in their estimate of *K*, and this could be a result of just a few unique individuals from unsourced populations. The softwares STRUCTURE and PARTITION could not identify these individuals whilst BAPS and GENELAND could assign them to their own populations. With these individuals removed 5 of 7 of the models produced the same estimate of *K* ($K = 2$), and the clustering patterns resulting from the interpolation maps (Figure 2.2 Chapter 2) were remarkably similar between the methods. However, with a 0.75 assignment threshold, population assignment only agreed among all the models for 18% of the individuals, although the models showed consistency in their assignment over multiple runs. This was an important factor - even when the models use the same *K* they may assign individuals differently, and the more clustering solutions that are sought the greater the likelihood of differences between programs.

In Chapter 3 the Bayesian clustering methods judged to have performed well in Chapter 2 (STRUCTURE (Pritchard *et al.* 2000), BAPS4 SPATIAL (Corander *et al.* 2004) and GENELAND SPATIAL (Guillot *et al.* 2005a)) were used to conduct analysis of the UK data set. The same difficulties described above were also encountered with different estimates of *K* and where *K* was the same different clusterings of individuals were identified.

5.2.2. Identification of population groupings at different scales of genetic differentiation

Taking the hierarchical analysis approach used by Coulon *et al.* (2008) it was found that the clusters identified by GENELAND SPATIAL could be further divided to produce populations similar to those identified by BAPS4 SPATIAL (Appendix 3.1). This indicates that these models were identifying population groupings at different scales of genetic differentiation.

5.2.3. Development of novel progressive partitioning approach

In response to the variability in clustering solutions a novel progressive partitioning approach was used to try to investigate how the populations divided at different scales of genetic differentiation, using the assumption that at lower values of *K* the genetic differentiation between population groups would be greatest. The objective

was to produce a technique that would allow replication of the results and to be able to compare and combine outputs between models. Unlike Coulon *et al.* (2008) the optimal *K* identified by the models was not used but used $K = 2$, allowing the models to identify the greatest degree of genetic differentiation at each step and to identify different levels of genetic differentiation between population clusters. By assuming $K = 2$ for each step the method is kept as simple as possible with the added benefit of removing the need to estimate *K* for all models, thus removing a laborious step especially in the case of STRUCTURE. This method worked extremely well, and allowed the identification of regional population groupings and sub-regions within those regions (Chapter 3), which were not possible to deduce with optimal *K* values only. The use of GIS to visualise assignments was invaluable and allowed comparison of the results and the identification of the clusters at various levels of genetic differentiation. There were some differences in the clusters identified and assignment of individuals, for example STRUCTURE could not identify unsourced individuals, and BAPS4 SPATIAL would further subdivide some clusters based on their unique allele frequencies.

5.2.3.1. Benefits of using a progressive portioning approach

When an estimate of *K* is used the models are forced to find solutions, clusters may be identified which are not at the same degree of genetic differentiation. The use of a progressive partitioning method for Bayesian clustering allows the researcher to identify population clusters at various spatial scales, identifying sub-divisions with the greatest degree of genetic variation first. This method provides a greater degree of information, allowing for a better understanding of the population clusters identified. Another benefit is that the models were able to produce clustering solutions that were similar enough that they could be compared and combined without the need to manipulate the results. This contrasts with Carmichael *et al.* (2007) who were forced to place individuals with conflicting assignment into the cluster according to their geographical locality.

Despite a majority of individuals being assigned to the same clusters by the different models there was some conflict for others, more so at lower degrees of differentiation. At the regional scale there was high agreement between the models with 80.2% of 566 individuals assigned to one of four regions by all of the models; at

a finer spatial scale 58.7% of the original 566 individuals were assigned to the same sub-regions. Conflicting individuals were left out of further analysis; although not ideal this may give a better account of the effective population in these areas. For example those that were not assigned to the population of origin may be transient migrants or translocated individuals which do not reflect local population history. Even though individuals were lost from summary statistics this method provides a simple way to compare and combine the assignments of individuals to populations using various Bayesian Clustering methods.

The spurious identification of populations by Bayesian clustering techniques has been a cause of concern (Frantz *et al.* 2009). The progressive partitioning method provides a structured format on which to interpret the clusters allowing the researcher to identify with confidence population clusters that find agreement between multiple softwares. This method shows that the variance in the number of populations (*K*) identified could be attributed to the fact that each model identifies genetic partitions at various scales of genetic differentiation. By removing the subjective step of identifying *K*, a consensus can be reached between all models to identify the most suitable population groupings whilst having the benefit of identifying genetic partitions at various scales of genetic differentiation adding strength and information to the analysis.

5.2.4. Spurious identification of populations along an isolation by distance gradient 'cline'

Another problem identified by users of Bayesian clustering methods is the identification of clusters when faced with deviations from random mating not caused by genetic discontinuities such as along an isolation by distance (IBD) gradient 'cline' (Frantz *et al.* 2009). It has been a recent matter of debate to decide whether clusters identified by Bayesian algorithms were artificially detected structures emerging from uneven sampling along clines or were actually well-differentiated groups (Serre & Pääbo 2004; Rosenberg *et al.* 2005; Frantz *et al.* 2009). This is an inherent problem with the progressive partitioning method as by its nature the more you split the populations the more chance you have of finding partitions along a cline.

An approach was tested in Chapters 2 & 3 that incorporated and built on ideas from Chen *et al.* (2007) and Sahlsten *et al.* (2008) to detect whether the partitions identified were found along a cline for each step of the progressive partitioning analysis. The average population assignment for each individual was identified (Chen *et al.* 2007) using the software CLUMPP for five runs of each model at $K = 2$. A summary plot of the estimated population assignments for each individual to each cluster was produced (Sahlsten *et al.* 2008). By sorting the plot by the estimated membership coefficient the strength of the genetic partition could be identified. In the early steps (that identify the greatest degree of differentiation) of progressive partitioning analysis, individuals were typically more unambiguously assigned to each cluster, with little or no admixture (Figure 5.4a). At later steps (by definition more closely related clusters) there is more admixture shown in individuals (Figure 5.4b) and the posterior probability of the cluster membership varies quite smoothly over all individuals. Therefore the strength of the genetic subdivision can be identified by this method and gene flow can be represented by the plots.

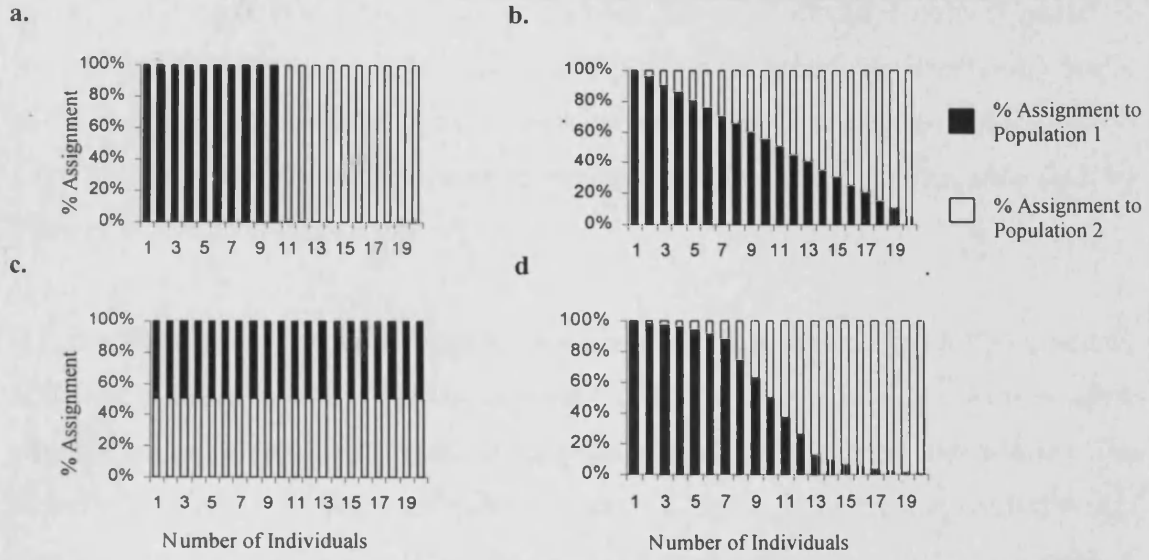


Figure 5.4. Using plots of individual population assignments as indicators of gene flow between populations.

a) Where there is little or no gene flow individuals will show unambiguous assignment and there will be a vertical divide between cluster assignment on the plots; b) As gene flow increases the number of individuals and the amount of admixture will increase, to get a gradual slope in assignment between the plots*; c) When the slope is horizontal and the membership coefficient is 50-50 there are no longer two clusters and only one cluster exists. d) exaggerated variation around the central cluster*. (Interpretation of the slope between vertical and horizontal will depend on the researcher, scale of study and the study organism). *Figures are modified from Chen *et al.* (2007) Figures 5 and 6.

Chen *et al.* (2007) use simulated datasets to test the ability of the models to identify an IBD effect. Using simulated data they displayed the estimates of the membership probabilities along a cline against the membership probability in one population as a function of the location along the cline. Using a nonlinear regression curve they found that STRUCTURE (which is aspatial and is more sensitive to admixture (Carmichael *et al.* 2007)) produced a quasi-linear estimate for the coefficient membership similar to Figure 5.4b, while models such as GENELAND which used spatial priors provided an exaggerated variation around the central cluster as Figure 5.4d.

Plotting the average individual population assignment to each partition at $K = 2$ with graphical display gives an indication of the degree of admixture between clusters. STRUCTURE has no priors therefore its assignment results best take into account gene flow and admixture (Analysis of the preliminary results of the plots for the Bayesian

techniques found that the clusters identified by GENELAND SPATIAL as a result of its strong spatial prior had little or no admixture). The STRUCTURE results (Chapter 3; Appendix 3.2) were as expected; at greater degrees of genetic differentiation (early partitions) the partitions are more distinct, at lower levels of genetic differentiation there is increased admixture between clusters approaching the cline identified by Chen *et al.* (2007) Figure 5.4b.

It is therefore unlikely that the regions are identifying clusters along an IBD gradient, however, at lower levels of genetic differentiation there could be IBD effects or some other factor that is restricting gene flow, such as landscape features (see below). The cause of the sub-divisions may differ between clusters therefore each sub-division should be investigated separately. The progressive partitioning of the data at steps of $K = 2$ provides researchers with an extra tool to investigate these phenomena and have a greater understanding of the differentiation between clusters produced using Bayesian clustering techniques. Further work to investigate the plausibility of this method to truly identify gene flow and possible IBD effects is required, however, the trends found here will provide a basis for advancement.

5.2.5. Identifying the effect of landscapes on gene flow

In nature populations occur in a landscape mosaic in which environmental features restrict or promote movement and dispersal of individuals which will influence the distribution of genetic variation within a population (Taylor *et al.* 1993; Storfer *et al.* 2007; Perez-Espona *et al.* 2008; Wang *et al.* 2009). In Chapter 4 the role that the landscape in Wales and surrounding borders played in the movement of otters was examined.

In Chapter 4 population groupings at 3 spatial scales were used, these spatial scales were based on population groupings created by the identification of population clusters using Bayesian clustering algorithms (Chapter 3). Within the sub-regions there was no correlation between genetic and geographic distance, there was however, significant correlation when sub-regions were combined.

The sampling regime depended on Road Traffic Accidents (RTAs), and although samples were from all over the Wales and Borders area, the interpolation maps show

that the density of samples decreases in areas where sub-regions meet. Therefore an IBD effect could be assumed to be present here, giving rise to artificial clusters. The plots of individual population membership (Appendix 3.2), identify the presence of admixture but do not support an IBD effect. Hardy & Vekemans (1999) state "*isolation by distance occurs within a continuously distributed population, when dispersal of gametes and/or zygotes is spatially restricted or in subdivided populations, when sub- populations exchange genes at a rate dependent upon the distance*". Otters are highly mobile organisms, spatial autocorrelation analysis indicate neighbourhood sizes of 40-80km, therefore it is unlikely that dispersal is spatially restricted especially on the scale of this study (232 X 223km). It is therefore unlikely that the sub-populations exchange genes at a rate dependent on the geographical distance, but rates may be dependent on effective distance. Effective distance is the cost to an organism to move across a landscape (Verbeylen *et al.* 2003) and can be used to reveal the effect of landscape features on microevolutionary processes in the context of isolation by distance (Ray 2005). Therefore, while there may not be an isolation by geographic distance effect the otters may be spatially restricted by landscape features causing an isolation by effective distance (IBED) effect.

Wildlife population structures are complex and linear geographic distances separating populations may have less influence on creating and maintaining genetic structure than features of the environment that affect dispersal (e.g. slope, roads and other climatic and topographical features) (Kozak *et al.* 2008). Organisms perceive the habitat differently (Holderegger & Wagner 2006), and landscape genetic studies have shown animal movement as a function of genetic distance is influenced to a great extent by topography, habitat types etc (Cushman *et al.* 2006; Perez-Espona *et al.* 2008, Kozak *et al.* 2008).

5.2.6. Identifying isolation by effective distance (IBED): The least cost approach

A least cost approach was used to incorporate detailed geographical information and behavioural aspects to derive the effective distance using a cost grid based on assumed habitat value following the studies of Coulon *et al.* (2004), Spear *et al.* (2005), Vignieri (2005), Cushman *et al.* (2006) and Epps *et al.* (2007). Problems

faced when using these approaches are the choice of habitats and the resistance values attributed to those habitats. Following the methods of Perez-Espona *et al.* (2008) arbitrary resistance values and Mantel tests were used to test for correlation between genetic and effective distance matrices. In addition this method was combined with the approach used in other studies (Roach *et al.* 2001; Spear *et al.* 2005; Cushman *et al.* 2006) and only attributed a significant effect to a landscape feature if it was still significant after a Partial Mantel test was performed which takes into account geographical distance

Ths study ascertained that it is important to identify the effect of landscape features on gene flow and dispersal at various spatial scales. It is also important to understand the distribution of the landscape features studied when interpreting the results. For example, otters are semi aquatic and have a dependence on rivers. Due to their abundance in the study area, their effect on gene flow could not be removed from the effect of geographic distance.

A problem in this analysis is the intrinsic correlation between landscape variables which may confound the effects of a particular landscape feature on the population genetic structure (Perez-Espona *et al.* 2008). The effect of a particular landscape feature can be dependant on other landscape features in the surrounding area. As a result the effects of individual variables have to be interpreted with caution or extrapolated from the background noise.

To quantify the effect of one particular variable in isolation from other variables is difficult and may be unrealistic, as Partial Mantel tests do not allow the simultaneous assessment of more than two predictor variables (Carmichael *et al.* 2007). Therefore alternative approaches are necessary.

5.2.7. Alternative approaches to correlate genetic structure to landscape features

5.2.7.1. Distance Based Redundancy Analysis (dbRDA)

An alternative recently applied to population genetic data in wolves is distance based redundancy analysis (dbRDA) (McArdle & Anderson 2001; Geffin *et al.* 2004; Pilot

et al. 2006). The dbRDA allows the user to test up to N-1 predictor variables (N= number of populations) either individually, or fitted in sequence to produce a combined model (Carmichael *et al.* 2007). The conditional tests used in the dbRDA examine the extent to which any of the sets of predictor variables (or their combination) explains genetic diversification among populations over and above that explained by geographical distance alone (Carmichael *et al.* 2007).

The studies which have successfully applied the dbRDA approach have found that it is able to identify populations that have adapted to different environments as the reason for their isolation. Similar to methods used to identifying species distribution, the dbRDA method attempts to identify environmental/landscape factors that are associated specifically with population 1 or population 2. For example, Carmichael *et al.* (2007) found that the genetic structure in wolves was correlated strongly to transitions in habitat type, probably as a result of specialising on prey with different behaviours restricting the differentiated wolf populations to the habitat of their prey. Pilot *et al.* (2006) found a similar result that ecological processes may strongly influence the amount of gene flow between populations as a result of natal-habitat biased dispersal.

dbRDA correlates environmental variables with population assignment and is useful for identifying populations of the same species that are restricted to different habitat types. For example by natal-habitat-biased dispersal where individuals can stay in the pack for an extended period learning to hunt on prey animals and as a result disperse to similar habitats, therefore the opportunity for gene flow between these populations is missed. This method does not lend itself to the ability to produce cost grids or identify effective distances and is not specific enough to be used to explain the genetic distance between individuals.

5.2.7.2. Causal modelling

Another approach is that of causal modelling, as used by Cushman *et al.* (2006). Cushman *et al.* (2006) use least cost modelling, and Partial Mantel tests to test multiple hypotheses of the effect of landscape features and environmental conditions on gene flow between individuals. They produced resistance surfaces based on the factorial combination of four landscape factors they found to influence Black bear

movement: elevation, slope, roads, and land cover. They modelled resistance of these factors (scaled 1-10) to gene flow across four levels for elevation and three levels for the other factors, producing 108 hypotheses. A significant Mantel correlation between the genetic matrix and a cost matrix indicates that a specific landscape-resistance hypothesis is correlated to the genetic structure of the population (Cushman *et al.* 2006).

Cushman *et al.* (2006) were able to use this method as they had previously identified the landscape variables that effected black bear movement, Where (as in the current study) landscape features affecting movement are unknown, an exploratory approach with multiple variables can be used, but is very analytically intensive. Now that the correlation of otter genetic structure with resistance surfaces created from individual landscape features has been tested, future work could combine significant factors to form one resistance surface, for use in a causal hypotheses framework such as that applied by Cushman *et al.* (2006). Whilst this provides an obvious starting point, significant resistance factors were in fact population grouping dependant. The presence or magnitude of the significant correlation between landscape features and gene flow is a result of the interaction of landscape features in that specific area. It may therefore not be realistic to identify the resistance of a particular landscape feature due to the complication of the interactions between all the landscape features.

5.2.8. Future work to identify the effect of landscape features on gene flow

Future work should try to combine the landscape features that impede or facilitate otter movement to provide the best explanation for the genetic variation. This approach would benefit from the ability to produce and run least cost path analysis on sets of resistance surfaces automatically as this is an extremely time consuming task. The production of a computer program that allows multiple hypotheses to be tested automatically will allow for progression in this area of landscape genetics.

An alternative approach could be to use a dbRDA type method that identifies the landscape features associated with population 1, the landscape features associated with areas of highest admixture between population 1 and 2 (or if no admixture the area defining the population boundary) and the landscape features associated with population 2. In this scenario the landscape variables or combination of landscape

variables that are important for the sub-structuring between the two populations may become evident. This would have to be conducted separately for each combination of populations.

5.3. Closing Statement

This thesis was successful in its aim to identify the UK otter population genetic structure. In order to reach its goal it required the use of techniques that are new and largely untested on wild animal population datasets. As a result this study provided an opportunity to investigate these techniques on a wild georeferenced dataset.

Landscape Genetics is a developing field and this study benefitted from and built on recent advances. The approaches used in this study; to combine molecular ecology, Bayesian clustering techniques and the presentation of the data using GIS software provides a powerful toolbox to investigate cryptic genetic structure in wild animal populations.

The techniques and methods conducted in this study (identifying areas of genetic subdivision by comparing and combining the results of Bayesian Clustering techniques; exploring the effects of landscape features on creating/preserving these genetic partitions by the visual identification of barriers or identifying correlations with gene flow) will provide a basis for subsequent analysis performed by the landscape genetic community.

The identification of the otter population structure in the UK into four regional populations, and their subsequent sub-regions will provide a basis for targeted management by conservationists and policy makers and hopefully contribute to the long term survival of one of Britain's most charismatic mammals.

References

Adriaensen F, Chardon JP, De Blust G, Swinnen E, Villalba S, Gulinck H, Matthysen E (2003) The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning,* **64,** 233-247.

Aho T, Rönn J, Piironen J, Bejörklund (2006) Impacts of effective population size on genetic diversity in hatchery Brown trout (*Salmo trutta* L.) populations. *Aquaculture,* **253,** 244-248.

Allen JD (2004) Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology Evolution and Systematics,* **35,** 257-84.

Allentoft ME, Siegismund HR, Briggs L, Andersen LW (2009) Microsatellite analysis of the natterjack toad (*Bufo calamita*) in Denmark: populations are islands in a fragmented landscape. *Conservation Genetics*, **10**, 15–28.

Altmann J, Alberts S, Haines SA, Dubach J, Muruth P, Coote T, Geffen E, Cheeseman DJ, Mututua RS, Saiyalel SN, Wayne RK, Lacy RC, Bruford MW (1996) Behaviour predicts genetic structure in a wild primate group. *Proceedings of the National Academy of Sciences*, **93**, 5797-5801.

Banks SC, Lindenmayer DB, Ward SJ, Taylor AC (2005) The effects of habitat fragmentation via forestry plantation establishment on spatial genotypic structure in the small marsupial carnivore, *Antechinus agilis. Molecular Ecology,* **14,** 1667-1680.

Barbara T, Martinelli G, Palma-Silva C, Fay MF, Mayo S, Lexer C (2009) Genetic relationships and variation in reproductive strategies in four closely related bromeliads adapted to neotropical 'inselbergs': *Alcantarea glaziouana, A. regina, A. geniculata and A. imperialis (Bromeliaceae). Annals of Botany,* **103**, 65-77.

Barbosa AM, Real R, Márquez AL, Rendón MA (2001) Spatial, environmental and human influences on the distribution of otter (*Lutra lutra*) in the Spanish provinces. *Diversity and Distributions*, **7**, 137–144.

Barbosa AM, Real R, Olivero J, Vargas JM (2003) Otter (*Lutra lutra*) distribution modelling at two resolution scales suited to conservation planning in the Iberian. *Biological Conservation*, **114**, 377-387.

Barnett JR, Ruiz-Gutierrez V, Coulon, A, Lovette IJ (2008) Weak genetic structuring indicates ongoing gene flow across White-ruffed Manakin

(*Corapipo altera*) populations in a highly fragmented Costa Rica landscape. *Conservation Genetics,* **9,** 1403-1412.

Basu N, Scheuhammer AM, O'Brien M (2007) Polychlorinated biphenyls, organochlorinated pesticides, and polybrominated diphenyl ethers in the cerebral cortex of wild river otters (*Lontra canadensis*). *Environmental Pollution,* **149,** 25-30.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics,* **5,** 251-261.

Bedford (2009) The effects of riparian habitat quality and biological water quality on the European Otter (*Lutra lutra*) in Devon. *Bioscience Horizons,* **2,** 125-133.

Berthier K, Galan M, Foltete JC, Charbonnel N, Cosson JF (2005) Genetic structure of the cyclic fossorial water vole (*Arvicola terrestris*): landscape and demographic influences. *Molecular Ecology*, **14,** 2861-2871.

Birks JDS (1997) A volunteer-based system for sampling variations in the abundance of polecats *(Mustela putorius). Journal of Zoology London,* **243,** 857-863.

Boitani (2000) Action Plan for the conservation of the wolves (*Canis lupus*) in Europe. Convention on the Conservation of European Wildlife and Natural Habitats (Bern Convention). *Nature and environment,* **113,** © Council of Europe Publishing.

Bonesi, Macdonald (2004) Differential habitat use promotes sustainable coexistence between the specialist otter and the generalist mink. *Oikos,* **106,** 509-519.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13,** 3261-3273.

Bonnet E, de Peer YV (2002) zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **7,** 1-12.

Booth DB, Jackson CR (1997) Urbanisation of aquatic systems: Degradation of thresholds, stormwater detection, and the limits of mitigation. *Journal of the American Water Resources Association*, **33,** 1077-1090.

Botkin DB, Saxe H, Araújo MB, Betts R, Bradshaw RHW, Cedhagen T, Chesson P, Dawson TP, Etterson JR, Faith DP, Ferrier S, Guisan A, Hansen AS, Hilbert DW, Loehle C, Margules C, New M, Sobel M, Stockwell DRB (2007)

Forecasting the Effects of global warming on biodiversity. *BioScience,* **57,** 227-236.

Bradshaw AV, Slater FM (2002) *A post-mortem study of otters (Lutra lutra) in England and Wales.* Environment Agency R & D Technical Report W1-084.

Breitenmoser U (1998) Large predators in the Alps: The fall and rise of man's competitors. *Biological Conservation,* **83,** 279-289.

Broadmeadow S, Nisbit, TR (2004) The effects of riparian forest management on the freshwater environment- a literature review of best practice management. *Hydrology and Earth System Sciences*, **8,** 286-305.

Broquet, Ray, Petit, Fryxell, Burel (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology,* **21,** 877–889.

Bruford MW, Wayne RK (1993) Microsatellites and their application to population genetic studies. *Current opinion in Genetics and Development*, **3**, 939-943.

Bruford MW, Cheesman DJ, Coote T, Green HAA, Haines SA., O'Ryan C, Williams TR (1996) Microsatellite and their approaches to conservation genetics. In: T.B. Smith and R.K. Wayne (eds), *Molecular Genetic Approaches in Conservation.* OUP New York.

Carmichael LE, Krizan J, Nagy JA, Fuglei E, Dumond M, Johnson D, Veitch A, Berteaux D, Strobeck C (2007) Historical and ecological determinants of genetic structure in arctic canids. *Molecular Ecology,* **16,** 3466–3483.

Carmichael LE, Nagy JA, Larter NC, Strobeck C (2001) Prey specialization may influence patterns of gene flow in wolves of the Canadian Northwest. *Molecular Ecology,* **10,** 2787-2798.

Castellano S, Balletto E (2002) Is the partial Mantel test inadequate? *Evolution*, **56,** 1871-1873.

Chadwick EA (2006) *A post mortem study of otters found dead in England and Wales between 1992 and 2003.* Environment Agency R & D Technical Report W1-084.

Chanin P (2006) Otter road casualties. *Hystrix Italian Journal of Mammology,* **17,** 79-90.

Chanin P, Jefferies D (1978) The decline of the otter *Lutra Lutra L.* in Britain: an analvsis of hunting records and discussion of causes. *Biological journal of the Linnean Society*, **10,** 305-328.

Chanin P (1985). The Natural History of Otters, Croom Helm, London, England

Chanin P (2003) *Monitoring the otter (*Lutra lutra*) Conserving Natura 2000 Rivers.* Monitoring Series No. 10, English Nature, Peterborough.

Chapman HP (2003) Rudston 'Cursus A' - Engaging with a Neolithic monument in its landscape setting using GIS. *Oxford Journal of Archaeology,* **22,** 345–356.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes,* **7,** 747-756.

Ciofi C, Beaumont MA, Swingland IR, Bruford MW (1999) Genetic Divergence and units for conservation in the Komodo dragon *Varanus komodoensis. Proceedings of the Royal Society B: Biological Sciences,* **266,** 2269-2274.

Clavero, Prendal, Delibes (2003) Trophic diversity of the otter (*Lutra lutra L.*) in temperate and Mediterranean freshwater habitats. *Journal of Biogeography,* **30,** 761–769.

Conroy JWH, Chanin PRF (2000) The status of the Eurasian otter (*Lutra lutra*) in Europe. A Review. In: Conroy J. W. H, Yoxon P. and Gutleb A.C. (eds). Proceedings of the First Otter Toxicology Conference. *Journal of the International Otter Survival Fund*, **1,** 7-28.

Cop, Frkovic (1998) Slovenia and its present status in Slovenia and Croatia. *Hystrix Italian Journal of Mammology,* **10,** 65-76.

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9,** 539-552.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology*, **15,** 2833-2843.

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20,** 2363-2369.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163,** 367-374.

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

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.

Coulon A, Guillot G, Cosson JF (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669-1679.

Coulon A, Fitzpatrick JW, Bowman R, Stith BM, Makarewich CA, Stenzler LM, Lovette IJ (2008) Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma coerulescens*). *Molecular Ecology,* **17**, 1685-1701.

Countryside Information System (CIS) version 8 geographical database applications http://www.cis-web.org.uk/

Couvet D (2002) Deleterious effects of restricted gene flow in fragmented populations. *Conservation Biology*, **16**, 369-376.

Coxon KE, Chanin PRF, Dallas JF, Sykes T (1999) The use of DNA fingerprinting to study the population dynamics of otters (*Lutra lutra*) in Southern Britain: Feasibility study. *The Environment Agency, R&D report W202.*

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

Crawford A (2003) *Fourth Otter Survey of England 2000-2002.* Environment Agency. Bristol.

Crawford A, Evans D, Jones A, McNulty J (1979) *Otter Survey of Wales 1977-78.* Society for the Promotion of Nature Conservation. Lincoln.

Crompton AE, Obbard ME, Petersen SD, Wilson PJ (2008) Population genetic structure in polar bears (*Ursus maritimus*) from Hudson Bay, Canada: Implications of future climate change *Biological Conservation*, **141**, 2528-2539.

Crosby KA, Licht LE, Fu J (2008) The effect of habitat fragmentation on finescale population structure of wood frogs (*Rana sylvatica*). *Conservation Genetics,* **10,** 1707-1718.

Croteau EK, Lougheed SC, Krannitz PG, Mahony NA, Walker BL, Boag PT (2007) Genetic population structure of the sagebrush Brewer's sparrow, *Spizella breweri breweri*, in a fragmented landscape at the northern range periphery. *Conservation Genetics,* **8,** 1453-1463.

Cushman SA, McKelvey KS, Hayden J and Schwartz MK (2006) Gene Flow in complex landscapes: Testing multiple hypotheses with causal modeling. *The American Naturalist,* **168**, 486-499.

Dallas JF, Piertney SB (1998) Microsatellite primers for the Eurasian otter. *Molecular Ecology*, **7**, 1248-1251.

Dallas JF, Bacon PJ, Carss DN, Conroy CWH, Green R, Jefferies DJ, Kruuk H, Marshall F, Piertney SB, Racey PA (1999) Genetic diversity in the Eurasian otter, *Lutra lutra*, in Scotland. Evidence from microsatellite polymorphism. *Biological Journal of the Linnean Society*, **68**, 73-86.

Dallas JF, Coxon KE, Sykes T, Chanin PRF, Marshall F, Carss DN, Bacon, PJ, Piertney SB, Racey PA (2003) Similar estimates of population genetic composition and sex ratio derived from carcasses and faeces of Eurasian otter *Lutra lutra*. *Molecular Ecology*, **12**, 275-282.

Dallas JF, Marshall F, Piertney SB, Bacon PJ, Racey PA (2002) Spatially restricted gene flow and reduced microsatellite polymorphism in the Eurasian otter *Lutra lutra* in Britain. *Conservation Genetics*, **3**, 15-29.

Danielson BJ, Hubbard MW (2000) The influence of corridors on the movement behavior of individual *Peromyscus polionotus* in experimental landscapes. *Landscape Ecology*, **15**, 323-331.

Datta S (2003) Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, **19**, 459-466.

Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research*, **78**, 59-77.

Dhuyvetter H, Gaublomme E, Desender K (2005) Bottlenecks, drift and differentiation: the fragmented population structure of the saltmarsh beetle *Pogonus chalceus*. *Genetica,* **124**, 167-177.

Crosby KA, Licht LE, Fu J (2008) The effect of habitat fragmentation on finescale population structure of wood frogs (*Rana sylvatica*). *Conservation Genetics,* **10,** 1707-1718.

Croteau EK, Lougheed SC, Krannitz PG, Mahony NA, Walker BL, Boag PT (2007) Genetic population structure of the sagebrush Brewer's sparrow, *Spizella breweri breweri*, in a fragmented landscape at the northern range periphery. *Conservation Genetics,* **8,** 1453-1463.

Cushman SA, McKelvey KS, Hayden J and Schwartz MK (2006) Gene Flow in complex landscapes: Testing multiple hypotheses with causal modeling. *The American Naturalist,* **168**, 486-499.

Dallas JF, Piertney SB (1998) Microsatellite primers for the Eurasian otter. *Molecular Ecology*, **7**, 1248-1251.

Dallas JF, Bacon PJ, Carss DN, Conroy CWH, Green R, Jefferies DJ, Kruuk H, Marshall F, Piertney SB, Racey PA (1999) Genetic diversity in the Eurasian otter, *Lutra lutra*, in Scotland. Evidence from microsatellite polymorphism. *Biological Journal of the Linnean Society*, **68**, 73-86.

Dallas JF, Coxon KE, Sykes T, Chanin PRF, Marshall F, Carss DN, Bacon, PJ, Piertney SB, Racey PA (2003) Similar estimates of population genetic composition and sex ratio derived from carcasses and faeces of Eurasian otter *Lutra lutra*. *Molecular Ecology*, **12**, 275-282.

Dallas JF, Marshall F, Piertney SB, Bacon PJ, Racey PA (2002) Spatially restricted gene flow and reduced microsatellite polymorphism in the Eurasian otter *Lutra lutra* in Britain. *Conservation Genetics*, **3**, 15-29.

Danielson BJ, Hubbard MW (2000) The influence of corridors on the movement behavior of individual *Peromyscus polionotus* in experimental landscapes. *Landscape Ecology*, **15**, 323-331.

Datta S (2003) Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, **19**, 459-466.

Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research*, **78**, 59-77.

Dhuyvetter H, Gaublomme E, Desender K (2005) Bottlenecks, drift and differentiation: the fragmented population structure of the saltmarsh beetle *Pogonus chalceus*. *Genetica,* **124**, 167-177.

Diniz-Filho JAF, Telles MPC, Bonatto SL, Eizirik E, de Freitas TRO, de Marco Jr P, Santos FR, Sole-Cava A, Soares TN (2008) Mapping the evolutionary twilight zone: molecular markers, populations and geography. *Journal of Biogeography,* **35**, 753-763.

Dionne, Caron, Dodson, Bernatchez (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon the interaction of gene flow and local adaptation. *Molecular Ecology,* **17**, 2382-2396.

Drielsma M, Ferrier S, Manion G (2007) A raster-based technique for analysing habitat configuration: The cost-benefit approach. *Ecological Modelling,* **202,** 324-332.

Dubes R, Jain AK (1976) clustering techniques: the user's dilemma. *Pattern Recognition,* **8**, 247-260.

Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution,* **26**, 1963-1973.

Durbin LS (1993). *Food and habitat utilisation of otters (Lutra lutra L) in a riparian habitat – the River Don in north-east Scotland.* Unpublished PhD thesis, University of Aberdeen.

Durbin (1998) Habitat selection by five otters *Lutra lutra* in rivers of northern Scotland. *Journal of Zoololgy,* **245**, 85-92.

Environmental Systems Research Institute (ESRI) (2000) Arcview 3.2/Spatial Analyst. Geographic Information System. Redlands, California.

Environmental Systems Research Institute (ESRI) (2007) ArcMap 9.2 Geographic Information System. Redlands, California.

Environment Agency (1998) *Understanding riverbank erosion.* Environment Agency report.

Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics*, **124**, 757-771.

Epps CW, Palsbøll J, Wehausen JD, Roderick GK, Ramey RR, McCullough DR (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters*, **8**, 1029-1038.

Epps CW, Wehausen JD, Bleich VC, Torres SG, Brashares JS (2007) Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology,* **44**, 714-724.

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.

Excoffier L, Estoup A, Cornuet J (2005) Bayesian Analysis of an Admixture Model With Mutations and Arbitrarily Linked Markers. *Genetics*, **169**, 1727-1738.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**, 745-758.

Fabbri E, Miquel C, Lucchini V, Santini A, Caniglia R, Duchamp C, Weber J, Lequette B, Marucco F, Boitani L, Fumagalli L, Taberlet P, Randi E (2007) From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. *Molecular Ecology*, **16**, 1661-1671.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Ferreras P, Delibes M, Palomares F, Fedriani J, Calzada J, Revillaa E (2004) Proximate and ultimate causes of dispersal in the Iberian lynx *Lynx pardinus*. *Behavioral Ecology*, **15**, 31-40.

Ferreras P (2001) Landscape structure and asymmetrical inter-patch connectivity in a metapopulation of the endangered Iberian lynx. *Biological Conservation*, **100**, 125-136.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149-1166.

Forman RTT, Alexander LE (1998) Roads and their major ecological effects. *Annual Review of Ecology and Systematics*, **29**, 207-231.

Francois O, Ancelet S, Guillot G (2006) Bayesian Clustering using hidden Markov random fields in spatial population genetics. *Genetics*, **174**, 805-816.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, United Kingdom.

Frantz AC, Pourtois JT, Heuertz M (2006) Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population. *Molecular Ecology*, **15**, 3191-3203.

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.

Funk WC, Blouin MS, Corn PS, Maxell BA, Pilliod DS, Amish D, Allendorf FW (2005) Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology*, **14**, 483-496.

Garcia LV (2004) Escaping the Bonferroni iron claw in ecological studies. *Oikos*, **105**, 657-663.

Geffin E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology*, **13**, 2481-90.

Girman DJ, Mills MGL, Geffen E, Wayne RK (1997) A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (*Lycaon pictus*). *Behavioral Ecology and Socio-biology*, **40**, 187-198.

Goodnight KF, Queller DC (1999) Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology*, **8**, 1231-1234.

Goossens B, Setchell JM, Vidal C, Dilambaka E, Jamart A (2003) Successful reproduction in wild-released orphan chimpanzees (*Pan troglodytes troglodytes*). *Primates*, **44**, 67-69.

Gorman ML, Kruuk H, Jones C, McClaren G, Conroy JWH (1998) The demography of European otters *Lutra lutra*. *Symposium. Zoological Society of London*, **71**, 107-118.

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

Green J, Green R (1980) *Otter survey of Scotland 1977-79.* London, The Vincent Wildlife Trust.

Green J, Green R, and Jefferies DJ (1984). A radio-tracking survey of otters *Lutra lutra* on a Perthshire river system. *Lutra*, **27**, 85-145.

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

Guillot G, Santos F, Estoup A (2005b). Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, **24**, 1406-1407.

Gula R, Hausknecht R, Kuehn R (2009) Evidence of wolf dispersal in anthropogenic habitats of the Polish Carpathian Mountains. *Biodiversity and Conservation*, **18**, 2173-2184.

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

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation and population genetic models. *Heredity*, **83**, 145-154.

Harrison CJO (1988) The Otter Trust and the River Otter Studbook. *IUCN Otter Specialist Group Bulletin*, **3**, 10-11.

Hartl DL, Clark AG (1997) *Principles of Population Genetics*. Sunderland, MA, Sinauers Associates.

Hauer S, Ansorge H, and Zinke O (2002) Mortality patterns of otters (*Lutra lutra*) from eastern Germany. *Journal of Zoology, London*, **256**, 361-368.

Hobbs GI, Chadwick EA, Slater FM, Bruford, MW (2006) Landscape genetics applied to a recovering otter (*Lutra lutra*) population in the UK: preliminary results and potential methodologies. *Hystrix Italian Journal of Mammalogy*, **17**, 47-63.

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

Holden J, Shotbolt L, Bonn A, Burt TP, Chapman PJ, Dougill AJ, Fraser EDG, Hubacek K, Irvine B, Kirkby MJ, Reed MS, Prell C, Stagl S, Stringer LC, Turner A, Worrall F (2007) Environmental change in moorland landscapes. *Earth-Science Reviews*, **82**, 75-100.

Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. *Landscape Ecology*, **21**, 793-796.

Huang C-C, Hsu Y-C, Lee L-L, Li S-H (2005) Isolation and characterization of tetramicrosatellite DNA markers in the Eurasian otter (*Lutra lutra*). *Molecular Ecology Notes*, **5**, 314-316.

Hung C-M, Li S-H, Lee L-L (2004) Faecal DNA typing to determine the abundance and spatial organisation of otters (*Lutra lutra*) along two stream systems in Kinmen. *Animal Conservation*, **7**, 301-311.

Hubisz MJ, Falush D, Stephens M and Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information *Molecular Ecology Resources,* **9**, 1322-1332.

Ihaka R Gentleman R (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics,* **5**, 299-314.

IUCN (1990) IUCN red list of threatened animals. IUCN. Gland.

IUCN (2004) The IUCN red list of threatened species. *www.redlist.org.*

IUCN (2008) 2008 red list of threatened species. *www.iucnredlist.org.*

Jain AK, Murty MN, Flynn PJ (1999) Data Clustering: A Review. *ACM Computing Surveys*, **31**, 264-323.

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.

Janssens X, Fontaine MC, Michaux JR, Libois R, de Kermabon J, Defourny P, Baret FV (2008) Genetic pattern of the recent recovery of European otters in southern France. *Ecography,* **31**, 176-186.

Jarne P, Lagoda JL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424-428.

Jedrzejewska, Sidorovich, Pikulik, Jedrzejewski (2001) Feeding habits of the otter and the American mink in Bialowieza Primeval Forest (Poland) compared to other Eurasian populations. *Ecography,* **24**, 165-180.

Jefferies DJ (1989) The changing otter population of Britain 1700-1989. *Biological Journal of the Linnean Society,* **38**, 61-69.

Jefferies DJ, Hanson HM (2001) The role of dieldrin in the decline of the otter (*Lutra lutra*) in Britain: the analytical data. In: Conroy JWH, Yoxon P, Gutleb AC (Eds.), Procceedings of the First Otter Toxicology Conference. *Journal of the International Otter Survival Fund*, No 1. Skye, September 2000. International Otter Survival Fund, Broadford, Scotland, 95-143.

Jenkins D, Burrows GO (1980) Ecology of otters in northern Scotland III. The use of faeces as indicators of otter (*Lutra lutra*) density and distribution. *Journal of Animal Ecology*, **49**, 755-744.

Jessop RM, Cheyne DL (1992) The re-introduction of the European otter *Lutra l. lutra* into lowland England carried out by the Otter Trust 1983/92: A progress report. *Otters, Journal of the Otter Trust*, 11-16.

Johansson MI, Banks MA, Glunt KD, Hassel-Finnegan HM, Buonaccorsi VP (2008) Influence of habitat discontinuity, geographical distance, and oceanography on fine-scale population genetic structure of copper rockfish (*Sebastes caurinus*). *Molecular Ecology,* **17**, 3051-3061.

Jones T, Jones D (2004). *Otter survey of Wales 2002.* Environment Agency Wales.

Kaczensky P, Blazic M, Gossow H (2004) Public attitudes towards brown bears (*Ursus arctos*) in Slovenia. *Biological Conservation,* **118**, 661-674.

Kays RW, Gittleman JL, Wayne RK (2000). Microsatellite analysis of kinkajou social organization. *Molecular Ecology*, **9**, 743-751.

Kie JG, Ager AA, Bowyer RT (2005) Landscape-level movements of North American elk (*Cervus elaphus*): effect of habitat patch structure and topography. *Landscape Ecology*, **20**, 289-300.

Kitchener AC, Yamaguchi N, Ward JM, MacDonald DW (2005) A diagnosis for the Scottish wildcat (*Felis silvestris*): a tool for conservation action for a critically-endangered felid. *Animal Conservation*, **8**, 223-237.

Koepfli KP, Wayne RK (1998) Phylogenetic relationships of otters (Carnivora: Mustelidae) based on mitochondrial cytochrome *b* sequences. *Journal of Zoology,* **246**, 401-416.

Konovalov DA, Manning C, Henshaw MT (2004) KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes,* **4**, 779-782.

Kowalik RA, Cooper DM, Evans CD, Ormerod SJ (2007) Acidic episodes retard the biological recovery of upland British streams from chronic acidification. *Global Change Biology*, **13**, 2439-2452.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology and Evolution*, **23**, 141-148.

Kruuk H (1992) Scent marking by otters *(Lutra lutra)*: signaling the use of resources. *Behavioural Ecology,* **3**, 133-140.

Kruuk H (2006) *Otters: ecology, behaviour, and conservation.* Oxford University Press.

Kyle CJ, Davison A Strobeck C (2003) Genetic structure of European pine martens (*Martes martes*), and evidence for introgression with *M. americana* in England. *Conservation Genetics,* **4,** 179-188.

Landguth EL, Cushman SA (2009) CDPOP: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, **10**, 156-161.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295-302.

Latch EK, Scognamillo DG, Fike JA, Chamberlain MJ, Rhodes Jr OE (2008) Deciphering ecological barriers to north American river otter (*Lontra canadensis*) gene flow in the louisiana landscape. *Journal of Heredity*, **99**, 265-274.

Lecis R, Ferrando A, Ruiz-Olmo J, Manas S, Domingo-Roura X (2008) Population genetic structure and distribution of introduced American mink (*Mustela vison*) in Spain, based on microsatellite Variation. *Conservation Genetics,* **9**, 1149-1161.

Lefèvre F, Fady B, Fallour-Rubio D, Ghosn D, Bariteau M (2004) Impact of founder population, drift and selection on the genetic diversity of a recently translocated tree population. *Heredity*, **93**, 542-550.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology,* **74**, 1659-1673.

Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. *Plant Ecology*, **80**, 107-138.

Lenton EJ, Chanin PRF, Jefferies DJ (1980) *Otter survey of England 1977-79.* Nature Conservancy Council London.

Lewis PO, Zaykin D (2001) Genetic Data Analysis: Computer program for the analysis of Allelic data, Version 1.1. Available from: URL http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php

Lindley SJ, Walsh T (2005) Inter-comparison of interpolated background nitrogen dioxide concentrations across Greater Manchester, UK. *Atmospheric Environment*, **39**, 2709-2724.

Linnell JDC, Swenson JE, Andersen R (2001) Predators and people: conservation of large carnivores is possible at high human densities if management policy is favourable. *Animal Conservation,* **4**, 345-349.

Litzkow M, Livny M, Mutka M (1998) Condor- a hunter of idle workstations, Proceedings of the 8th International Conference of Distributed Computing Systems.

Lopez MCG (2004) Riparian zone characteristics, fluvial attributes and watershed land-use, and the utility of river otter *(Lutra lutra)* as indicator species, in rural Aragon, Spain. MSC thesis. University of Victoria.

Lucchini V, Fabbri E, Marucco F, Ricci S, Boitani L, Randi E (2002) Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. *Molecular Ecology*, **11**, 857-868.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution*, **20**, 136-142.

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.

Mank JE, Avise, JC (2004) Individual organisms as units of analysis: Bayesian-clustering alternatives in population genetics. *Genetics Research, Cambridge,* **84**, 135-143.

Manning AD, Lindenmayer B, Nix HA (2004) Continua and umwelt: novel perspectives on viewing landscapes. *Oikos,* **104,** 621-628.

Mantel N (1967) Detection of disease clustering and a generalized Regression Approach. *Cancer Research,* **27**, 209-220.

Mason CF, Macdonald SM (1986) *Otters: ecology and conservation*. Cambridge University Press.

Mason CF, Macdonald SM (1989) Acidification and otter *(Lutra lutra)* distribution in Scotland. *Water, Air, and Soil Pollution,* **43**, 365-374.

Mason CF, Macdonald SM (2004) Growth in otter (*Lutra lutra*) populations in the UK as shown by long-term monitoring. *Ambio*, **33**, 148-152.

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, **82**, 290-297.

McGarigal K, Tagil S, Cushman SA (2009) Surface metrics: an alternative to patch metrics for the quantification of landscape structure. *Landscape Ecology*, **24**, 433-450.

McIntyre S, Barrett GW (1992) Habitat variegation, an alternative to fragmentation. *Conservation Biology,* **4**, 197-202.

Mech LD (1995) The challenge and opportunity of recovering wolf populations. *Conservation Biology*, **9**, 270-278.

Mills L, Allendorf FW (1996) The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, **10**, 1509-1518.

Mills L, Citta J, Lair K, Schwartz M, Tallmon D (2000) Estimating animal abundance using non-invasive DNA sampling: promise and pitfalls. *Ecological Applications*, **10**, 283-294.

Monteith DT, Evans CD (2000) *UK Acid Waters Monitoring Network: 10 year report*. ENSIS Publishing, London.

Montgomery DR, Buffington JM (1997) Channel-reach morphology in mountain drainage basins. *GSA Bulletin,* **109**, 596-611.

Moran (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos,* **100**, 403-405.

Mucci N (2008) *Assessing the patterns of genetic diversity in otter (Lutra lutra) populations in Europe.* Unpublished PhD thesis. Università di Bologna.

Myers M, Knoll AH (2001) The biotic crisis and the future of evolution. *Proceedings of the National Academy of Sciences,* **98**, 5389-5392.

Nakagawa S (2004) A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology,* **15**, 1044-1045.

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591-605.

O'Ryan C, Bruford M, Beaumont M, Wayne R, Cherry M, Harle E (1998) Genetics of fragmented populations of African buffalo (*Syncerus caffer*) in South Africa. *Animal Conservation*, **1**, 85-94.

Pastorini J, Zaramody A, Curtis, DJ, Nievergelt CM, Mundy IM (2009) Genetic analysis of hybridisation and introgression between wild mongoose and brown lemurs. *BMC Evolutionary Biology,* **9**,

Paul MJ, Meyer JL (2001) Streams in the urban landscape. *Annual Review of Ecology and Systematics*, **32**, 333-365.

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, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Molecular Ecology*, **4,** 135-147.

Pearse DE, Crandall KA (2004) Beyond F-ST: Analysis of population genetic data for conservation. *Conservation Genetics*, **5,** 585-602.

Pérez-Espona S, Pérez-Barbería FJ, Mcleod JE, Jiggins CD, Gordon IJ, Pemberton JM (2008) Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology,* **17,** 981-996.

Pertoldi C, Hansen MM, Loeschcke V, Madsen AB, Jacobsen L, Baagoe H (2001) Genetic consequences of population decline in the European otter (*Lutra lutra*): an assessment of microsatellite DNA variation in Danish otters from 1883 to 1993. *Proceedings of the Royal Society B: Biological Sciences*, **268,** 1775-1781.

Philcox CK, Grogan AL, Macdonald DW (1999) Patterns of otter *Lutra lutra* road mortality in Britain. *Journal of Applied Ecology*, **36,** 748-762.

Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C (2008) Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula genetics. *Genetics*, **180,** 1009-1021.

Pilot M, Jedrzejewski W, Branicki W, Sidorovich VE, Jedrzejewska B, Stachura K, Funk SM (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology*, **15,** 4533-4553.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) Computer note GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95,** 536-539.

Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the atlantic cod, *Gadus Morhua,* at large and small geographic scales. *Evolution,* **55,** 131-146.

Prigioni C, Remonti L, Balestrieri A, Sgrosso S, Priore G, Misin C, Viapiana M, Spada S, Anania R (2005). Distribution and sprainting activity of the otter (*Lutra lutra*) in the Pollino National Park (southern Italy). *Ethology Ecology & Evolution*, **17,** 171-180.

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

Pritchard JK, Wen W (2003) Documentation for STRUCTURE software: Version 2. Available from http://pritch.bsd.uchicago.edu.

QIAGEN, #206143, QIAGEN® Multiplex PCR Handbook, Protocol: Amplification of microsatellite loci using multiplex PCR, 19-21.

QIAGEN, #65906, DNeasy Tissue handbook, Protocol: Purification of total DNA from animal tissues, 18-20.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258-275.

Ralls K, Ballou JD, Templeton AR (1988) Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology*, **2**, 185-193.

Rand WM (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association,* 846-850.

Randi E, Davoli F, Pierpaoli M, Pertoldi C, Madsen AB, Loeschck V (2003) Genetic structure in otter (*Lutra lutra*) populations in Europe: implications for conservation. *Animal Conservation*, **6**, 93-100.

Rannala B, Hartigan JA (1996) Estimating gene flow in island populations. *Genetics Research*, **67**, 147-58.

Raufaste N, Rousset F (2001) Are partial Mantel tests adequate? *Evolution*, **55**, 1703-1705.

Ray N (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes,* **5,** 177-180.

Ray N, Lehmann A, Joly P (2002) Modeling spatial distribution of amphibian populations: a GIS approach based on habitat matrix permeability. *Biodiversity and Conservation,* **11,** 2143-2165.

Raymond M, Rousett F (1995) Genepop (version 1.2) population genetics software for exact tests and ecumenicalism. *Journal of Heredity,* **86,** 248-249.

de Ruiter JR, Geffen E (1998) Relatedness of Matrilines, Dispersing Males and social groups in long-tailed macaques (*Macaca fascicularis*). *Proceedings of the Royal Society B: Biological Sciences*, **265**, 79-87.

Richards C, Johnson LB, Host GE (1996) Landscape scale influences on stream habitats and biota. *Canadian Journal of Fisheries and Aquatic Science*, **53**, 295-311.

Richardson JS (2004) Meeting the conflicting objectives of stream conservation and land use through riparian management: another balancing act. Pages 1-6 in GJ Scrimgeour, G Eisler, B McCulloch, U Silins and M Monita. Editors. Forest Land–Fish Conference II – Ecosystem stewardship through collaboration. *Proceedings of the Forest-Land-Fish Conference II*, April 26-28, Edmonton, Alberta.

Roach JL, Stapp P, Van Horne B, Antolin MF (2001) Genetic structure of a metapopulation of black-tailed prairie dogs. *Journal of Mammalogy*, **82**, 946-959.

Robin K (1987) European River Otter Studbook. *IUCN Otter Specialist Group Bulletin*, **2**, 42-44.

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics,* **1**, 660- 671.

Roth NE, Allan JD, Erickson DL (1996) Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology*, **11**, 141-156.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219-1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.

Rousset F (2002) Partial Mantel tests: reply to Castellano and Balleto. *Evolution*, **56**, 1874-1875.

Rowe G, Beebee TJC (2007) Defining population boundaries: use of three Bayesian approaches with microsatellite data from British natterjack toads (*Bufo calamita*). *Molecular Ecology*, **16**, 785-796.

Ruiz-Olmo J, Calvo A, Palazón S, Arqued V (1998) Is the otter a bioindicator? *Galemys,* **10,** 227-237.

Ruiz-Olmo J, López-Martin JN, Palazón S (2001) The influence of fish abundance on the otter (*Lutra lutra*). Populations in Iberian Mediterranean habitats. *Journal of Zoology,* **254**, 325-336.

Ruiz-Olmo J, Saavedra D, Jiménez J (2001b) Testing the surveys and visual and track censuses of Eurasian otters (*Lutra lutra*). *Journal of Zoology,* **253**, 359-369.

Ryman N, Laikre L (1991) Effects of supportive breeding on the genetically effective population size. *Conservation Biology*, **5**, 325-329.

Sacks BN, Mitchell BR, Williams CL, Ernest HB (2005) Coyote movements and social structure along a cryptic population genetic subdivision. *Molecular Ecology*, **14**, 1241-1249.

Sahlsten J, Thörngren H, Höglund J (2008) Inference of hazel grouse population structure using multilocus data: a landscape genetic approach. *Heredity*, **101**, 475-482.

Sansom AL (1999) Upland Vegetation Management: The impacts of over stocking. *Water Science and Technology*, **39**, 85-92.

Schwartz CC, Swenson JE, Miller SD (2003) Large carnivores, moose, and humans: A changing paradigm of predator management in the 21st century. *Alces*, **39**, 41-63.

Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, **14**, 1679-1685.

Sherrard-Smith E, Cable J, Chadwick EA (2009) Distribution of Eurasian otter biliary parasites *Pseudamphistomum truncatum* and *Metorchis albidus* (Family Opistorchiidae), in England and Wales. *Parasitology*, **136**, 1015-1022.

Simpson VR (1998). A post mortem study of otters (*Lutra lutra*) found dead in south west England. *R&D Technical report W148*. Environment Agency, Swindon.

Sjoasen T (1997) Movements and establishment of reintroduced European otters *Lutra lutra*. *Journal of Applied Ecology*, **34**, 1070-1080.

Smith D, Meier T, Geffen E, Mech LD, Burch JW, Adams LG, Wayne RK (1997) Is incest common in gray wolf packs? *Behavioral Ecology*, **8**, 384-391.

Smouse P, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561-573.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627-632.

Snelder TH (2002) Multiscale river environment classification for water resources. *Journal of the American Water Resources Association*, **38**, 1225-1239.

Snyder NF, Derrickson SR, Beissinger SR, Wiley JW, Smith TB, Toone WD, Miller B (1996) Limitations of captive breeding in endangered species recovery. *Conservation Biology*, **10**, 338-348.

Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology*, **14**, 2553-2564.

Stanton DWG, Hobbs GI, Chadwick EA, Slater FM, Bruford MW (2009) Mitochondrial genetic diversity and structure of the European otter (*Lutra lutra*) in Britain. *Conservation Genetics*, **10**, 733-737.

Storfer A (1996) Quantitative genetics: a promising approach for the assessment of genetic variation in endangered species. *Trends in Ecology & Evolution*, **11**, 343-348.

Storfer A (1999) Gene flow and endangered species translocations: a topic revisited. *Biological Conservation*, **87**, 173-180.

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**, 128-142.

Strachan R (2007) *National survey of otter* Lutra lutra *distribution in Scotland 2003-2004.* Scottish Natural Heritage Commissioned Report No. 211.

Strachan R, Jefferies DJ (1996) *Otter Survey of England 1991-1994.* Vincent Wildlife Trust, London.

Strasburg JL (2006) Conservation biology: Roads and genetic connectivity. *Nature*, **440**, 875-876.

Sumner J, Rousset F, Estoup A, Moritz C (2001) 'Neighbourhood' size, dispersal and density estimates in the prickly forest skink (*Gnypetoscincus queenslandiae*) using individual genetic and demographic methods. *Molecular Ecology*, **10**, 1917-1927.

Taberlet P, Camarra J, Griffin S, Uhrès E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J (1997). Noninvasive Genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, **6**, 869-876.

Taberlet P (2003) Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation*, **6**, 83-92.

Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity is a vital element of landscape structure. *Oikos,* **68**, 571-573.

Telles MP, Diniz-Filho JAF (2005) Multiple Mantel tests and isolation-by-distance, taking into account long-term historical divergence. *Genetics and Molecular Research*, **4**, 742-748.

Thalamuthu A, Mukhopadhyay I, Zheng XJ, Tseng GC (2006) Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics,* **22**, 2405-2412.

Topchy A, Jain AK, Punch W (2003) *Combining Multiple Weak Clusterings* In Proceedings IEEE International Conference on data mining, Melbourne, FL, 331-338.

Utami SS, Goossens B, Bruford MW, de Ruiter JR, van Hooff JARAM (2002) Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology,* **13**, 643-652.

Valière N, Fumagalli L, Gielly L, Miquel C, Lequette B, Poulle M-L, Weber J-M, Arlettaz R (2003) Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation,* **6**, 83-92.

Verbeylen G, De Bruyn L, Adriaensen F, Matthysen E (2003) Does matrix resistance influence red squirrel (*Sciurus vulgaris* L. 1758) distribution in an urban landscape? *Landscape Ecology,* **18**, 791-805.

Verhoeven JJF, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control: increasing your power. *Oikos,* **108**, 643-647.

Vignieri SN (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Molecular Ecology*, **14**, 1925-1937.

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of earth's ecosystems. *Science*, **277**, 494-499.

Wang IJ, Savage WK, Shaffer HB (2009) Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*) *Molecular Ecology*, **18**, 1365-1374.

Ward, JV (1998) Riverine landscapes: Biodiversity patterns, disturbance regimes, and aquatic conservation. *Biological Conservation,* **83**, 269-278.

Wayre P (1992) The first twenty-one years of the otter trust and its re-introduction programme. *Otters, Journal of the otter trust,* **2,** 17-22.

White PCL, McCleana CJ, Woodroffe GL (2003) Factors affecting the success of an otter (*Lutra lutra*) reinforcement programme, as identified by post-translocation monitoring. *Biological Conservation,* **112,** 363-371.

White PCL, Gregory KW, Lindley PJ, Richards G (1997) Economic values of threatened mammals in Britain: A case study of the otter *Lutra Lutra* and the water vole *Arvicola terrestris. Biological Conservation,* **82,** 345-354.

With KA, Gardner RH, Turner MG (1997) Landscape connectivity and population distributions in heterogeneous environments. *Oikos,* **78,** 151-169.

Wright S (1943) Isolation by distance. *Genetics,* **28,** 114-138.

Wright S (1951) The genetical structure of populations. Annals of Eugenics. **15,** 323-354.

Yeo MJM, Blackstock TH (2002) A vegetation analysis of the pastoral landscapes of upland Wales, UK. *Journal of Vegetation Science,* **13,** 803-816.

Yeung KY, Ruzzo WL (2001) Principal component analysis for clustering gene expression data. *Bioinformatics,* **17,** 763-774.

Zalewski A, Piertney SB, Zalewska H, Lambin X (2009) Landscape barriers reduce gene flow in an invasive carnivore: geographical and local genetic structure of American mink in Scotland. *Molecular Ecology,* **18,** 1601-1615.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology,* **11,** 1-16.

Zhivotosky LA, Feldman MW (1995) Microsatellite variability and genetic distances. *Proceedings of the National Academy of Sciences of the United States of America,* **92,** 11549-11552.

Appendices

Appendix 2.1: Average Assignments of individuals to each population from the Wales and Borders Dataset used in Chapter 2 produced by each of the Bayesian Clustering Techniques.

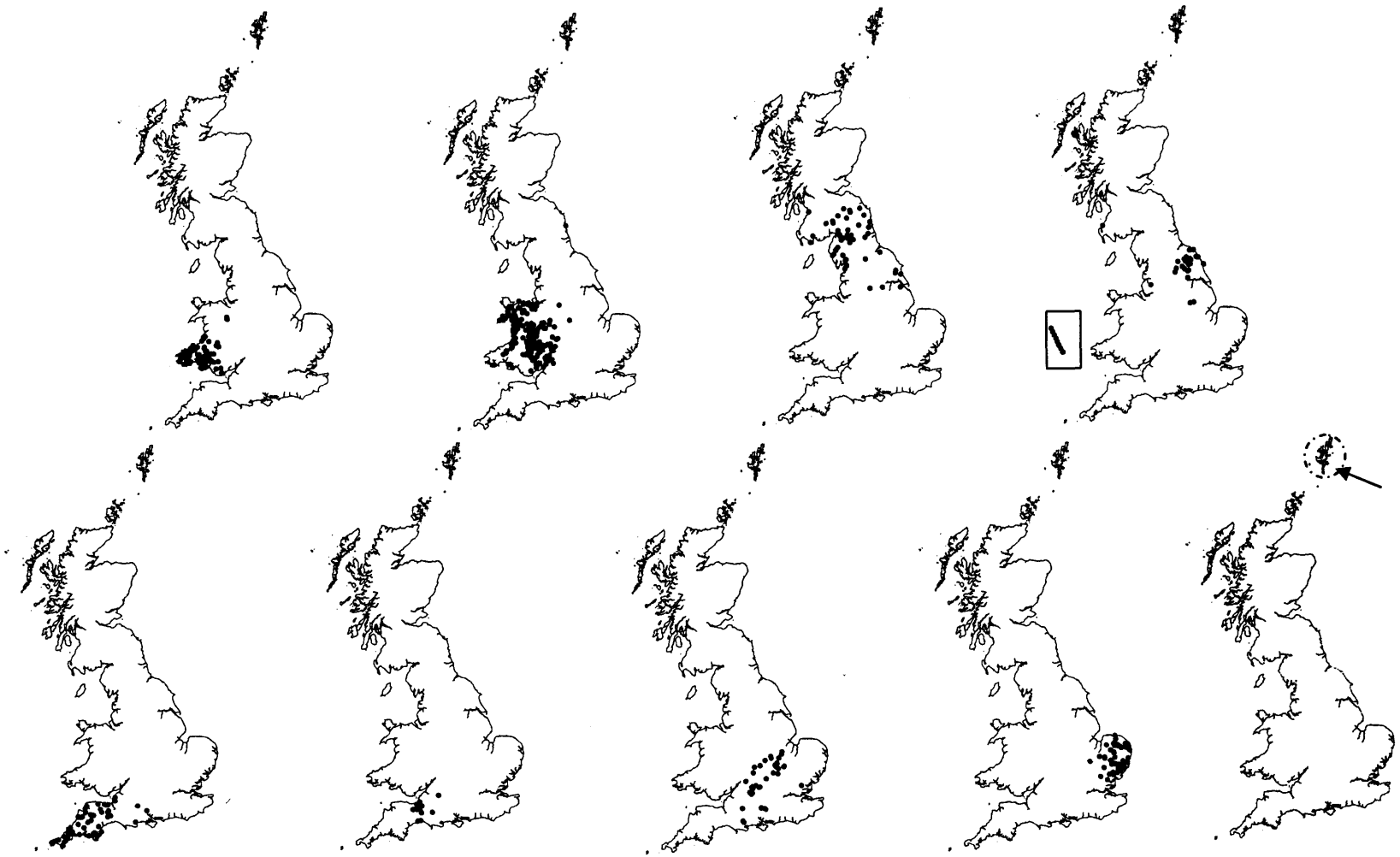
UWCREF	X	Y	PARTITION		STRUCTURE		BAPS 2			BAPS 4 SPATIAL			BAPS 4 NON SPATIAL							SPATIAL GENELAND				NON SPATIAL GENELAND				
			Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
7	296100	254700	1	0	0.90	0.10	1.00	0.00	0.00	1.00	0.00	0.00	0.71	0.00	0.28	0.00	0.01	0.00	0.00	0.97	0.03	0.00	0.00	0.99	0.00	0.00	0.01	0.00
8	304000	228000	0	1	0.07	0.93	0.00	1.00	0.00	0.02	0.98	0.00	0.01	0.03	0.00	0.00	0.96	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
11	253000	339500	1	0	0.78	0.22	1.00	0.00	0.00	0.79	0.22	0.00	0.86	0.13	0.01	0.00	0.00	0.00	0.01	0.01	0.99	0.00	0.00	0.64	0.01	0.00	0.34	0.00
21	201000	200000	0	1	0.07	0.93	0.00	1.00	0.00	0.02	0.98	0.00	0.01	0.02	0.00	0.00	0.97	0.00	0.00	0.95	0.05	0.00	0.00	0.00	1.00	0.00	0.00	0.00
25	293000	186000	1	0	0.90	0.10	1.00	0.00	0.00	1.00	0.00	0.00	0.90	0.07	0.04	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.96	0.01	0.00	0.03	0.00
26	286400	205800	1	0	0.95	0.05	1.00	0.00	0.00	1.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.96	0.00	0.00	0.04	0.00
27	327200	292100	1	0	0.91	0.09	1.00	0.00	0.00	0.77	0.18	0.05	0.95	0.00	0.00	0.03	0.00	0.00	0.02	0.00	1.00	0.00	0.00	0.34	0.00	0.00	0.66	0.00
31	313000	222000	1	0	0.86	0.14	1.00	0.00	0.00	0.95	0.05	0.00	0.92	0.02	0.02	0.00	0.05	0.00	0.00	0.44	0.56	0.00	0.00	0.76	0.01	0.00	0.23	0.00
32	213300	224900	0	1	0.55	0.45	1.00	0.00	0.00	0.65	0.35	0.00	0.25	0.33	0.24	0.00	0.18	0.00	0.00	0.96	0.04	0.00	0.00	0.69	0.19	0.00	0.12	0.00
34	237800	233500	0	1	0.12	0.88	0.00	1.00	0.00	0.00	1.00	0.00	0.13	0.70	0.01	0.01	0.16	0.00	0.00	0.03	0.97	0.00	0.00	0.00	0.90	0.00	0.10	0.00
35	206000	235300	0	1	0.54	0.46	0.00	1.00	0.00	0.41	0.57	0.02	0.04	0.06	0.01	0.76	0.08	0.00	0.05	0.98	0.02	0.00	0.00	0.11	0.06	0.00	0.83	0.00
36	195500	220100	1	0	0.96	0.04	1.00	0.00	0.00	1.00	0.00	0.00	0.04	0.00	0.96	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
37	287000	230000	1	0	0.96	0.04	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
38	181100	208200	1	0	0.90	0.10	1.00	0.00	0.00	1.00	0.00	0.00	0.90	0.07	0.04	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.96	0.00	0.00	0.04	0.00
42	199900	200900	1	0	0.96	0.04	1.00	0.00	0.00	0.98	0.02	0.00	0.89	0.00	0.09	0.02	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.00	0.00	0.05	0.00
47	202000	215000	0	1	0.55	0.45	0.00	1.00	0.00	0.52	0.48	0.00	0.21	0.03	0.00	0.76	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.01	0.09	0.00	0.90	0.00
48	201300	200600	1	0	0.93	0.07	1.00	0.00	0.00	1.00	0.00	0.00	0.11	0.01	0.87	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00
50	230300	241400	1	0	0.97	0.03	1.00	0.00	0.00	1.00	0.00	0.00	0.03	0.00	0.98	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
51	242800	202800	1	0	0.75	0.25	1.00	0.00	0.00	0.65	0.35	0.00	0.74	0.04	0.03	0.00	0.20	0.00	0.00	0.97	0.03	0.00	0.00	0.56	0.04	0.00	0.40	0.00
56	299400	262800	1	0	0.67	0.33	1.00	0.00	0.00	0.49	0.51	0.00	0.18	0.03	0.46	0.03	0.31	0.00	0.00	0.51	0.49	0.00	0.00	0.46	0.11	0.00	0.43	0.00
57	304000	228000	0	1	0.07	0.93	0.00	1.00	0.00	0.01	0.99	0.00	0.00	0.09	0.02	0.00	0.90	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
58	217000	245900	1	0	0.90	0.10	1.00	0.00	0.00	0.93	0.07	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.67	0.00	0.00	0.33	0.00
60	243200	227900	0	1	0.64	0.36	1.00	0.00	0.00	0.88	0.12	0.00	0.12	0.22	0.61	0.01	0.04	0.00	0.01	0.74	0.26	0.00	0.00	0.67	0.15	0.00	0.18	0.00
62	341300	307400	0	1	0.11	0.89	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.05	0.00	0.96	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.97	0.00	0.02	0.00
71	297900	228600	0	1	0.10	0.90	0.00	1.00	0.00	0.05	0.95	0.00	0.00	0.00	0.05	0.00	0.95	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00
72	318100	297800	1	0	0.92	0.08	1.00	0.00	0.00	1.00	0.00	0.00	0.68	0.00	0.27	0.05	0.00	0.00	0.00	0.10	0.90	0.00	0.00	0.88	0.00	0.00	0.13	0.00
73	337000	250000	0	1	0.17	0.83	0.00	1.00	0.00	0.00	1.00	0.00	0.04	0.00	0.00	0.25	0.71	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.70	0.00	0.30	0.00
75	260700	202400	1	0	0.86	0.14	1.00	0.00	0.00	0.97	0.03	0.01	0.03	0.06	0.90	0.00	0.00	0.01	0.00	0.96	0.04	0.00	0.00	0.94	0.01	0.00	0.05	0.00
76	308000	188000	1	0	0.80	0.20	1.00	0.00	0.00	0.98	0.02	0.00	0.00	0.01	0.99	0.00	0.01	0.00	0.00	0.97	0.03	0.00	0.00	0.96	0.00	0.00	0.03	0.00
77	261500	280900	1	0	0.84	0.16	1.00	0.00	0.00	0.76	0.23	0.01	0.68	0.00	0.07	0.09	0.14	0.01	0.00	0.97	0.03	0.00	0.00	0.76	0.02	0.00	0.22	0.00
78	250000	200000	0	1	0.67	0.33	0.00	1.00	0.00	0.48	0.51	0.01	0.16	0.01	0.00	0.84	0.00	0.00	0.00	0.81	0.19	0.00	0.00	0.02	0.02	0.00	0.96	0.00
80	276000	207600	1	0	0.90	0.10	1.00	0.00	0.00	0.98	0.00	0.02	0.86	0.01	0.01	0.09	0.00	0.00	0.02	0.99	0.01	0.00	0.00	0.61	0.00	0.00	0.38	0.00
81	336200	230200	0	1	0.08	0.92	0.00	1.00	0.00	0.01	0.99	0.00	0.00	0.00	0.01	0.00	0.99	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
83	265800	291800	0	1	0.61	0.39	0.00	1.00	0.00	0.18	0.71	0.12	0.01	0.02	0.03	0.90	0.00	0.03	0.02	0.00	1.00	0.00	0.00	0.00	0.01	0.00	0.98	0.00
84	307900	260400	0	1	0.14	0.86	0.00	1.00	0.00	0.06	0.94	0.00	0.00	0.01	0.09	0.00	0.91	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.93	0.00	0.05	0.00
85	317500	297300	0	1	0.12	0.88	0.00	1.00	0.00	0.04	0.96	0.00	0.00	0.87	0.00	0.00	0.12	0.00	0.00	0.01	0.99	0.00	0.00	0.00	1.00	0.00	0.00	0.00
89	320200	219600	1	0	0.61	0.39	1.00	0.00	0.00	0.74	0.26	0.00	0.03	0.00	0.56	0.00	0.42	0.00	0.00	0.39	0.61	0.00	0.00	0.93	0.04	0.00	0.03	0.00
91	305000	258000	0	1	0.25	0.75	0.00	1.00	0.00	0.04	0.94	0.02	0.01	0.94	0.00	0.00	0.00	0.01	0.04	0.03	0.98	0.00	0.00	0.01	0.84	0.00	0.15	0.00
92	317700	319200	0	1	0.13	0.87	0.00	1.00	0.00	0.03	0.98	0.00	0.04	0.01	0.01	0.03	0.91	0.00	0.00	0.17	0.83	0.00	0.00	0.02	0.80	0.00	0.17	0.00
93	263400	234300	0	1	0.06	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.72	0.00	0.00	0.29	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
95	311500	224500	0	1	0.47	0.53	0.00	1.00	0.00	0.33	0.67	0.00	0.55	0.26	0.00	0.17	0.02	0.00	0.00	0.17	0.83	0.00	0.00	0.09	0.20	0.00	0.70	0.00
102	352000	320000	1	0	0.80	0.20	1.00	0.00	0.00	0.90	0.10	0.00	0.06	0.15	0.74	0.05	0.00	0.00	0.00	0.05	0.95	0.00	0.00	0.62	0.02	0.00	0.37	0.00
103	256400	339800	0	1	0.83	0.17	0.00	1.00	0.00	0.18	0.76	0.06	0.03	0.00	0.02	0.93	0.00	0.02	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
104	323000	304000	0	1	0.43	0.57	0.00	1.00	0.00	0.12	0.88	0.00	0.12	0.00	0.02	0.73	0.14	0.00	0.00	0.07	0.93	0.00	0.00	0.02	0.17	0.00	0.81	0.00
105	238200	245500	1	0	0.94	0.06	1.00	0.00	0.00	0.98	0.01	0.01	0.01	0.00	0.96	0.01	0.00	0.01	0.01	1.00	0.00	0.00	0.00	0.67	0.00	0.00	0.32	0.00
109	249500	221900	1	0	0.90	0.10	1.00	0.00	0.00	0.99	0.01	0.00	0.02	0.00	0.98	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00

112	303400	227800	0	1	0.03	0.97	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.02	0.00	0.00	0.99	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
113	324500	275700	0	1	0.14	0.86	0.00	1.00	0.00	0.02	0.98	0.00	0.15	0.05	0.00	0.00	0.80	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.95	0.00	0.03	0.00
114	256900	190200	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
117	306800	250400	1	0	0.85	0.15	1.00	0.00	0.00	0.69	0.27	0.04	0.05	0.00	0.24	0.63	0.03	0.03	0.01	0.73	0.26	0.00	0.00	0.12	0.00	0.00	0.88	0.00
118	201400	238900	1	0	0.60	0.40	1.00	0.00	0.00	0.53	0.47	0.00	0.17	0.68	0.12	0.03	0.00	0.00	0.00	0.98	0.02	0.00	0.00	0.31	0.11	0.00	0.59	0.00
119	210000	215000	1	0	0.87	0.13	1.00	0.00	0.00	0.89	0.09	0.02	0.91	0.00	0.00	0.00	0.06	0.00	0.04	1.00	0.00	0.00	0.00	0.73	0.00	0.00	0.27	0.00
123	261900	283700	0	1	0.75	0.25	0.00	1.00	0.00	0.39	0.60	0.02	0.33	0.02	0.00	0.65	0.00	0.00	0.00	0.23	0.77	0.00	0.00	0.00	0.01	0.00	0.99	0.00
124	241200	205800	1	0	0.93	0.07	1.00	0.00	0.00	1.00	0.00	0.00	0.11	0.00	0.89	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
125	195000	235000	1	0	0.91	0.09	1.00	0.00	0.00	1.00	0.00	0.00	0.84	0.00	0.15	0.00	0.01	0.00	0.00	1.00	0.00	0.00	0.00	0.93	0.00	0.00	0.07	0.00
127	316000	235000	0	1	0.53	0.47	0.00	1.00	0.00	0.06	0.87	0.07	0.00	0.01	0.00	0.96	0.03	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.01	0.00	0.99	0.00
130	275000	215000	0	1	0.48	0.52	1.00	0.00	0.00	0.21	0.75	0.04	0.22	0.45	0.00	0.29	0.03	0.03	0.00	0.85	0.15	0.00	0.00	0.03	0.14	0.00	0.83	0.00
131	379900	214500	0	1	0.12	0.88	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.95	0.00	0.01	0.05	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.87	0.00	0.12	0.00
136	260400	296300	1	0	0.78	0.22	1.00	0.00	0.00	0.89	0.11	0.00	0.22	0.02	0.58	0.00	0.17	0.00	0.01	0.40	0.60	0.00	0.00	0.82	0.03	0.00	0.16	0.00
143	218900	236800	1	0	0.94	0.06	1.00	0.00	0.00	0.91	0.00	0.09	0.64	0.00	0.21	0.01	0.00	0.01	0.13	1.00	0.00	0.00	0.00	0.81	0.00	0.00	0.19	0.00
147	264700	230400	1	0	0.97	0.03	1.00	0.00	0.00	1.00	0.00	0.00	0.94	0.00	0.05	0.00	0.00	0.01	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
148	254500	210700	1	0	0.62	0.38	1.00	0.00	0.00	0.89	0.11	0.00	0.05	0.14	0.74	0.00	0.06	0.01	0.00	1.00	0.00	0.00	0.00	0.94	0.06	0.00	0.00	0.00
149	260000	220000	1	0	0.92	0.08	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
151	278000	364400	0	1	0.16	0.84	0.00	1.00	0.00	0.02	0.99	0.00	0.00	0.88	0.09	0.01	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.76	0.00	0.23	0.00
152	325900	214300	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
155	238400	237900	1	0	0.88	0.12	1.00	0.00	0.00	0.75	0.19	0.07	0.86	0.01	0.00	0.09	0.01	0.01	0.03	0.46	0.54	0.00	0.00	0.33	0.00	0.00	0.67	0.00
163	260000	250000	1	0	0.92	0.08	1.00	0.00	0.00	1.00	0.00	0.00	0.88	0.00	0.12	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.00	0.03	0.00
166	273500	266300	1	0	0.90	0.10	1.00	0.00	0.00	0.86	0.12	0.02	0.01	0.00	0.73	0.25	0.00	0.00	0.01	0.88	0.12	0.00	0.00	0.40	0.00	0.00	0.60	0.00
169	318000	268000	1	0	0.69	0.31	1.00	0.00	0.00	0.21	0.68	0.12	0.19	0.25	0.02	0.44	0.00	0.04	0.06	0.00	1.00	0.00	0.00	0.00	0.03	0.00	0.96	0.00
185	237400	227500	1	0	0.76	0.24	1.00	0.00	0.00	0.96	0.04	0.00	0.01	0.03	0.93	0.00	0.04	0.00	0.00	0.83	0.17	0.00	0.00	0.99	0.01	0.00	0.00	0.00
187	332500	258800	0	1	0.07	0.93	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.21	0.00	0.00	0.78	0.01	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
188	297900	228700	0	1	0.15	0.85	0.00	1.00	0.00	0.04	0.96	0.00	0.16	0.58	0.02	0.01	0.24	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.88	0.00	0.12	0.00
193	180000	220000	1	0	0.92	0.08	1.00	0.00	0.00	1.00	0.00	0.00	0.61	0.00	0.39	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
194	195900	224100	1	0	0.86	0.14	1.00	0.00	0.00	0.83	0.17	0.00	0.09	0.01	0.80	0.00	0.10	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00
195	246500	220400	1	0	0.88	0.12	1.00	0.00	0.00	0.91	0.08	0.01	0.78	0.01	0.13	0.01	0.07	0.00	0.01	1.00	0.00	0.00	0.00	0.69	0.00	0.00	0.31	0.00
206	259100	245400	1	0	0.90	0.10	1.00	0.00	0.00	0.83	0.12	0.05	0.00	0.01	0.79	0.14	0.02	0.03	0.00	0.99	0.01	0.00	0.00	0.38	0.00	0.00	0.62	0.00
208	303200	263200	0	1	0.13	0.87	0.00	1.00	0.00	0.02	0.98	0.00	0.03	0.08	0.08	0.00	0.82	0.00	0.00	0.00	1.00	0.00	0.00	0.03	0.92	0.00	0.06	0.00
210	264300	268200	0	1	0.62	0.38	0.00	1.00	0.00	0.34	0.65	0.02	0.00	0.00	0.19	0.63	0.18	0.00	0.01	0.23	0.77	0.00	0.00	0.05	0.06	0.00	0.90	0.00
211	249700	216800	1	0	0.90	0.10	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
214	244900	354900	0	1	0.14	0.86	0.00	1.00	0.00	0.01	0.99	0.00	0.09	0.91	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.82	0.00	0.17	0.00
215	252500	268400	0	1	0.78	0.22	0.00	1.00	0.00	0.15	0.82	0.03	0.04	0.01	0.00	0.92	0.01	0.02	0.00	0.18	0.82	0.00	0.00	0.00	0.01	0.00	0.99	0.00
216	280000	210000	1	0	0.90	0.10	1.00	0.00	0.00	1.00	0.00	0.00	0.84	0.00	0.15	0.00	0.01	0.00	0.01	1.00	0.00	0.00	0.00	0.95	0.00	0.00	0.05	0.00
218	250100	186500	1	0	0.92	0.08	1.00	0.00	0.00	0.96	0.04	0.00	0.60	0.00	0.35	0.00	0.05	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
225	309000	267000	1	0	0.89	0.11	1.00	0.00	0.00	0.99	0.01	0.00	0.85	0.00	0.00	0.14	0.00	0.00	0.01	0.01	0.99	0.00	0.00	0.46	0.00	0.00	0.54	0.00
226	331200	330900	1	0	0.92	0.08	1.00	0.00	0.00	1.00	0.00	0.00	0.91	0.00	0.09	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.00	0.00	0.05	0.00
242	228800	340200	0	1	0.56	0.44	0.00	1.00	0.00	0.01	0.82	0.18	0.00	0.08	0.07	0.43	0.28	0.12	0.03	0.00	1.00	0.00	0.00	0.00	0.05	0.00	0.95	0.00
243	256800	219600	1	0	0.64	0.36	1.00	0.00	0.00	0.72	0.29	0.00	0.71	0.03	0.00	0.09	0.16	0.00	0.00	1.00	0.00	0.00	0.00	0.31	0.12	0.00	0.57	0.00
244	194300	232100	1	0	0.85	0.15	1.00	0.00	0.00	0.95	0.06	0.00	0.01	0.01	0.97	0.00	0.01	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
248	329900	331600	1	0	0.86	0.14	1.00	0.00	0.00	0.86	0.14	0.00	0.00	0.02	0.96	0.00	0.02	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00
254	341980	204820	0	1	0.17	0.83	0.00	1.00	0.00	0.23	0.77	0.00	0.01	0.09	0.21	0.00	0.69	0.00	0.00	0.00	1.00	0.00	0.00	0.04	0.95	0.00	0.01	0.00
256	218800	240400	1	0	0.89	0.11	1.00	0.00	0.00	0.74	0.26	0.01	0.87	0.09	0.01	0.03	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.30	0.00	0.00	0.70	0.00
259	320000	250000	0	1	0.05	0.95	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.02	0.00	0.00	0.98	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
260	316700	290300	0	1	0.07	0.93	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.94	0.00	0.00	0.07	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00
295	399600	275400	0	1	0.20	0.80	0.00	1.00	0.00	0.01	0.99	0.00	0.00	0.00	0.17	0.01	0.82	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.90	0.00	0.09	0.00
304	290400	229000	0	1	0.21	0.79	0.00	1.00	0.00	0.11	0.88	0.01	0.11	0.72	0.06	0.03	0.07	0.00	0.02	0.07	0.93	0.00	0.00	0.01	0.88	0.00	0.11	0.00
315	350000	320000	0	1	0.23	0.77	0.00	1.00	0.00	0.11	0.89	0.00	0.15	0.59	0.16	0.00	0.11	0.00	0.00	0.02	0.98	0.00	0.00	0.03	0.82	0.00	0.14	0.00
319	269500	262900	0	1	0.64	0.36	0.00	1.00	0.00	0.01	0.79	0.20	0.09	0.16	0.00	0.57	0.01	0.18	0.00	0.27	0.73	0.00	0.00	0.00	0.02	0.00	0.98	0.00
324	334300	372300	0	1	0.69	0.31	0.00	1.00	0.00	0.18	0.77	0.06	0.15	0.32	0.00	0.47	0.00	0.00	0.06	0.00	1.00	0.00	0.00	0.02	0.08	0.00	0.90	0.00
327	300400	207200	1	0	0.91	0.09	1.00	0.00	0.00	0.91	0.04	0.05	0.94	0.04	0.00	0.00	0.00	0.02	0.00	0.85	0.15	0.00	0.00	0.57	0.00	0.00	0.43	0.00
342	310000	200000	0	1	0.03	0.97	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
350	360700	254400	0	1	0.08	0.92	0.00	1.00	0.00	0.01	0.99	0.01	0.00	0.96	0.00	0.00	0.03	0.00	0.02	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
351	350700	256800	0	1	0.15	0.85	0.00	1.00	0.00	0.01	0.99	0.00	0.00	0.03	0.02	0.70	0.25	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.42	0.00	0.58	0.00
354	305800	230000	0	1	0.07	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.03	0.95	0.02	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.00	0.05	0.00
356	254600	247300	0	1	0.11	0.89	0.00	1.00	0.00	0.01	0.98	0.01	0.00	0.00	0.02	0.06	0.91	0.01	0.01	0.01	0.99	0.00	0.00	0.00	0.94	0.00	0.05	0.00
374	310400	174500	1	0	0.85	0.15	1.00	0.00	0.00	0.73	0.26	0.01	0.00	0.12	0.82	0.00	0.03	0.00	0.03	0.99	0.01	0.00	0.00	0.66	0.00	0.00	0.34	0.00

393	330000	200000	0	1	0.18	0.82	0.00	1.00	0.00	0.04	0.96	0.00	0.00	0.00	0.00	0.55	0.45	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.59	0.00	0.41	0.00
397	376500	260000	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.05	0.00	0.00	0.95	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
408	322000	271000	0	1	0.06	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
409	330400	360200	0	1	0.05	0.95	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.25	0.00	0.00	0.75	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
410	197500	194400	1	0	0.96	0.05	1.00	0.00	0.00	0.99	0.00	0.01	0.97	0.01	0.02	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.93	0.00	0.00	0.08	0.00
412	353000	317000	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.51	0.00	0.00	0.50	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
413	263800	275600	1	0	0.71	0.29	1.00	0.00	0.00	0.31	0.70	0.00	0.30	0.03	0.00	0.61	0.07	0.00	0.00	0.34	0.66	0.00	0.00	0.14	0.04	0.00	0.82	0.00
415	384300	202900	0	1	0.15	0.85	0.00	1.00	0.00	0.08	0.92	0.00	0.01	0.00	0.10	0.00	0.89	0.00	0.00	0.01	0.99	0.00	0.00	0.02	0.89	0.00	0.09	0.00
417	235600	334400	1	0	0.78	0.22	1.00	0.00	0.00	0.63	0.37	0.00	0.20	0.25	0.25	0.30	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.60	0.01	0.00	0.40	0.00
425	217900	243600	1	0	0.69	0.31	1.00	0.00	0.00	0.77	0.23	0.00	0.42	0.13	0.34	0.00	0.12	0.00	0.00	1.00	0.00	0.00	0.00	0.94	0.05	0.00	0.01	0.00
433	399300	219800	1	0	0.97	0.03	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.51	0.49	0.00	0.00	0.47	0.00	0.53
434	302700	266200	0	1	0.18	0.82	0.00	1.00	0.00	0.02	0.97	0.01	0.00	0.00	0.11	0.00	0.88	0.00	0.01	0.00	1.00	0.00	0.00	0.00	0.96	0.00	0.03	0.00
436	280000	230000	1	0	0.91	0.09	1.00	0.00	0.00	0.91	0.09	0.00	0.18	0.00	0.30	0.52	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.33	0.00	0.00	0.67	0.00
441	301700	251100	1	0	0.97	0.03	0.00	0.00	1.00	0.04	0.01	0.95	0.01	0.01	0.00	0.00	0.00	0.01	0.97	0.00	0.00	0.48	0.52	0.00	0.00	0.48	0.00	0.52
442	266400	225600	0	1	0.38	0.62	0.00	1.00	0.00	0.45	0.55	0.00	0.34	0.00	0.02	0.04	0.61	0.00	0.00	0.99	0.01	0.00	0.00	0.14	0.49	0.00	0.37	0.00
446	275200	305300	0	1	0.10	0.90	0.00	1.00	0.00	0.00	0.94	0.06	0.00	0.76	0.00	0.17	0.02	0.05	0.01	0.00	1.00	0.00	0.00	0.00	0.81	0.00	0.19	0.00
456	332600	247200	0	1	0.23	0.77	0.00	1.00	0.00	0.11	0.89	0.00	0.00	0.92	0.07	0.00	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.06	0.88	0.00	0.06	0.00
457	380000	240000	0	1	0.13	0.87	0.00	1.00	0.00	0.00	0.98	0.02	0.00	0.08	0.00	0.85	0.08	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.32	0.00	0.68	0.00
459	329900	361900	0	1	0.06	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.96	0.00	0.04	0.00
463	229200	229700	0	1	0.24	0.76	0.00	1.00	0.00	0.02	0.97	0.01	0.00	0.10	0.01	0.61	0.24	0.00	0.04	0.41	0.59	0.00	0.00	0.01	0.50	0.00	0.49	0.00
466	271300	329800	0	1	0.07	0.93	0.00	1.00	0.00	0.00	0.99	0.01	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.92	0.00	0.08	0.00
467	275800	319800	0	1	0.09	0.91	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.75	0.00	0.07	0.18	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.82	0.00	0.18	0.00
468	246600	352100	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
469	262500	314800	0	1	0.17	0.83	0.00	1.00	0.00	0.04	0.97	0.00	0.00	0.27	0.03	0.38	0.33	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.69	0.00	0.31	0.00
477	375900	327700	0	1	0.05	0.95	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
479	384800	240900	0	1	0.26	0.74	0.00	1.00	0.00	0.01	0.98	0.02	0.00	0.06	0.00	0.93	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.04	0.00	0.96	0.00
485	219000	328200	0	1	0.30	0.70	0.00	1.00	0.00	0.24	0.76	0.00	0.00	0.08	0.10	0.69	0.13	0.00	0.00	0.00	1.00	0.00	0.00	0.03	0.27	0.00	0.70	0.00
511	238400	249800	1	0	0.95	0.05	1.00	0.00	0.00	0.82	0.10	0.08	0.00	0.00	0.00	0.79	0.13	0.00	0.05	0.03	1.00	0.00	0.00	0.00	0.30	0.00	0.70	0.00
513	336300	300400	0	1	0.09	0.91	0.00	1.00	0.00	0.04	0.96	0.00	0.05	0.03	0.00	0.00	0.92	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
519	236200	334800	0	1	0.09	0.91	0.00	1.00	0.00	0.01	0.99	0.00	0.00	0.03	0.12	0.00	0.85	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00
523	334700	291300	0	1	0.06	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.39	0.00	0.00	0.61	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
528	308600	281500	0	1	0.41	0.59	1.00	0.00	0.00	0.17	0.82	0.01	0.02	0.91	0.01	0.04	0.01	0.02	0.00	0.01	0.99	0.00	0.00	0.04	0.67	0.00	0.29	0.00
529	324500	247400	0	1	0.21	0.79	0.00	1.00	0.00	0.07	0.93	0.01	0.19	0.00	0.00	0.00	0.80	0.01	0.00	0.00	1.00	0.00	0.00	0.00	0.84	0.00	0.16	0.00
531	331300	190300	0	1	0.07	0.93	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.79	0.00	0.01	0.18	0.02	0.00	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.03	0.00
534	252400	341500	0	1	0.43	0.58	0.00	1.00	0.00	0.31	0.70	0.00	0.14	0.01	0.15	0.37	0.33	0.00	0.00	0.00	1.00	0.00	0.00	0.08	0.33	0.00	0.59	0.00
535	304820	348670	0	1	0.53	0.47	0.00	1.00	0.00	0.26	0.73	0.01	0.01	0.14	0.10	0.70	0.03	0.00	0.02	0.00	1.00	0.00	0.00	0.09	0.15	0.00	0.76	0.00
536	304000	348000	0	1	0.29	0.71	0.00	1.00	0.00	0.18	0.83	0.00	0.02	0.86	0.07	0.01	0.00	0.00	0.04	0.00	1.00	0.00	0.00	0.00	0.49	0.00	0.51	0.00
541	320000	362000	1	0	0.73	0.27	1.00	0.00	0.00	0.58	0.41	0.01	0.23	0.01	0.21	0.49	0.05	0.00	0.00	0.10	0.90	0.00	0.00	0.29	0.03	0.00	0.68	0.00
542	250700	362500	0	1	0.07	0.93	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.00	0.05	0.00
545	237003	213491	1	0	0.95	0.05	1.00	0.00	0.00	0.99	0.00	0.01	0.92	0.00	0.06	0.00	0.00	0.00	0.02	1.00	0.00	0.00	0.00	0.96	0.00	0.00	0.04	0.00
548	380100	246400	0	1	0.11	0.89	0.00	1.00	0.00	0.07	0.93	0.00	0.00	0.00	0.17	0.00	0.83	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
554	250800	187700	1	0	0.90	0.10	1.00	0.00	0.00	0.90	0.01	0.09	0.91	0.00	0.00	0.00	0.01	0.07	0.01	1.00	0.00	0.00	0.00	0.72	0.00	0.00	0.28	0.00
557	260800	282400	1	0	0.96	0.04	1.00	0.00	0.00	1.00	0.00	0.00	0.93	0.00	0.07	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.95	0.00	0.00	0.05	0.00
562	327400	296500	0	1	0.08	0.92	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.88	0.00	0.10	0.02	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.93	0.00	0.07	0.00
564	238300	335300	0	1	0.10	0.90	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.78	0.00	0.04	0.18	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.96	0.00	0.04	0.00
569	325000	270600	0	1	0.15	0.85	0.00	1.00	0.00	0.18	0.83	0.00	0.06	0.07	0.01	0.00	0.87	0.00	0.00	0.00	1.00	0.00	0.00	0.03	0.95	0.00	0.02	0.00
571	372200	255900	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
572	352200	241900	0	1	0.08	0.92	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.07	0.00	0.00	0.93	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.93	0.00	0.08	0.00
573	267550	294300	0	1	0.75	0.25	0.00	1.00	0.00	0.03	0.81	0.16	0.03	0.11	0.00	0.74	0.00	0.01	0.11	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
581	262200	288400	0	1	0.18	0.82	0.00	1.00	0.00	0.15	0.81	0.05	0.16	0.81	0.00	0.01	0.00	0.00	0.02	0.01	0.99	0.00	0.00	0.01	0.68	0.00	0.32	0.00
582	310500	291800	0	1	0.68	0.32	0.00	1.00	0.00	0.39	0.54	0.08	0.39	0.00	0.00	0.11	0.43	0.00	0.06	0.01	0.99	0.00	0.00	0.16	0.47	0.00	0.37	0.00
584	350200	244100	0	1	0.41	0.59	0.00	1.00	0.00	0.06	0.95	0.00	0.07	0.00	0.00	0.86	0.07	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.08	0.00	0.92	0.00
593	212600	205200	1	0	0.96	0.04	1.00	0.00	0.00	1.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.00	0.03	0.00
594	216270	207030	1	0	0.92	0.08	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
597	257170	275080	1	0	0.93	0.07	1.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.96	0.00	0.00	0.04	0.00
601	262010	281050	1	0	0.79	0.21	1.00	0.00	0.00	0.71	0.29	0.00	0.62	0.09	0.03	0.26	0.01	0.00	0.00	0.70	0.29	0.00	0.00	0.20	0.01	0.00	0.78	0.00
603	313100	371500	1	0	0.77	0.23	1.00	0.00	0.00	0.37	0.64	0.00	0.05	0.27	0.36	0.32	0.00	0.00	0.00	0.05	0.95	0.00	0.00	0.28	0.05	0.00	0.67	0.00
604	265200	351500	0	1	0.16	0.84	0.00	1.00	0.00	0.03	0.96	0.01	0.00	0.82	0.00	0.17	0.00	0.00	0.01	0.00	1.00	0.00	0.00	0.00	0.82	0.00	0.18	0.00
605	287500	331200	0	1	0.67	0.33	1.00	0.00	0.00	0.38	0.61	0.01	0.03	0.76	0.20	0.01	0.00	0.00	0.00	0.03	0.97	0.00	0.00	0.19	0.10	0.00	0.71	0.00

610	361200	305400	0	1	0.10	0.90	0.00	1.00	0.00	0.01	0.99	0.00	0.00	0.00	0.03	0.00	0.97	0.00	0.01	0.00	1.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00
611	268000	277000	0	1	0.67	0.33	0.00	1.00	0.00	0.20	0.78	0.02	0.01	0.10	0.09	0.72	0.09	0.00	0.00	0.13	0.88	0.00	0.00	0.01	0.02	0.00	0.96	0.00
612	320000	212200	0	1	0.63	0.37	1.00	0.00	0.00	0.22	0.78	0.00	0.56	0.13	0.01	0.02	0.29	0.00	0.00	0.02	0.98	0.00	0.00	0.14	0.41	0.00	0.46	0.00
614	367300	222700	0	1	0.13	0.87	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.52	0.02	0.17	0.30	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.88	0.00	0.12	0.00
619	266000	238620	1	0	0.94	0.06	1.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.01	0.00	1.00	0.00	0.00	0.00	0.81	0.00	0.00	0.19	0.00
625	303800	251700	1	0	0.94	0.06	1.00	0.00	0.00	0.92	0.03	0.06	0.02	0.00	0.15	0.82	0.00	0.02	0.00	0.87	0.13	0.00	0.00	0.13	0.00	0.00	0.87	0.00
627	343500	238600	0	1	0.06	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.03	0.00	0.01	0.96	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00
630	264380	280510	0	1	0.54	0.46	0.00	1.00	0.00	0.33	0.66	0.02	0.00	0.00	0.28	0.46	0.25	0.00	0.01	0.12	0.88	0.00	0.00	0.05	0.10	0.00	0.85	0.00
633	254300	368000	0	1	0.28	0.72	0.00	1.00	0.00	0.23	0.78	0.00	0.00	0.68	0.17	0.10	0.05	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.76	0.00	0.23	0.00
634	268200	341400	0	1	0.63	0.37	0.00	1.00	0.00	0.14	0.85	0.01	0.35	0.60	0.00	0.03	0.00	0.01	0.00	0.00	1.00	0.00	0.00	0.04	0.14	0.00	0.82	0.00
635	322500	216700	0	1	0.05	0.95	0.00	1.00	0.00	0.00	0.99	0.01	0.00	0.24	0.00	0.01	0.71	0.04	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
637	316500	235700	0	1	0.48	0.52	0.00	1.00	0.00	0.24	0.75	0.01	0.23	0.00	0.01	0.31	0.43	0.02	0.00	0.00	1.00	0.00	0.00	0.02	0.33	0.00	0.65	0.00
641	303800	251700	1	0	0.66	0.34	1.00	0.00	0.00	0.33	0.66	0.02	0.01	0.03	0.10	0.76	0.09	0.02	0.00	0.87	0.13	0.00	0.00	0.13	0.00	0.00	0.87	0.00
642	330700	262500	1	0	0.90	0.10	1.00	0.00	0.00	0.87	0.12	0.02	0.97	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.99	0.00	0.00	0.78	0.00	0.00	0.22	0.00
644	352590	191430	0	1	0.24	0.76	0.00	1.00	0.00	0.00	0.94	0.06	0.00	0.01	0.00	0.01	0.93	0.02	0.03	0.00	1.00	0.00	0.00	0.00	0.96	0.00	0.04	0.00
645	340400	202000	0	1	0.07	0.93	0.00	1.00	0.00	0.00	1.00	0.00	0.02	0.02	0.00	0.08	0.88	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
646	340850	254000	0	1	0.14	0.86	0.00	1.00	0.00	0.00	0.96	0.04	0.00	0.02	0.00	0.05	0.89	0.04	0.00	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.03	0.00
648	330454	220539	0	1	0.43	0.57	1.00	0.00	0.00	0.14	0.84	0.03	0.02	0.11	0.02	0.64	0.19	0.01	0.02	0.00	1.00	0.00	0.00	0.00	0.28	0.00	0.71	0.00
649	341220	203250	0	1	0.14	0.86	0.00	1.00	0.00	0.11	0.90	0.00	0.00	0.77	0.19	0.00	0.05	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.92	0.00	0.08	0.00
660	307400	246800	0	1	0.15	0.85	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.08	0.00	0.44	0.48	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
661	307400	246800	0	1	0.05	0.95	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
666	263000	282000	0	1	0.67	0.33	0.00	1.00	0.00	0.22	0.72	0.07	0.03	0.00	0.00	0.87	0.07	0.03	0.00	0.15	0.85	0.00	0.00	0.00	0.01	0.00	0.99	0.00
667	322800	216200	0	1	0.16	0.84	0.00	1.00	0.00	0.00	0.96	0.04	0.00	0.17	0.00	0.00	0.78	0.00	0.05	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
669	329500	253850	0	1	0.11	0.89	0.00	1.00	0.00	0.00	1.00	0.00	0.02	0.00	0.04	0.00	0.95	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.03	0.00
670	281020	322360	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.85	0.00	0.00	0.15	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
671	261860	338660	0	1	0.45	0.55	0.00	1.00	0.00	0.19	0.82	0.00	0.02	0.95	0.00	0.03	0.00	0.00	0.01	0.00	1.00	0.00	0.00	0.02	0.24	0.00	0.74	0.00
672	290600	350500	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.01	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
673	291200	277400	0	1	0.47	0.53	0.00	1.00	0.00	0.21	0.77	0.02	0.52	0.26	0.02	0.02	0.13	0.05	0.00	0.03	0.97	0.00	0.00	0.05	0.43	0.00	0.52	0.00
678	241200	213300	0	1	0.49	0.51	0.00	1.00	0.00	0.61	0.39	0.00	0.39	0.00	0.04	0.00	0.58	0.00	0.00	1.00	0.00	0.00	0.00	0.45	0.39	0.00	0.17	0.00
680	304000	228500	0	1	0.50	0.50	0.00	1.00	0.00	0.16	0.81	0.03	0.49	0.14	0.00	0.05	0.30	0.01	0.00	0.00	1.00	0.00	0.00	0.08	0.19	0.00	0.72	0.00
684	303200	232200	0	1	0.05	0.95	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.75	0.00	0.00	0.25	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00
685	302300	228700	0	1	0.05	0.95	0.00	1.00	0.00	0.00	0.99	0.01	0.00	0.02	0.00	0.00	0.94	0.04	0.01	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
686	363600	223300	0	1	0.46	0.54	1.00	0.00	0.00	0.46	0.54	0.00	0.00	0.00	0.69	0.00	0.31	0.00	0.00	0.00	1.00	0.00	0.00	0.62	0.34	0.00	0.04	0.00
691	283700	212240	1	0	0.87	0.13	1.00	0.00	0.00	0.96	0.04	0.00	0.62	0.04	0.19	0.08	0.08	0.00	0.00	1.00	0.00	0.00	0.00	0.86	0.00	0.00	0.14	0.00
702	353600	250900	1	0	0.74	0.26	1.00	0.00	0.00	0.49	0.46	0.05	0.00	0.00	0.66	0.04	0.25	0.01	0.04	0.00	1.00	0.00	0.00	0.43	0.06	0.00	0.51	0.00
704	372300	226400	0	1	0.09	0.91	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.02	0.06	0.92	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.94	0.00	0.05	0.00
705	254600	260000	0	1	0.04	0.96	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.86	0.00	0.00	0.15	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
706	256500	365700	0	1	0.27	0.73	0.00	1.00	0.00	0.19	0.80	0.01	0.05	0.89	0.00	0.01	0.00	0.00	0.04	0.00	1.00	0.00	0.00	0.04	0.44	0.00	0.52	0.00
708	259700	335300	0	1	0.06	0.94	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00
710	294300	378600	0	1	0.22	0.78	0.00	1.00	0.00	0.10	0.90	0.00	0.00	0.85	0.15	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.06	0.86	0.00	0.08	0.00
715	358400	218800	0	1	0.08	0.92	0.00	1.00	0.00	0.00	1.00	0.00	0.01	0.74	0.02	0.00	0.23	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.03	0.00
722	289800	312400	0	1	0.36	0.64	0.00	1.00	0.00	0.02	0.91	0.07	0.00	0.62	0.00	0.30	0.02	0.03	0.03	0.00	1.00	0.00	0.00	0.01	0.33	0.00	0.66	0.00
723	303800	318300	1	0	0.87	0.13	1.00	0.00	0.00	0.80	0.21	0.00	0.94	0.00	0.01	0.05	0.00	0.00	0.00	0.26	0.74	0.00	0.00	0.50	0.01	0.00	0.49	0.00
725	230600	328700	0	1	0.42	0.58	0.00	1.00	0.00	0.36	0.64	0.00	0.06	0.13	0.15	0.61	0.06	0.00	0.00	0.01	0.99	0.00	0.00	0.11	0.27	0.00	0.62	0.00
726	282700	363800	0	1	0.35	0.65	0.00	1.00	0.00	0.17	0.83	0.00	0.02	0.92	0.03	0.03	0.00	0.00	0.01	0.00	1.00	0.00	0.00	0.03	0.32	0.00	0.65	0.00
730	305600	374930	1	0	0.88	0.12	1.00	0.00	0.00	0.52	0.48	0.00	0.75	0.00	0.02	0.23	0.00	0.00	0.00	0.05	0.95	0.00	0.00	0.28	0.00	0.00	0.71	0.00
732	368300	214000	0	1	0.08	0.92	0.00	1.00	0.00	0.02	0.98	0.00	0.00	0.93	0.03	0.00	0.04	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
733	313600	181380	0	1	0.13	0.87	0.00	1.00	0.00	0.01	0.99	0.00	0.02	0.00	0.00	0.00	0.98	0.00	0.00	0.15	0.85	0.00	0.00	0.01	0.98	0.00	0.02	0.00
735	343600	309100	0	1	0.11	0.89	0.00	1.00	0.00	0.07	0.93	0.00	0.05	0.04	0.00	0.00	0.91	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.96	0.00	0.03	0.00
736	356500	321300	0	1	0.38	0.62	0.00	1.00	0.00	0.41	0.60	0.00	0.52	0.15	0.00	0.00	0.33	0.00	0.00	0.03	0.97	0.00	0.00	0.15	0.51	0.00	0.34	0.00
751	361200	201300	0	1	0.29	0.71	0.00	1.00	0.00	0.28	0.72	0.00	0.19	0.00	0.02	0.01	0.77	0.00	0.01	0.00	1.00	0.00	0.00	0.03	0.51	0.00	0.45	0.00
584A	350200	244100	0	1	0.51	0.49	1.00	0.00	0.00	0.36	0.64	0.00	0.70	0.01	0.00	0.01	0.28	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.08	0.00	0.92	0.00

Appendix 3.1. GENELAND SPATIAL K=9. The four main sub regions identified in GENELAND SPATIAL K = 6 (Chapter 3; Figure 6) were further divided into 2 clusters each. The Shetland Isles individual is shown here as 1 population.



Appendix 3.2. Clusters identified by STRUCTURE for each progressive partition and the plots of population assignment for each K=2 (populations and graphs are colour coded).

Appendix 3.2 continued: Detecting Clinal Variation

There has been much debate about the effect of isolation by distance (IBD) (Wright 1943), and whether Bayesian clustering programs are actually identifying true clusters or are artificially detecting structures emerging from uneven sampling along a cline (Serre & Pääbo 2004; Rosenberg *et al.* 2005; Frantz *et al.* 2009).

The authors of STRUCTURE concede that there may be difficulties detecting structures if IBD is present (Pritchard & Wen 2003). To combat this Frantz *et al.* (2006) recommend using spatial data in the analysis and found that genetic clusters identified by BAPS 4.1 spatial were the most biologically meaningful out of three models tested, and that the model was robust when faced with isolation-by-distance relationships in the genetic data set. Chen *et al.* (2007) found that all the Bayesian clustering methods they tested that included spatial data as a prior could identify a cline, however they found STRUCTURE, despite not incorporating spatial data, showed the best estimation of a cline to the actual allele frequencies. The ability of the program to identify true clusters and not artificial clusters along a cline can also depend on study design and the number of markers (Corander *et al.* 2004; Swhartz *et al.* 2008; Rosenburg *et al.* 2005; Serre & Pääbo 2004). Frantz *et al.* (2009) recommend caution when interpreting results of populations characterised by IBD as this can lead to an overestimation of genetic structure and the identification of erroneous population units.

In Chapter 2 I investigated the possibility that the clusters identified may not be populations but are actually artificial structures produced by clines in allele frequencies. Evidence of clinal variation can be identified by plotting the population adherence sorted by Q (the most likely population for any individual) (Sahlsten *et al.* 2008). Clinal variation may not be possible to identify with only one run, and the average of multiple runs must be used to detect this type of variation (Chen *et al.* 2007). Clinal variation was investigated by using the average population membership coefficients created in CLUMPP and displaying them using Microsoft EXCEL.

The progressive partitioning method used in Chapter 3 lends itself well to identifying clines using this approach. As it splits the dataset into two each time, for each sub-division the average assignments of individuals to the two populations can be plotted and sorted to give an indication of admixture.

Despite the risk that by using a progressive partitioning approach the more you split the populations the more chance you have of finding partitions along a cline (Frantz *et al.* 2009). The progressive partitioning approach may actually provide the researcher with a better understanding of the gene flow between populations.

Since STRUCTURE has no spatial priors the results of the STRUCTURE progressive partitioning were used here (figure 3.2) to produce an unbiased estimate of admixture and look for evidence of clines, using the plots of individual assignment to a cluster at $K = 2$.

Figure 3.2 shows the clusters identified by STRUCTURE for each progressive partition and also the displays the plot of population assignment for each K=2 (populations and graphs are colour coded). At greater degrees of genetic differentiation (early partitions) the partitions are more distinct, i.e. individuals are assigned to one population or another with little/ no admixture. As the sub divisions progress the genetic differentiation between the $K = 2$ clusters decreases and the amount of admixture increases, showing individuals with different degrees of admixture. Clusters showed no more subdivision when average population assignments were approximately 0.5 (Plots in figure 3.2 representing half colour of population and half black).

Those individuals that show admixture in the initial $K = 2$ split all show the same amount of admixture and are a product of the fact that the split is constrained to $K = 2$. Individuals from a third population are forced to be assigned to one population or another, since these histograms are produced by averaging the assignment of multiple runs this third population was assigned to either cluster 1 or 2 in different runs therefore the average indicates admixture.

If there is no mixture of admixture levels present, the researcher can be confident that there is little possibility of the clusters being identified along a cline and produced as a result of IBD, if there is admixture present this could be a sign of IBD or even restricted gene flow between populations as a result of landscape features (IBED).

Appendix 4. Summary statistics of geographic distances between individuals within population groupings.

To further understand the distances between individuals within the population groupings summary statistics of the geographic distances between individuals for each population are given in Appendix 4.1.

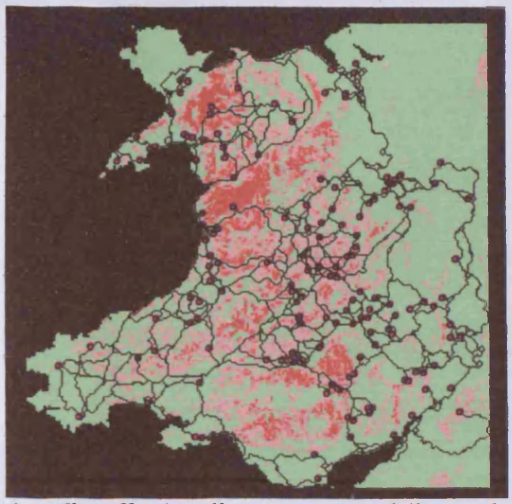
Appendix 4.1: Summary statistics of distance (km) between individuals within each population grouping. The x-intercepts (neighbourhood size) identified through autocorrelation analysis is also added for reference.

	1) Southwest Wales	2) Mid-eastern Wales	3) Northwest Wales	4) Southwest and northwest Wales	5) Mid- eastern and northwest Wales	6) Southwest and mid-eastern Wales	7) Wales and Borders region
0 (min)	0	0	0	0	0	0	0
1st quartile	37.3	45.6	20.3	46.3	49.8	50.8	53.0
2nd quartile (median)	61.7	72.3	36.3	79.0	80.2	81.3	84.3
3rd quartile	95.0	105.9	61.4	121.3	116.2	116.9	119.7
4th quartile (max)	223.8	222.7	110.6	224.5	222.7	231.2	228.5
Average	69.6	77.5	40.9	84.8	84.3	85.6	87.5
X-intercept spatial autocorrelation	48.9	59.0	38.3	70.5	84.2	74.9	82.2

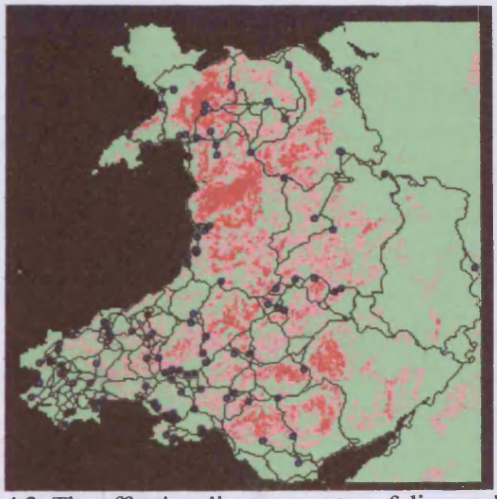
Appendix 4.2- 4.15 The effective distance routes of dispersal between individuals for the population groupings with landscape features that showed significantly correlated *r* values.

Mantel tests were used to identify correlations between genetic distance and the effective distance matrices produced by resistance surfaces of landscape features. Taking into account the fact those individuals that are geographically close can be expected to be genetically relatively similar, partial Mantel tests were also conducted to control for this effect of geographic distance. For each population grouping, the effective distances that produced the highest significantly correlated *r* values (Chapter 3: Tables 4-8) were plotted on maps of that landscape features (Figures 4.2-4.15). Figures 4.2- 4.15 show the landscape feature, resistance value of that landscape feature and the population grouping.

4.2: The effective distance routes of dispersal between individuals for southwest Wales sub region (1). Resistance-to-movement surface for slope [(% slope/km^2) ^{2}].



4.4: The effective distance routes of dispersal between individuals for Mid-eastern and northwest Wales population grouping (5). Resistance-to-movement surface for slope [(% slope/km^2) ^{2}].



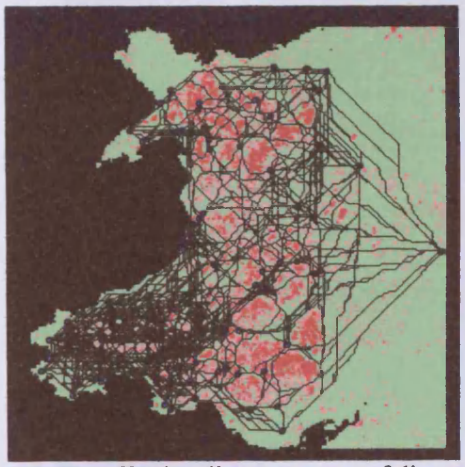
4.3: The effective distance routes of dispersal between individuals for southwest and northwest Wales population grouping (4). Resistance-to-movement surface for slope [(% slope/km^2) ^{2}].

4.5 : The effective distance routes of dispersal between individuals for southwest and mid-eastern population grouping (6). Resistance-to-movement surface for slope [(% slope/km^2) ^{2}].

4.6: The effective distance routes of dispersal between individuals for Wales region (7). Resistance-to-movement surface for slope [(% slope/km^2) ^{2}].

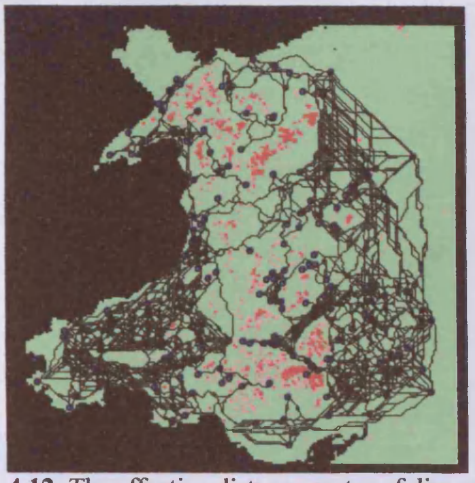
4.8: The effective distance routes of dispersal between individuals for mid-eastern Wales sub-region (3). Resistance-to-movement surface for upland habitat (% upland cover/km^2) ^{2}].

4.7: Anthropogenic map and all individuals from all populations



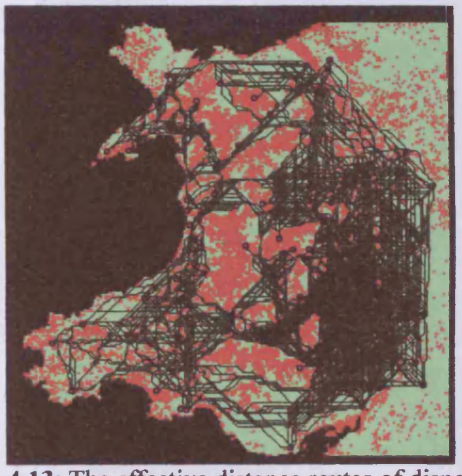
4.9: The effective distance routes of dispersal between individuals for southwest and northwest Wales population grouping (4). Resistance-to-movement surface for upland habitat [(% upland cover/km^2) x (10)].

4.10: The effective distance routes of dispersal between individuals for mid-eastern and northwest Wales population grouping (5). Resistance-to-movement surface for upland habitat [(% upland land cover/km^2) x (1000)].

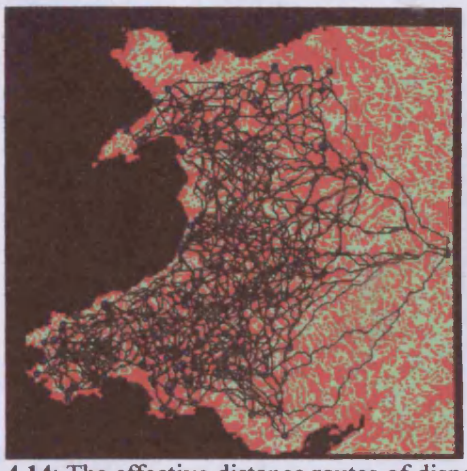


4.12: The effective distance routes of dispersal between individuals for Wales region (7). Resistance-to-movement surface for upland habitat [(% upland cover/km^2) ^{3}].

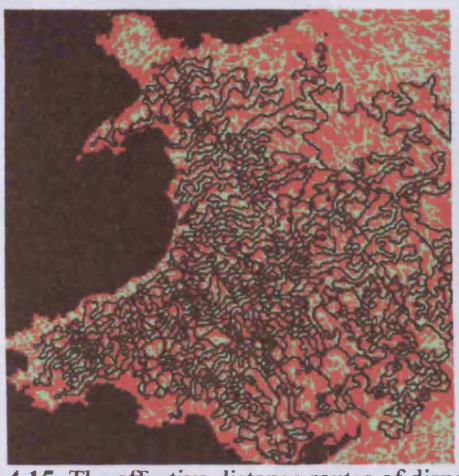
4.11: The effective distance routes of dispersal between individuals for southwest and mid-eastern sub-regions combined population grouping (6). Resistance-to-movement surface for upland habitat [(% upland cover/km^2) ^{2}].



4.13: The effective distance routes of dispersal between individuals for Mideast and north Wales population grouping (5). Resistance-to-movement surface for Broadleaf habitat [(% Broadleaf cover/km^2) x (0.03)].



4.14: The effective distance routes of dispersal between individuals for southwest and north Wales population grouping (4). Resistance-to-movement surface for rivers [(% river cover/km^2) x (0.3)].



4.15: The effective distance routes of dispersal between individuals for Mideast and north Wales population grouping (5). Resistance-to-movement surface for rivers[(% river cover/km x (0.0001)].

Appendix 5

Hobbs et al. (2009) Landscape Genetics applied to a recovering otter (*Lutra lutra)* population in the UK: Preliminary results and potential methodologies. *Hystrix Journal of Mammalogy*. **17**, 47-63.

Hystrix It. J. Mamm (n.s.) 17 (1) (2006): 47-63

LANDSCAPE GENETICS APPLIED TO A RECOVERING OTTER (*LUTRA LUTRA*) POPULATION IN THE UK: PRELIMINARY RESULTS AND POTENTIAL METHODOLOGIES

GEOFFREY I. HOBBS[a], ELIZABETH A. CHADWICK, FRED M. SLATER, MICHAEL W. BRUFORD

Cardiff University, CF10 3TL, Wales, UK
[a]Corresponding author: Hobbsgi@cardiff.ac.uk

ABSTRACT - The Eurasian otter (*Lutra lutra*) has declined significantly across its European range. In the UK, the decline was particularly severe during the late 1950's and early 1960's, and by the mid 1970's the population was largely confined to strongholds in parts of Scotland, Northern Ireland, mid and West Wales and south west England. In recent years the otter population has started to recover, with otter surveys confirming an increased distribution of otters in Wales, Scotland and England. In England, population expansion and recolonisation is believed to be occurring both through breeding and by dispersal, from the west (south west England and the Welsh borders) and from the north (Scotland). However, little is known about the degree of genetic loss due to the decline, potential barriers to recolonisation, routes of dispersal, or the contribution of reintroduction programmes to population increases. This project aims to use tissues collected since 1994 (complete with geographic location) from over 500 otters found dead on roads in Wales and England, to analyse the genetic diversity and structure of otter populations. Using molecular genetic analysis of the otter population, we will identify whether and when bottlenecks occurred, whether population decline has resulted in a loss of genetic variability, and to what degree. Preliminary analysis from 177 otters has shown that observed is generally lower than expected heterozygosity, and that the population is in Hardy Weinberg equilibrium for 11 out of the 15 loci. Spatial patterns in genetic data will be analysed, to identify clines, isolation by distance and genetic boundaries to gene flow, the contribution of released animals will also be assessed. Geographical information systems (GIS) will be used to map spatial genetic patterns and to generate hypotheses about the potential cause of genetic boundaries such as landscape or environmental features.

Key words: *Lutra lutra*, microsatellites, spatial genetic patterns, barriers to dispersal, genetic variation

RIASSUNTO – ***La genetica del paesaggio applicata allo studio di una popolazione di Lontra* (Lutra lutra) *in fase di espansione nel Regno Unito: risultati preliminari e metodologie potenziali.*** Le popolazioni di Lontra (*Lutra lutra*) sono significativamente diminuite in tutto il loro areale europeo. Nel Regno Unito il declino è stato particolarmente severo nell'ultima parte degli anni '50 ed all'inizio degli anni '60, e a metà degli anni '70 la popolazione era sostanzialmente confinata in alcune aree della Scozia, nord Irlanda, Galles centrale ed occidentale, ed Inghilterra del sud-ovest. In anni più recenti la popolazio-

ne di Lontra ha iniziato a ri-espandersi, come indicato dai censimenti che hanno confermato la maggior presenza di lontre in Galles, Scozia ed Inghilterra. L'espansione e la ricolonizzazione in Inghilterra potrebbe essere sostenuta sia dalla riproduzione che dalla dispersione da ovest (dall'Inghilterra del sud-ovest e dai confini col Galles), e da nord (Scozia). Tuttavia si conosce poco circa il declino di diversità genetica dovuto alla contrazione demografica, alle potenziali barriere alla ricolonizzazione, le vie di dispersione, o il contributo dei programmi di reintroduzione all'incremento della popolazione. Scopo di questo progetto è di usare tessuti raccolti fin dal 1994 (completi di localizzazioni geografiche) da più di 500 carcasse di lontre raccolte a seguito di incidenti stradali in Galles ed Inghilterra, per analizzare la diversità genetica e la struttura delle popolazioni. Tramite analisi genetiche molecolari, ci si propone di identificare se e quando si siano stati *bottlenecks*, se il declino della popolazione abbia prodotto perdite di variabilità genetica, ed in quale misura. Analisi preliminari da 177 lontre hanno mostrato che l'eterozigosi osservata è generalmente minore dell'attesa, e che la popolazione è in equilibrio di Hardy-Weinberg per 11 su 15 loci. Si analizzeranno i *patterns* spaziali dei dati genetici, al fine di identificare clini, isolamento per distanza e flusso genico. Si identificherà anche l'eventuale contributo di lontre rilasciate. Sistemi informatici geografici (GIS) saranno usati per mappare i pattern geografici e per generare ipotesi sulle cause potenziali di barriere genetiche quali componenti ambientali o paesaggistiche.

Parole chiave: *Lutra lutra*, microsatelliti, *pattern* genetici spaziali, barriere al *dispersal*, variazione genetica

INTRODUCTION

1. Otter distribution and declines

The Eurasian otter (*Lutra lutra*) is a member of the family Mustelidae and its vast range extends from the west coast of Ireland to Japan and from Arctic Finland to North Africa and Indonesia (Chanin, 1985). The Eurasian otter has declined significantly throughout its European range (Barbosa *et al.*, 2003) and in the UK this occurred particularly during the late 1950's and early 1960's, throughout much of Wales, England and the Scottish borders (Coxon *et al.*, 1999; Conroy and Chanin, 2000; Mason and Macdonald, 2004). By the mid 1970's the UK population was largely confined to strongholds in parts of Scotland, Northern Ireland, mid and west Wales and south west England (Jones and Jones, 2004). There are a number of reasons proposed for this decline, such as a loss of riparian habitat, hunting, water pollution, fish traps, road traffic accidents and general disturbance (Mason and Macdonald, 2004). The most likely factor, given the suddenness of the decline, was the introduction of the organochlorine group of insecticides (particularly dieldrin), and polychlorinated biphenyls (PCBs) (Conroy and Chanin, 2000; Mason and Macdonald, 2004). The suggested combination of factors has contributed to this species being listed as either vulnerable or endangered throughout much of its current range (Ruiz-Olmo *et al.*, 2001).

2. Trends in the recovery of otters

Detailed monitoring programmes have shown that since the late 1970's there has been a slow expansion of the otter population in the UK (Ruiz-Olmo and Delibes, 1998; Conroy and Chanin, 2000), which may be the result of reduced pollution. For example, Mason (1998) shows a decline in the level of PCBs found in otter tissues from England and Wales between 1983 and 1992, to a level that no longer poses a threat to otter populations and thus should no longer act as a constraint on recolonisation. In Wales, otter surveys confirm that there has been an increase in range, with recolonisation rates exceeding Biodiversity Action Plan (BAP) targets (Jones and Jones, 2004). Scotland has also shown signs of recovery, but there are still large areas, particularly of central and southern England, where the species remains absent, or is very rare. Population expansion and recolonisation is believed to be occurring in this area both through breeding and by dispersal, from the west (south west England and the Welsh borders) and from the north (Scotland) (Coxon *et al.*, 1999; Conroy and Chanin, 2000).

3. Otter population fragmentation and its genetic consequences

Little is known about otter ecology and population dynamics in the UK outside Scotland, and organisations such as the Environment Agency have channelled resources into schemes such as habitat enhancement for otter conservation with little knowledge of their long term effectiveness (Coxon *et al.*, 1999). The need for more information about otter populations and recolonisation processes has been recognised by conservation bodies such as the Joint Nature Conservation Committee (JNCC), and incorporated into the UK Otter BAP (Biodiversity Action Plan). Anthropogenic factors have caused habitat fragmentation and a reduction in total habitat area. In most species, habitat fragmentation causes a reduction in population size and increased isolation of populations (Hooftman *et al.*, 2003). Fragmentation can result in reduced migration and gene flow, which can have deleterious effects on genetic diversity, and increase the risk of inbreeding and extinction (Charlesworth and Charlesworth, 1987; Ralls *et al.*, 1988). One of the main goals of conservation should be to mitigate fragmentation of natural habitats to increase population sizes and connectivity (Hooftman *et al.*, 2003).

The JNCC Framework for Otter Conservation in the UK identified the need to assess genetic variation within and between otter populations (Coxon *et al.*, 1999). Dallas *et al.* (2002) studied the genetic structure of the British otter populations using microsatellite markers. They had two major findings, that "*populations in Scotland, regarded as continuous according to distributions of signs, were to some extent genetically subdivided and populations in mainland Scotland showed a strong pattern of isolation by distance (IBD)...*" And "*populations in southern Britain regarded as biologically equivalent to those in Scotland contained significantly reduced levels of*

microsatellite polymorphism".
Statistical assignment tests performed by Dallas *et al.*, (2002) suggest there was no gene flow between populations in Scotland, Wales and SW England at the time of study. The different levels of microsatellite polymorphism were associated mainly with the discontinuity between populations in mainland Scotland, and those in Wales and SW England. It was unclear whether the reduced microsatellite polymorphism in Wales and SW England was the result of recent or long-term population fragmentation (Dallas *et al.*, 2002). It was suspected that the reduced polymorphism reflected a long history of low effective population size rather than recent declines (Dallas *et al.*, 2002). However, assessment of the loss of variability was hampered by lack of information about the genetic composition of the same populations prior to their fragmentation and bottleneck (c.f. Pertoldi *et al.*, 2001).
Pertoldi *et al.* (2001) investigated whether the recent otter population decline in Denmark had resulted in a loss of genetic variability, using samples from the contemporary otter population, and from historical (museum) specimens collected between 1880 and 1960. The otter population in Denmark has experienced a severe population decline in the last four decades, similar to that in the UK. However, analyses of microsatellite DNA variation in the contemporary population showed surprisingly few signs of a recent bottleneck, and indicated that the extant otter population has not suffered a recent severe loss of genetic variability (Pertoldi *et al*, 2001). The study also showed that some geographical subdivision was present in historical specimens. There were indications of a drastic population decline, but this was shown to have had happened on a time scale covering hundreds or thousands of years, not during the last few decades. It was concluded that otter populations, at least those from northern Europe, generally exhibit low genetic variability. The study suggested that the variation in the Danish otters was likely to have been low even before the recent decline in otter populations and was explained either by post-glacial founder events or a decline which started *ca.* 2,000-3,000 years ago. These findings support Dallas *et al.*'s (2002) hypothesis that the low genetic variation found in the otter populations of the UK is the result of historical rather than recent population declines.
It is nonetheless important that the long-term viability of UK otter population is likely to depend upon recolonisation and the establishment of corridors for gene flow between populations. Mitigation should therefore be considered against the potentially negative effects of population fragmentation.

4. Monitoring otter populations

The UK Otter BAP identified the need to monitor populations, distribution of otters and to monitor the expansion of fringe populations to ensure the successful management and conservation of this species (Coxon *et al.*, 1999). However, in addition to its status as an endangered species, which brings with it logistical and ethical problems that

hamper data collection, otters live at low densities and are often nocturnal or crepuscular, so their study is not straightforward (Ruiz-Olmo *et al.*, 2001). As a result, monitoring techniques encounter many difficulties (Ruiz-Olmo *et al.*, 2001).

There have been a handful of studies in which direct, systematic visual observations have been used to gain information about European otter populations (Ruiz-Olmo *et al.*, 2001, Chanin, 2003). These methods involve a large investment of time and experienced personnel, and given the secretive nature of this species, systematic watches have limited value in monitoring otter populations, especially where there is overhanging vegetation (Chanin, 2003). Direct observations using cameras are a possibility, however, the cost and difficulty in getting clear pictures renders this option impractical (Chanin, 2003). Studies have been conducted using radio-tracking, focusing mainly on space use i.e. range sizes and rates of travel (Sjoasen, 1997). This requires the trapping of individuals, which may be problematic due to the low capture rate, small population sizes, or potential for injuries caused by handling (Mills *et al.*, 2000). Radio-tracking has been successful, but is more suited to monitoring introduced and translocated individuals, providing data without the risk associated with trapping wild animals (Sjoasen, 1997). Results of such a study showed that radio-tracked translocated otters spent a high proportion of their time exploring, apparently searching for a suitable area to establish their home ranges away from occupied sites (Sjoasen, 1997).

The most frequently used technique in Europe for detecting the presence, abundance or relative abundance of otters, is to search for spraints (faeces). Otters leave spraints in visible spots (e.g. stones, rocks, tree-trunks) and in predictable places (e.g. under bridges, at junctions of rivers, in basins) which facilitates survey work. This allows the possibility to differentiate between positive and negative sites and to count the number of signs (Ruiz-Olmo *et al.*, 2001; Hung *et al.*, 2004; Prigioni *et al.*, 2005). Over the past 25 years detecting spraints has become the standard survey method and has been used on a large scale for the national surveys of Britain and Ireland (Chanin, 2003).

Mason and Macdonald (2004) tested the method of predicting abundance of otters from spraints, using river catchments where colonisation by otters was assisted by the release of a known number of captive animals. These authors showed that there was a relationship between the number of otters, the number of sprainting sites and the spraint density. Although this method cannot be used to determine the exact number of otters present, it does provides evidence that the number of positive sites and the intensity of sprainting can be used to give a broad estimation of the performance of the otter population.

5. Genetic analysis from non-invasive biological samples

DNA can be recovered from non-invasive samples such as faeces, potentially allowing genetic analysis of otter spraints. Thus the genetic identity of

individuals can be characterised, providing an abundance of information on the population (Chanin, 2003; Dallas *et al.*, 2003; Hung *et al.*, 2004). A positive identification provides the location of an individual at a particular point in space and time, but provides no information on whether it is resident or transient, adult or juvenile. A distinction must be made between areas of frequent use/sedentary presence, and areas through which otters move quickly (Ruiz-Olmo *et al.*, 2001). A pilot study was performed by Coxon *et al.* (1999) in 1997-98. It allowed the identification of a minimum number of individuals within the study area, and repeated identification allowed the calculation of home range size for one of the individuals. To estimate the population size in elusive or rare species, a new technique of mark-recapture using non-invasive genetic sampling (i.e. faeces) has been developed by Miller *et al.*, (2005): the method is implemented through the software package *capwire*. The data generated from this sampling method differ from traditional mark–recapture data in that individuals may be captured multiple times within a session or there may only be a single sampling event. Preliminary studies of this method have shown it provides estimates with small bias and good coverage, along with high accuracy and precision, providing an improved way to estimate N for some DNA-based data sets (Zhan et al, in press).

There are problems associated with the use of spraints. For example, the collection of spraints involves a lot of effort, not only in the field (where it has been calculated that it can take two man hours per spraint) but also in the lab, where analysis can take ten man hours per DNA profile (Chanin, 2003). New techniques for DNA extraction from faeces are, however, reducing the time spent in the lab and improving its success (Chanin, 2003). Another limitation of this technique is the difficulty of obtaining a sufficient quantity and quality of DNA from spraints (Dallas *et al.*, 2003; Hung *et al.*, 2004). If spraints are not collected fresh they may become degraded and unusable (Chanin, 2003). Also, genotyping of DNA from faeces is prone to several problems. Due to the scarcity of the template DNA, stochastic amplification of only one out of two alleles at a heterozygous locus can cause 'allelic dropout'. Artefacts are sometimes generated during amplification to produce a 'false allele', and sometimes a 'counterfeit' or third allele is produced. Contaminant DNA can cause serious problems when the target DNA is rare and may lead to mistyping of the genotype (Huang *et al.*, 2005). These errors need to be detected and resolved and this can mean repeating the DNA amplification independently several times in order to obtain reliable genotypes (Taberlet *et al.*, 1997; Dallas *et al.*, 2003; Hung *et al.*, 2004).

6. Genetic analysis from otter tissue

With an increasing otter population in Britain, the likelihood of an encounter with humans increases. Unfortunately in the last 15 – 20 years, mortality due to road traffic accidents has increased, and has become one of the most important causes of death of otters in most

European countries (Hauer *et al.*, 2002; Philcox *et al.*, 1999). Although unfortunate, where carcasses are collected they provide an ideal source of samples for genetic analysis, because the extraction of DNA from tissue samples is much more reliable than from faeces.

The collection of genetic data from many individuals of known geographic origin, in combination with recently developed statistical tools, potentially allows the identification of spatial genetic patterns (Manel *et al.*, 2003). This approach enables the spatial mapping of allele frequencies and potential correlation with landscape or environmental features. This 'landscape genetic approach' combines landscape ecology with population genetics, allowing the examination of biogeography at a fine spatial and temporal scale. This provides information on the interaction between environmental or landscape features, and microevolutionary processes such as genetic drift, gene flow and selection (Manel *et al.*, 2003; Berthier, 2005). Geographical information systems (GIS) can be used in conjunction with statistical tests to visualise spatial genetic patterns, by overlaying landscape variables and genetic data (Manel *et al.* 2003). An important feature of this approach is that it aids in the identification of cryptic genetic discontinuities (barriers to gene flow) across populations which have no obvious cause and can identify secondary contact between previously isolated populations. Spatial delineation of genetic discontinuities within a species can also allow for the formation of operational units, important for management purposes (Manel *et al.*, 2003).

7. Molecular approaches

Microsatellites consist of tandemly repeated units, generally less than 5bp (base pairs) in length such as (TG)n or (ATT)n (Bruford and Wayne, 1993). These repeat units are often highly polymorphic with many different alleles segregating in a population. Due to their attributes they have been used in many different areas of study ranging from ancient and forensic DNA studies, to population genetics and conservation/management of biological resources (Jarne and Lagoda, 1996; Zhivotosky and Feldman, 1995; Zane *et al.*, 2002). Locus-specific PCR primers are designed to recognise sequences flanking the tandem repeats (Bruford *et al.*, 1996).

8. Background and aims of study

The otter population in England and Wales is known to be growing (Coxon *et al.*, 1999; Conroy and Chanin 2000; Jones and Jones, 2004) but little is known about the dynamics of recolonisation events associated with this expansion. Using genetic data available from otter carcasses found and collected in this area since 1994, the genetic structure of remnant and newly established populations will be investigated. This information can be used to analyse the origin, rate and direction of recolonisation into formerly vacant regions using spatial genetic analysis and population assignment tests (e.g. Piry *et al.*, 2004).

In a 'source-sink' situation such as recolonisation into a vacant habitat, where otters are expected to spread

from stronghold populations, a correlation between genetic and geographic distance from the source can be expected (Bertorelle and Barbujani, 1995), with a continuous increase of genetic distance with geographic distance (isolation by distance). The identification of spatial genetic patterns will show both the degree and direction of spread of the otter population from strongholds to adjacent unpopulated areas, and demonstrate the success and spread of any otters introduced. GIS will be used to visualise spatial genetic patterns and to generate hypotheses about the cause and consequence of genetic boundaries, which can then be explicitly tested.

Our study will concentrate initially on the genetic structure of the Welsh otter population, to identify if genetic differences exist at local and regional levels. If sub-structures do exist, GIS will be used to identify whether genetic boundaries are associated with physical obstacles such as roads and other landscape features. Later in the study, we aim to include English otter populations, again to investigate the genetic structure but also to assess the relative contribution of source populations in Wales, SW England and Scotland. We also aim to use spatial genetic patterns to identify the degree, direction and routes of dispersal as well as identify barriers. Genotype mapping will also demonstrate the origin and success of otters that have been introduced.

PRELIMINARY ANALYSIS

As a first step in this study we have analysed samples from Wales and bordering catchments to establish molecular methodologies, and to examine the genetic structure of the Welsh otter populations.

METHODS

1. Sampling

Over the past two decades in the UK, the Environment Agency along with other regional organisations have recorded the geographical location and collected otter road casualties (over 500 individuals) throughout England and Wales. Muscle samples have been removed from otters and stored in ethanol at -20°C. Of these, 177 samples have been selected from Wales and bordering catchments (Fig. 1), for use in this preliminary analysis.

2. DNA extraction

DNA was extracted from muscle tissue, using the QIAGEN DNeasy tissue kit following the 'isolation of total DNA from animal tissues' protocol (QIAGEN, #65906).

3. Primers

Using primers that have been designed for the Eurasian otter, we identified the genotypes of individuals for 21 loci. The microsatellite loci used comprise lut 435, 453, 457, 604, 615, 701, 715, 717, 782, 818, 832, 833 (Dallas and Piertney, 1998) lut 902 (Dallas *et al.*, 1999) and 04OT02, 04OT04, 04OT05, 04OT07, 04OT14, 04OT17, 04OT19 and 04OT22 (Huang *et al.*, 2005). (Following preliminary analyses, the number of loci will be reduced using rarefaction analysis, see below).

4. Multiplex design

For more efficient analysis, four PCR mul-

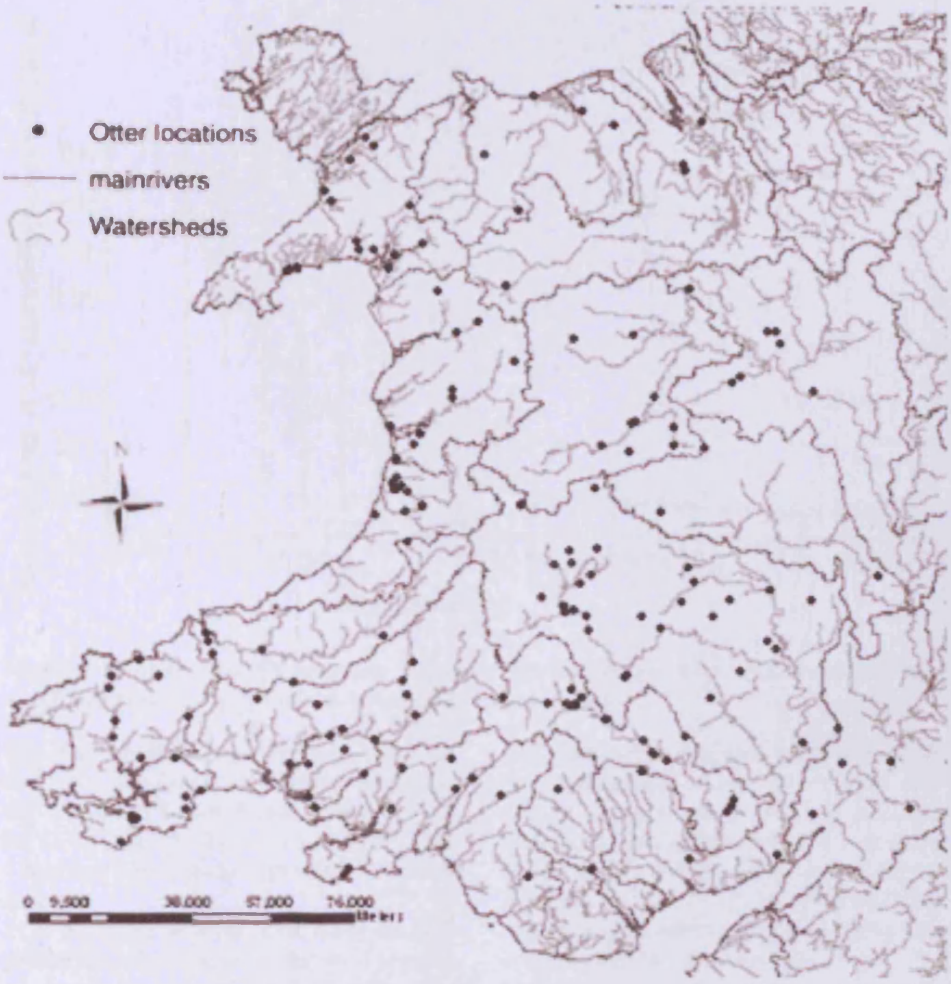


Figure 1 - Map of Wales and Borders showing major rivers, watersheds and otter location.

tiplex groups were designed and optimised. The Forward primers of each primer pair were labelled with a fluorescent dye (Ned, Hex or Fam). The dye used to label each primer was chosen as part of the design of the multiplex group which also took into account the allele size, to ensure that each locus was distinct. Two multiplex groups contained five primer pairs and two contained six. PCR reactions were conducted with a QIAGEN Multiplex PCR kit following the 'amplification of microsatellite loci using multiplex PCR' protocol (QIAGEN, #206143). Amplification of DNA extracts was performed using a GeneAmp® PCR system 9700 (Applied Biosystems) in 6.5 µl reactions containing DNA template, 1x QIAGEN Multiplex PCR Master Mix (containing HotStarTaq® DNA polymerase, Multiplex PCR buffer (contains 3 mM $MgCl_2$) and dNTP Mix), 10x Primer Mix (0.2 µM of each primer) and sterile water.

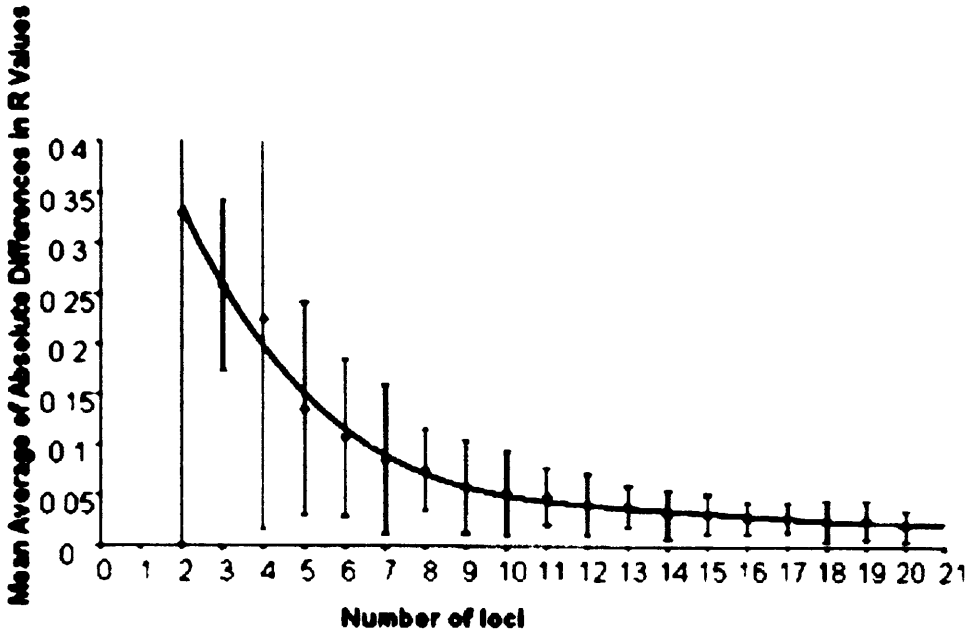


Figure 2 - The decrease in the mean difference between consecutive relatedness estimates as a function of the number of microsatellite loci analysed.

The PCR profile was identical for each multiplex and included an initial denaturation step of 95 °C for 15 min, 29 cycles with 94 °C for 30 s, 58 °C for 90 s and 72 °C for 1 min and a final extension of 60 °C for 30 minutes.

PCR products were analysed using an ABI PRISM® 3100 Genetic Analyser (Applied Biosystems) and gel analysis was performed using the software Genescan v 3.7 and Genotyper version 3.6 (Applied Biosystems).

5. Rarefaction analysis

A random sample of 100 otters from the Wales and Borders region were genotyped for all 21 loci using the methods described above. These genotypes were input into the program POPASSIGN version 4.3a (http://www.darwinfox..org/fulvipes/EnHome.htm) to conduct rarefaction analysis. Rarefaction analysis aims to identify the combination of loci which most efficiently recover data, enabling accurate relatedness and genetic diversity estimation (Kays *et al.*, 2000; Smith *et al.*, 1997; Altmann *et al*, 1996). In POPASSIGN, relatedness is assessed by simulating first order relative datasets based on the observed allele frequencies, estimating 'Queller and Goodnight (1989) relatedness' (*R*) using the simulated data, and repeating the process for all possible combinations of loci to be used. Standard errors are generated by permuting loci without replacement. The number of loci was increased by addition without replacement until all 21 loci were selected (Girman *et al.*, 1997; Kays *et al.*, 2000). This procedure was repeated 1000 times. The mean difference in relatedness estimate *R* for different numbers of loci and jackknifed standard errors were calculated as the average of absolute differ-

ences in *R* values calculated between steps (Altmann *et al.*, 1996).

6. Genetic variability

Genotyping using 15 loci (the optimal combination identified by rarefaction analysis; lut435, lut453, lut717, lut604, lut733, lut615, lut902, lut782, lut701, lut833, lut818, lut715, lut832 (Dallas and Piertney, 1998), 04OT05, 04OT22 (Huang *et al.*, 2005)) was conducted for 177 individuals. POPASSIGN was used to identify the allelic diversity and the observed (*Ho*) and expected (*He*) heterozygosity of the loci. Significant deviations from Hardy-Weinberg equilibrium (HWE) for each locus in the population were tested using the software GENEPOP Version 3.3 (Raymond and Rousett, 1995).

RESULTS

1. Rarefaction analysis

The difference between consecutive sampling in the outcome of *R* was expressed as a function of the total number of loci drawn, and showed that mean and variance estimates of relatedness *(R)* stabilised after 15 loci (Fig. 2). Therefore 15 loci can be used to provide consistent measures of relatedness.

2. Genetic variability

The microsatellite loci for Wales and Borders otters are polymorphic with an average of 5.1 alleles per locus (min-max: 3-7). Comparison with the results of other studies of the European otter (Table 1) shows that the larger sampling area of the European population studied by Randi *et al,* (2003) had a higher average number of alleles per locus of 7.8. The smaller island populations of Kinmen (China) and Sealand (Denmark) showed fewer alleles per locus averaging 0.35-0.39 and 3.6 alleles per locus respectively.

The Wales and Borders otter population had an average expected heterozygosity (*He*) of 0.53 over the 15 loci. This was somewhat lower than the European average *He* = 0.74 (Randi *et al*, 2003), and also lower than the island population of Kinmen *He* = 0.61, 0.70 (Hung *et al.*, 2004, Huang *et al*, 2005). The *He* of the Wales and Borders population was however, similar to Sealand in Denmark, *He* = 0.51 (Pertoldi *et al.*, 2001) despite having 40 % more alleles on average per locus.

The pooled European samples (Randi *et al*, 2003) showed significant deviation from HWE, with significantly positive *Fis* values for 9 out of 11 loci. In contrast, they found that most local populations were actually in HWE (over all loci) when analysed separately. However, French and German samples still showed significant deviations from HWE which Randi *et al.,* (2003) suggested could be due to the Wahlund effect.(artifactual deviation due to a sample that is composed of sub-samples from separate populations; Hartl and Clark, 1997). Pertoldi *et al.*, (2001), Hung *et al.*, (2004) and Huang *et al.*, (2005) studied populations over smaller areas than Randi *et al.*, (2003) and found little evidence for deviations from HWE.

In this study the Wales and Borders samples show that the observed were generally lower than the expected heterozygosities. Significant deviations from HWE were observed at four out of

Table 1 - Summary of observed (Ho) and expected (He) heterozygosity and observed allele number (n alleles) for the 15 loci chosen by rarefaction analysis over five studies of European otters (* $p<0.05$, *** $p<0.001$ significant difference between Ho & He).

	Hobbs *et al.*, unpublished		Hung *et al.*, 2004		Huang *et al.*, 2006		Randi *et al.*, 2003		Pertoldi *et al.*, 2001	
	Wales and Borders		Kinmen (island), China		Kinmen (island), China		Europe		Sealand, Denmark	
Locus	H_o *(in alleles)*	H_e	H_o *(in alleles)*	H_e	H_o *(in alleles)*	H_e	H_o *(in alleles)*	H_e	H_o *(in alleles)*	H_e
lut435	0.44 (5)	0.47					0.61 (12)*	0.83	0.33 (5)	0.60
lut453	0.27 (5)	0.31					0.69 (9)*	0.82		
lut604	0.54 (4)	0.63					0.43 (9)*	0.75		
lut615	0.55 (6)	0.63					0.63 (1)*	0.83		
lut701	0.46 (3)	0.42	0.61 (5)	0.56			0.58 (8)*	0.76	0.50 (3)	0.42
lut715	0.55 (6)*	0.57	0.89 (6)	0.76			0.46 (6)*	0.64		
lut717	0.35 (5)	0.41	0.71 (3)	0.52					0.56 (2)	0.55
lut733	0.47 (5)	0.46	0.89 (4)	0.69			0.57 (8)	0.69	0.39 (4)	0.46
lut782	0.46 (4)	0.47	0.79 (2)	0.5			0.54 (8)	0.55	0.33 (3)	0.38
lut818	0.64 (7)	0.67					0.49 (6)*	0.76	0.69 (4)	
lut733	0.47 (5)	0.46	0.89 (4)	0.69			0.57 (8)	0.69	0.69 (4)	0.62
lut832	0.26 (5)**	0.35	0.66 (3)	0.55			0.48 (6)*	0.69	0.56 (4)	0.49
lut833	0.71 (5)	0.71	0.74 (4)	0.7			0.54 (6)*	0.78		
lut902	0.55 (7)***	0.65							0.60 (4)	0.57
04OT05	0.63 (6)	0.67			0.83 (4)	0.72				
04OT22	0.50 (4)	0.53			0.59 (3)	0.68				
Mean	0.49 (5.1)	0.53	0.76 (3.9)	0.61	0.71 (3.5)	0.70	0.55 (7.8)	0.74	0.50 (3.6)	0.51

teen loci (see Table 1), with loci lut733 standing out as the only locus having significantly more observed heterozygotes than expected. The three other loci showed a significant deficit in observed heterozygotes with lut832 and lut902 showing highly significant deviations. This could be due to a number of reasons, such as allelic dropout or DNA degradation, however, neither of these seem likely given the quality and quantity of DNA extracted from muscle tissue. Randi *et al.* (2003) suggested that significant deviations from HWE in their samples from France and Germany could be due to the Wahlund effect as a result of differentiation at a lower geographical scale. If this was the case it would be expected to see more loci showing significant deviation from HWE. This was the case when nine English samples were added to the analysis of the Wales and Borders population (results not presented here), when eleven loci showed deviations from HWE. Likewise if inbreeding was a cause more loci would be expected to show significant deviations from HWE.

FUTURE DIRECTIONS

Future work will identify the reasons for the anomalies for these two loci, using the suggestions made by Wondji *et al.* (2002), for example focusing on locus-specific constraints such as null alleles (Callen *et al.*, 1993), limited allelic range (Epplen *et al.*, 1993) or preferential amplification of one allele in heterozygotes (Wattier *et al.*, 1998), rather than population substructure or inbreeding (Wondji *et al.*, 2002).

In addition, further analyses will be undertaken. Using the perspectives of landscape genetics, spatial genetic patterns will be assessed at an individual level without defining populations in advance (Manel *et al.*, 2003).

Methods that can be used for analysis of the results include Mantel's test, to identify the presence of an isolation-by-distance pattern between individuals using genetic differentiation and geographical distance (Manel *et al.*, 2003). Multivariate analysis and synthesis maps, using principal component analysis (PCA) vectors can also be used. PCA summarises all the variation for many loci in the study area, and can accommodate individuals as the operational units. The interpolation of the major principal components derived from the PCA leads to a synthesis map (Manel *et al.*, 2003).

There are specific methods to infer genetic boundaries from allele frequency spatial distributions. Monmonier's algorithm visualises data contained on a genetic distance matrix on a geographical map. A Womble approach locates boundaries across a surface for an interpolated variable (i.e. allele frequency surface) by searching for regions in which the absolute value of the surface slope is large (Manel *et al.*, 2003). Delaunay triangulations and Voronoi diagrams can be used for surface modelling by using a finite set of points scattered over a surface to construct a three-dimensional model (Attali and Boissonnat, 2004).

Once the genetic pattern is identified it must be correlated with environmental and landscape variables. In parallel to statistical tests, GIS will be used to visualise spatial genetic patterns and

also generate hypotheses about the cause of genetic boundaries because it allows landscape variables to be overlaid onto genetic data.

ACKNOWLEDGEMENTS

I (H.G.I.) would like to thank my supervisors for their guidance, Vic Simpson for otter samples for the south west of England and the Environment Agency, for their funding of the Cardiff University Otter Project. We would also like to thank the Environment Agency, CCW and UK Wildlife Trusts for their collection of otter carcasses.

REFERENCES

Altmann, J., Alberts, S., Haines, S.A., Dubach J., Muruth P., Coote T., Geffen E., Cheeseman D.J., Mututua R.S., Saiyalel S.N., Wayne R.K., Lacy R.C. and Bruford M.W. 1996 Behaviour predicts genetic structure in a wild primate group. Proceedings of the National Academy of Sciences, 93: 5797-5801

Attali D. and Boissonnat J.D. 2004. A linear bound on the complexity of the delaunay triangulation of points on polyhedral surfaces. *Discrete and Computational Geometry*, 31:369–384.

Barbosa A.M., Real R., Olivero J. and Vargas J.M. 2003. Otter (*Lutra lutra*) distribution modelling at two resolution scales suited to conservation planning in the Iberian. *Biological Conservation*, 114(3): 377-387.

Berthier K. 2005. Genetic structure of the cyclic fossorial water vole (*Arvicola terrestris*): landscape and demographic influences. *Molecular Ecology*, 14: 2861–2871.

Bertorelle G. and Barbujani G. 1995. Analysis of DNA diversity by spatial autocomelation. *Genetics*, 140: 811-819.

Bruford M.W. and Wayne R. K. 1993. Microsatellites and their application to population genetic studies. Current opinion in Genetics and Development, 3: 939-943.

Bruford M.W., Cheesman D.J., Coote T., Green H.A.A., Haines S.A., O'Ryan C. and Williams T.R. 1996. Microsatellite and their approaches to conservation genetics. In: T.B. Smith and R.K. Wayne (eds), Molecular Genetic Approaches in Conservation. OUP New York.

Callen D.F., Thompson A.D., Shen Y., Phillips H.A., Richards R.I., Mulley J.C. and Sutherland G.R. 1993 Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics*, 22: 1–10.

Charlesworth D. and Charlesworth B., 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, 18: 237-268.

Chanin P. 1985. The Natural History of Otters, Croom Helm, London, England

Chanin P. 2003. Monitoring the otter *Lutra lutra*. Conserving Natura 2000 Rivers Monitoring Series No. 10, English Nature, Peterborough.

Conroy J.W.H. and Chanin P.R.F., 2000. The status of the Eurasian otter (*Lutra lutra*) in Europe. A Review. In: Conroy J. W. H, Yoxon P. and Gutleb A.C. (eds), Proceedings of the First Otter Toxicology Conference, Journal of the International Otter Survival Fund, 1: 7-28.

Couvet D. 2002. Deleterious effects of restricted gene flow in fragmented populations. *Conservation Biology*, 16(2): 369-376.

Coxon K.E, Chanin P.R.F, Dallas J.F. and Sykes T. 1999. The use of DNA fingerprinting to study the population dynamics of otters (*Lutra lutra*) in Southern Britain: Feasibility study. The Environment Agency, R&D report W202.

Dallas J.F. and Piertney S.B. 1998. Microsatellite primers for the Eurasian otter. *Molecular Ecology*, 7: 1248-1251.

Dallas J.F., Bacon P.J., Carss D.N., Conroy C.W.H., Green R., Jefferies D.J., Kruuk H., Marshall F., Piertney S.B. and Racey P.A. 1999. Genetic diversity in the Eurasian otter, *Lutra lutra*, in Scotland. Evidence from microsatellite polymorphism. *Biological Journal of the Linnean Society*, 68(12): 73-86.

Dallas J.F., Marshall F., Piertney S.B., Bacon P.J. and Racey P.A. 2002. Spatially restricted gene flow and reduced microsatellite polymorphism in the Eurasian otter *Lutra lutra* in Britain. *Conservation Genetics*, 3: 15–29.

Dallas J.F., Coxon K.E., Sykes T., Chanin P.R.F., Marshall F., Carss D.N., Bacon, P.J., Piertney S.B. and Racey P.A. 2003. Similar estimates of population genetic composition and sex ratio derived from carcasses and faeces of Eurasian otter *Lutra lutra*. *Molecular Ecology*, 12 (1): 275-282.

de Ruiter, J.R. and Geffen E. 1998 Relatedness of Matrilines, Dispersing males and social groups in Long-tailed macaques (*Macaca fascicularis*) Proceedings: *Biological Sciences*, 265 (1391): 79-87

Epplen C., Melmer G., Siedlaczck I., Schwaiger F.W., Mäueler W. and Epplen J.T. 1993. On the essence of 'meaningless' simple repetitive DNA in Eucaryotes genomes. In: Pena S. D. J, Chakraborty R.,. Epplen J. T and Jeffreys A. J (eds.), DNA Fingerprints: State of the Science. Birkhauser Verlag, Basel, Switzerland. 29–45.

Girman D. J., Mills M.G.L. Geffen E. and Wayne R.K. 1997. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (*Lycaon pictus*). *Behavioral ecology and socio-biology*, (40): 187 - 198

Hauer S. Ansorge H. and Zinke O. 2002. Mortality patterns of otters (*Lutra lutra*) from eastern Germany. *Journal of Zoology* London, 256: 361-368.

Hartl D. L. and Clark A. G. 1997. Principles of Population Genetics. Sinauer Associates, Sunderland MA, USA

Hooftman D.A., Billeter R.C., Schmid B. and Diemer M. 2003. Genetic effects of habitat fragmentation on common species of Swiss fen meadows. *Conservation Biology*, 18(4): 1043-1051.

Hung C-M., Li S-H and Lee L-L. 2004. Faecal DNA typing to determine the abundance and spatial organisation of otters (*Lutra lutra*) along two stream systems in Kinmen. *Animal Conservation*, 7: 301–311.

Huang C-C., Hsu Y-C., Lee L-L and Li S-H. 2005. Isolation and characterization of tetramicrosatellite DNA markers in the Eurasian otter (*Lutra lutra*). *Molecular Ecology Notes*, 5(2): 314-316.

Jarne P. and Lagoda J. L. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11: 424-428.

Jones T. and Jones D. 2004. Environment Agency Wales, Otter survey of Wales 2002.

Kays R.W., Gittleman J.L., Wayne R.K. 2000. Microsatellite analysis of kinkajou social organization. *Molecular Ecology*, 9 (6): 743-751

Manel S., Schwartz M.K., Luikart G. and

Taberlet P. 2003. Landscape Genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, 18(4): 189-197.

Mason C.F. 1998. Decline in PCB levels in otters (*Lutra lutra*). *Chemosphere*, 36(9): 1969-1971.

Mason C.F. and Macdonald S.M. 2004. Growth in otter (*Lutra lutra*) populations in the UK as shown by long-term monitoring. *Ambio*, 33(3): 148–152.

Miller C.R., Joyce. P. and Waits L.P. 2005. A new method for estimating the size of small populations from genetic mark–recapture data. *Molecular Ecology*, 14(7), 1991-2005.

Mills L., Citta J., Lair K., Schwartz M. and Tallmon D. 2000. Estimating animal abundance using non-invasive DNA sampling: promise and pitfalls. *Ecological applications*, 10: 283–294.

Pertoldi C., Hansen M.M., Loeschcke V., Madsen A.B., Jacobsen L. and Baagoe H. 2001. Genetic consequences of population decline in the European otter (*Lutra lutra*): an assessment of microsatellite DNA variation in Danish otters from 1883 to 1993. Proceedings of the Royal Society B: *Biological Sciences*, 268: 1775-1781.

Philcox C.K., Grogan A.L. and Macdonald D.W. 1999. Patterns of otter *Lutra lutra* road mortality in Britain. *Journal of Applied Ecology*, 36: 748-762.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L. and Estoup A. 2004. Computer note GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, 95(6):536–539.

Prigioni C., Remonti L., Balestrieri A., Sgrosso S., Priore G., Misin C., Viapiana M., Spada S. and Anania R. 2005. Distribution and sprainting activity of the Otter (*Lutra lutra*) in the Pollino National Park (southern Italy). *Ethology Ecology & Evolution*, 17: 171-180.

Queller D.C. and Goodnight K.F. 1989. Estimating Relatedness Using Genetic Markers. *Evolution*, 43(2): 258-275

QIAGEN, #65906, DNeasy Tissue handbook, Protocol: Purification of total DNA from animal tissues, 18-20

QIAGEN, #206143, QIAGEN® Multiplex PCR Handbook, Protocol: Amplification of microsatellite loci using multiplex PCR, 19-21

Randi E., Davoli F., Pierpaoli M., Pertoldi C., Madsen A. B. and Loeschck V. 2003. Genetic structure in otter (*Lutra lutra*) populations in Europe: implications for conservation. *Animal Conservation*, 6(2): 93-100.

Ralls K., Ballou J.D. and Templeton A.R. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biololgy*, 2: 185-193.

Raymond M. and Rousett F. 1995. Genepop (version 1.2) population genetics software for exact tests and ecumenicalism. *Journal of Heredity*, 86: 248-249.

Ruiz-Olmo J. and Delibes M. (eds) 1998. La nutria en España ante el horizonte del año 2000. SECEM, Málaga.

Ruiz-Olmo J., Saavedra D. and Jimenez J. 2001. Testing the surveys and visual and track censuses of Eurasian otters (*Lutra lutra*). *Journal of Zoology* London, 253: 359-369.

Sjoasen T. 1997. Movements and establishment of reintroduced European otters *Lutra lutra*. *Journal of Applied Ecology*, 34: 1070-1080.

Smith D., Meier T., Geffen E., Mech L.D., Burch J.W., Adams L.G. and Wayne R.K. 1997. Is incest common in gray wolf packs? *Behavioral Ecology*, 8: 384–391.

Taberlet P, Camarra J., Griffin S., Uhrès E., Hanotte O., Waits L. P., Dubois-Paganon C., Burke T. and Bouvet J. 1997. Noninvasive genetic tracking of the endangered Pyrenean brown bear

population. *Molecular Ecology*, 6(9): 869-876.

Wattier R., Engel C.R., Saumitou-Laprade P. and Valero M. 1998. Short allele dominance as a source of heterozygote deficiency at microsattellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, 7: 1569–1573.

Wimmer B., Tautz D. and Kappeler P.M. 2002. The genetic population structure of the gray mouse lemur (*Microcebus murinus*), a basal primate from Madagascar. *Behavioral ecology and socio-biology*, 52:166–175

Wondji C., Simard F. and Fontenille D. 2002. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Molecular Biology*, 11 (1): 11-19.

Zane L., Bargelloni L., Patarnello T. 2002 Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11:1-16.

Zhivotosky L.A. and Feldman M.W. 1995. Microsatellite variability and genetic distances. Proceedings of the National Academy of Sciences of the United States of America, 92: 11549-11552.

Appendix 6

Stanton DWG, Hobbs GI, Chadwick EA, Slater FM and Bruford MW (2009) Mitochondrial genetic diversity and structure of the European otter (*Lutra lutra*) in Britain. *Conservation Genetics*, **10**, 733-737.

Conserv Genet
DOI 10.1007/s10592-008-9633-y

TECHNICAL NOTE

Mitochondrial genetic diversity and structure of the European otter (*Lutra lutra*) in Britain

D. W. G. Stanton · G. I. Hobbs · E. A. Chadwick · F. M. Slater · M. W. Bruford

Received: 11 December 2007 / Accepted: 12 June 2008

Abstract The European otter (*Lutra lutra*) is a focus for conservation efforts throughout Europe due to a population decline in recent decades and because of its importance as a biological indicator of the health of rivers and waterways. The aim of this study was to aid the conservation of this species by adding genetic information from samples originating in the United Kingdom (UK), to help build up a picture of the phylogeographic structure of the European otter throughout Europe. This was done by a comparison of 299 base pairs of the mitochondrial DNA control region. Four haplotypes were identified in the UK, one of which has not been found outside the west of the UK in the wild, and one of which was unique. Populations in the UK, and in particular the west were shown to have a higher haplotype diversity than previously found for the European otter in Europe ($h = 0.7338$ for the 58 UK otters sampled in this study) and an overall nucleotide diversity of $\pi = 0.003$. The western UK population was shown to have a high level of genetic distinctiveness. We discuss possible contributory population processes, the importance of the western UK population for the future conservation of the species and comment on future conservation strategies.

Keywords *Lutra lutra* · Control region · Conservation · Haplotype · Genetic structure

D. W. G. Stanton · G. I. Hobbs · E. A. Chadwick · F. M. Slater · M. W. Bruford (✉)
School of Biosciences, Cardiff University, Cathays Park, Cardiff CF10 3US, UK
e-mail: BrufordMW@cardiff.ac.uk

Introduction

The otter is a top predator and important biological indicator of the health of rivers and wetlands. The monitoring of this species is therefore a priority for the continued conservation of these ecosystems (Crawford 2002). Recent decades have shown a decline in numbers and distribution of the European otter (*Lutra lutra*) (Mason and Macdonald 1986; Foster-Turley et al. 1990). A number of factors have been suggested as a cause for the decline, such as a reduction in fish stocks, the loss of riparian habitat, hunting, road traffic accidents and fish traps (Macdonald and Mason 1994; cited in Mason and Macdonald 2004). However, the most significant cause is usually attributed to water contamination by organochlorine pesticides and polychlorinated biphenyls (PCBs; Mason 1995; Murk et al. 1998). Mason and Macdonald (2004) reported a slow recovery in the United Kingdom (UK) otter population over recent years, attributed to a reduction in levels of these pollutants. It is likely that this recovery has also been helped by human intervention. The otter trust in Earsham, Suffolk re-introduced 117 otters between 1983 and 1999 at several locations, including 56 in East Anglia. These otters were captive bred but of unknown origin (Pers. Comm., Woodroffe 2007).

Previous studies of mitochondrial DNA control region (mtDNA CR) variation of European otter in Europe have found generally low levels of haplotype diversity. Within Europe, Mucci et al. (1999) identified two haplotypes, Perez-Haro et al. (2005) identified three and Cassens et al. (2000) and Ferrando et al. (2004) identified five each. Haplotypes in these studies are often the same and the haplotype variously designated DK/Lut1/H1 has repeatedly been found to be the most common, although a second—UK/Lut3/H4 was identified as the most common in East

Germany (Cassens et al. 2000). A total of eight different haplotypes have previously been identified in Europe, for the mtDNA CR fragment analyzed in this study. Thus far, European phylogenetic analysis has given networks showing a star-like structure, interpreted as evidence of a population bottleneck followed by rapid expansion.

This study aimed to characterize the mtDNA CR genetic structure in European otter throughout the UK. We combined sequences from the UK identified in this study with sequences already obtained from the rest of Europe (Mucci et al. 1999; Cassens et al. 2000; Ferrando et al. 2004; Perez-Haro et al. 2005) and describe the implications of these results for the future conservation of this species.

Materials and methods

DNA was extracted from the muscle tissue of road-killed otters from Eastern England ($n = 12$), Scotland ($n = 2$), Gloucestershire ($n = 3$) and Wales ($n = 41$) using the Qiagen DNeasy tissue extraction kit, following the manufacturer's instructions. A 299 base pair segment of the mtDNA CR from 58 samples was analyzed. Amplifications were run on a PerkinElmer GeneAmp PCR System 9700 with the following conditions: 94°C for 2 min; 35 cycles—94°C for 15 s, 50°C for 15 s, 72°C for 15 s; 72°C for 5 min. Primers used were L-Pro (5′-CGT CAG TCT CAC CAT CAA CCC CCA AAG C-3′) and H-Phe (5′-GGG AGA CTC ATC TAG GCA TTT TCA GTG-3′), which bind to the flanking tRNA-Pro (L-primer) and flanking tRNA-Phe (H-primer) regions, respectively. PCR reactions were performed in a final volume of 25 µl, including 20–50 ng target DNA, 10× PCR buffer (Invitrogen), 3 mM $MgCl_2$, 0.05 mM dNTPs, 0.1 µM each primer and 0.625 units of *Taq* DNA polymerase (Invitrogen). Negative controls, where DNA was substituted with water, were used for each PCR. Samples were precipitated in 75% isopropanol and sequenced using a BigDye® Terminator v1.1 cycle sequencing kit (following manufacturer's instructions) and an Applied Biosystems 3130XL genetic analyser.

Two hundred and ninety-nine base pairs of sequence from the 5′ end of the mtDNA CR from the 58 samples in this study was aligned with equivalent European otter sequences derived by Mucci et al. (1999), Cassens et al. (2000), Ferrando et al. (2004) and Perez-Haro et al. (2005) using DAMBE v4.2.13 (Xia and Xie 2001). All samples ($n = 357$) were then analyzed to give a description of the distributions of known European otter CR haplotypes across Europe. The connection length between all haplotypes detected was calculated in Arlequin v3.01 (Excoffier et al. 2005) in a pairwise fashion and a minimum spanning network of British samples was drawn by eye. Haplotype diversity was estimated in DnaSP v4.10.9 (Rozas et al. 2003) and nucleotide diversity in MEGA v3.1 (Kumar et al. 2004). Arlequin v3.01 was used to carry out an analysis of molecular variance to investigate partitioning of genetic variation in European otter populations in Britain and Europe. Cassens et al. (2000) found haplotype Lut4 to have an insertion (cytosine at position 101) and when aligned, the gaps at this position in the other samples were considered when calculating haplotype diversity by DnaSP. MtDNA sequences were not available for samples sequenced by Ferrando et al. (2004). However, the haplotypes H1 and H4 were described by Ferrando et al. (2004) to match Lut1 and Lut3 (Cassens et al. 2000) respectively. The remaining Ferrando et al. (2004) haplotypes could then be reconstructed from Fig. 1 in the text, relative to these two. All the sequences obtained in this study have been submitted to GenBank (Accession No. EU294255–EU294258).

Results

We identified four haplotypes in UK populations. Two of these (Lut1 and Lut3) are identical to the two most dominant haplotypes already identified in Europe (Mucci et al. 1999; Cassens et al. 2000; Ferrando et al. 2004; Perez-Haro et al. 2005). A third had only previously been found in captive bred otters (Lut6, Perez-Haro et al. 2005) and in a single individual, in Wales (H2, Ferrando et al. 2004). The fourth has not previously been described, and will be referred to as Lut7. Lut7 possesses a C to T transition at position 152 and a T to C transition at position 236, from haplotype Lut1. For the fragment we compared, haplotypes H5 and H6 were identical, as were H1 and H7 (Ferrando et al. 2004). H5 and H6 will be referred to as H5. Three samples were found in Gloucestershire and were grouped with the Welsh samples for simplicity, and will be referred to as western UK. Two variable nucleotides were identified, both C-T transitions at positions 152 and 236. A haplotype diversity of $h = 0.7338$ and an overall nucleotide diversity of $\pi = 0.003$ was estimated for the 58 UK otter samples sequenced in this study. A haplotype diversity of $h = 0.4712$ and a nucleotide diversity of $\pi = 0.002$ was calculated for all 357 UK and European samples. Among group variation between the UK (east England, western UK and Scotland—Group 1) and European (all other samples—Group 2) populations accounted for 18.83% of the total variation ($P = 0.126$), among populations within groups accounted for 35.73% ($P < 0.001$) and within population variation accounted for 45.45% ($P < 0.001$). Among group variation between the western UK and all other populations accounted for 52.27% of the total ($P = 0.062$). The geographic distribution of the haplotypes identified in this study

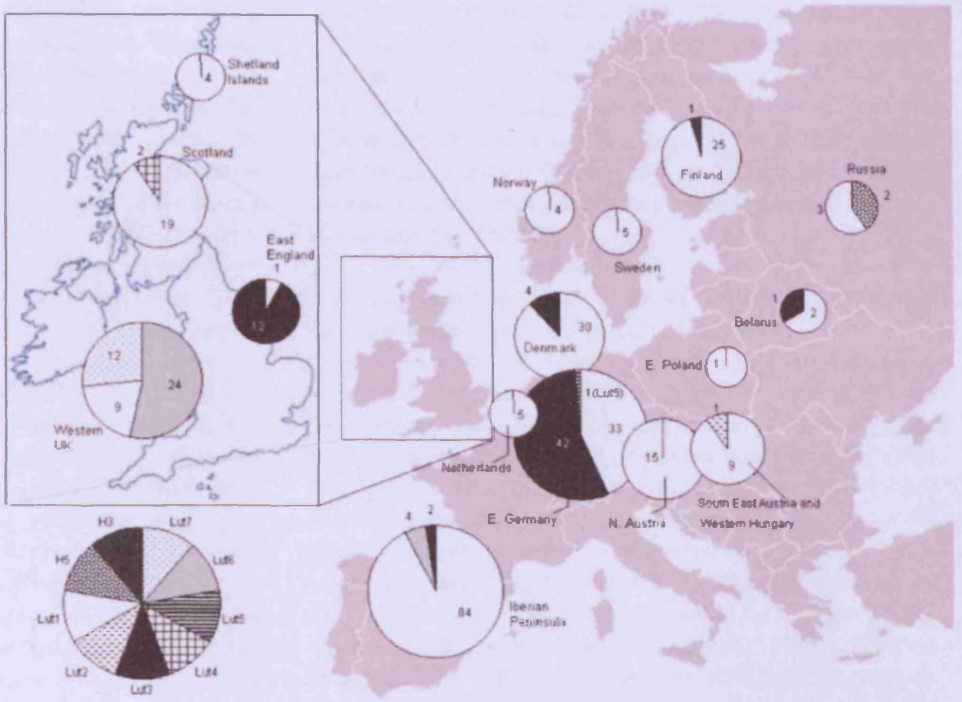


Fig. 1 Haplotype distribution of European otter throughout Europe. Numbers on circle segments are the number of individuals with that particular haplotype. Circle size is approximately proportional to sample number

is shown in Fig. 1 along with those already defined throughout Europe. A minimum spanning network of all known haplotypes (Mucci et al. 1999; Cassens et al. 2000; Ferrando et al. 2004; Perez-Haro et al. 2005) is shown in Fig. 2. The individuals with Lut3 in Denmark (Fig. 1) are of presumed English origin (Mucci et al. 1999) and the individuals with Lut3 and Lut6 in the Iberian Peninsula are captive otters originating from France and England respectively (Perez-Haro et al. 2005).

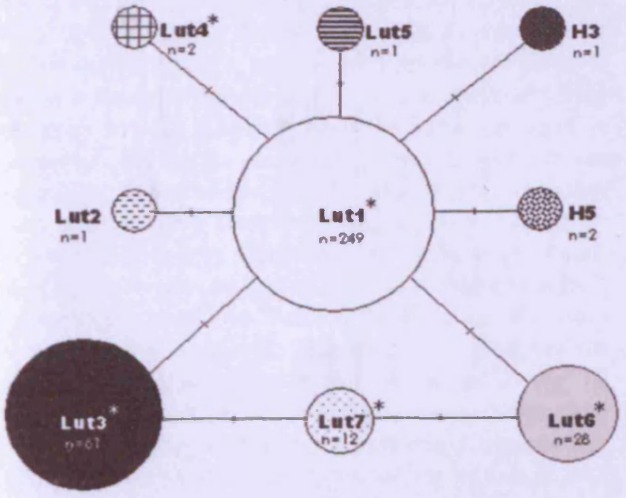


Fig. 2 Minimum spanning network showing all known haplotypes (Mucci et al. 1999; Cassens et al. 2000; Ferrando et al. 2004; Perez-Haro et al. 2005). Haplotypes that have been identified in Britain have been marked with asterisks

Discussion

The most unexpected result in this study was the discovery of such high frequencies of haplotypes Lut6, and the novel haplotype, Lut7 in European otter populations in western UK, relative to the rest of Europe. High frequencies of these haplotypes contribute to a high genetic diversity in the mtDNA CR of UK, and in particular, western UK otters compared to the rest of Europe. This can be shown by the high haplotype diversity (UK $h = 0.7338$) relative to the overall European otter haplotype diversity previously estimated by Ferrando et al. (2004) of $h = 0.360$. The haplotype diversity of the European otter in the UK is more consistent with other European mustelid species, e.g. *Martes martes*, $h = 0.76$ (Davison et al. 2001). When UK samples are included with all other known European samples, a haplotype diversity of $h = 0.4721$ is obtained. The minimum spanning network (Fig. 2) shows Lut1 as the central haplotype with the frequency implying its ancestral status within the European sequences. However, the population bottleneck experienced by the European otter in the UK and elsewhere indicates that drift will have played a significant role in determining the relative frequencies of these haplotypes. The structure of UK samples in this network is somewhat different to previous studies of this

otter in Europe (Cassens et al. 2000; Ferrando et al. 2004), which identify a "star-like" phylogeny, usually interpreted as being evidence of a rapid expansion following a severe population bottleneck during Pleistocene glaciations. The UK population, and in particular western UK, shows far higher diversity than previously detected. The distinctiveness of the western UK population is shown by the among group variation between the UK and Europe, which accounts for 18.83% of the variation, and western UK and Europe, which accounts for 52.27% of the variation. These results are conservative as the captive bred otters sampled in Europe have been included in the European group.

The dominance of Lut6 in western UK (53.3%) in comparison with the most dominant haplotype in Europe—Lut1 (Lut1 = 36.7% in the UK as opposed to 76.9% in the rest of Europe) implies that these populations have been demographically isolated, with the novel haplotypes possibly arising relatively recently. Martinkova et al. (2007) provided evidence that the stoat (*Mustela erminea*) colonized Ireland during the last glacial maximum and may not have had land connections to continental Europe and Britain thereafter, with Britain being colonized later by a replacement event. This is one potential explanation for our European otter results, with the haplotypes Lut6 and Lut7 arising either in currently unidentified southern refugia or in situ in the western UK before recent anthropogenic population declines. Further sampling, especially from Ireland would help to investigate this hypothesis.

Western UK is also known to have been a stronghold for European otters during the widespread twentieth century population declines (Mason and Macdonald 1986; Foster-Turley et al. 1990). Due to the clonal maternal inheritance of mtDNA, its effective size in many diploid populations is predicted to be approximately 25% that of nuclear DNA and its frequencies are therefore sensitive to genetic drift. For this reason mtDNA haplotypes may become fixed in small populations relatively rapidly. A comparatively large population size is therefore likely to have remained in western UK during the twentieth century to result in the frequencies of each of the three haplotypes found. Finally, Mucci et al. (1999) discussed a haplotype possessed by four individuals within their sampling in Denmark described as "captive reared otters of presumed English origin". This haplotype (Lut3) is found in the majority of eastern England samples analyzed. This supports the hypothesis that these Danish otters are of English origin. Within the UK, Lut3 is unique to eastern UK. It is also the most common haplotype in this area which might suggest that re-introductions have had an influence on haplotype distribution in some parts of the UK. However, haplotypes are unknown for otters re-introduced by the otter trust in Earsham. In addition, drift may still have a large impact on haplotype distribution in otter populations following re-introductions (Arrendal et al. 2004). This study shows Lut3 to be more common and widely distributed than previously thought. This haplotype has now been identified as the most common in both Germany (Lut3 $n = 42$, 55.3% Cassens et al. 2000) and England, although our sample distribution is biased towards the east of England. It is still evident however that Europe as a whole has relatively low mitochondrial diversity ($h = 0.4712$ for all known European sequences), although there is evidence of other countries showing higher levels of nuclear genetic diversity, for example Ireland (Randi et al. 2003).

The presence of Europe's most common haplotype, Lut1, in the western UK also suggests ancestral gene flow between western UK and the rest of Europe. This together with the apparently limited gene flow between geographic regions in Britain (Dallas et al. 2002) and the relatively high haplotype diversity of the European otter in the UK shows the importance, in terms of conservation, of western UK's otter population. The western UK's otter population should be further investigated to see if it could be a distinct management unit for conservation purposes. Poorly planned reintroductions can potentially have adverse repercussions for the conservation of isolated populations. However this must be balanced with the genetic and demographic risks associated with small isolated populations (Edmands 2007). Current evidence suggests that reintroductions of European otter in the UK have been demographically successful, and haplotypes found in Eurasian otters only differ by single base pair substitutions, suggesting a common evolutionary origin. However, further reintroductions of European otters in the UK would appear to be unnecessary at the present time. Populations appear to be recovering, albeit slowly (Mason and Macdonald 2004), and natural re-colonization through habitat protection and restoration of corridors, reconnecting isolated populations throughout Europe will allow natural demographic recovery of this key European mammal (Reuther 1994).

References

Arrendal J, Walker CW, Sundqvist A et al (2004) Genetic evaluation of an otter translocation program. Conserv Genet 5:79–88. doi:10.1023/B:COGE.0000014059.49606.dd

Cassens I, Tiedemann R, Suchentrunk F et al (2000) Mitochondrial DNA variation in the European otter (*Lutra lutra*) and the use of spatial autocorrelation analysis in conservation. J Hered 91:31–35. doi:10.1093/jhered/91.1.31

Crawford A (2002) Fourth otter survey of England 2000–2002. Available via Conservation Section, Environment Agency, Waterside Drive, Aztec West, Almondsbury, Bristol BS32 4UD. http://www. jncc.gov.uk

Dallas JF, Marshall F, Piertney SB et al (2002) Spatially restricted gene flow and reduced microsatellite polymorphism in the Eurasian otter *Lutra lutra* in Britain. Conserv Genet 3:15–29. doi:10.1023/A:1014259218632

Davison A, Birks JDS, Brookes RC et al (2001) Mitochondrial phylogeography and population history of pine martens *Martes martes* compared with polecats *Mustela putorius*. Mol Ecol 10:2341–2347. doi:10.1046/j.1365-294X.2001.01381.x

Edmands S (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol Ecol 16:463–475. doi:10.1111/j.1365-294X.2006.03148.x

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Ferrando A, Ponsà M, Marmi J et al (2004) Eurasian otters, *Lutra lutra*, have a dominant mtDNA haplotype from the Iberian Peninsula to Scandinavia. J Hered 95:435–440. doi:10.1093/jhered/esh066

Foster-Turley P, Macdonald S, Mason C (1990) Otters. An action plan for their conservation. IUCN/SSC Otter Specialist Group Broadview, Kelvyn Press

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163. doi:10.1093/bib/5.2.150

Martinkova N, McDonald RA, Searle JB (2007) Stoats (*Mustela erminea*) provide evidence of natural overland colonization of Ireland. Proc R Soc B Biol Sci 274:1387–1393. doi:10.1098/rspb.2007.0334

Mason CF (1995) Habitat quality, water quality and otter distribution. Hystrix 7:195–208

Mason CF, Macdonald S (1986) OTTERS: ecology and conservation. Cambridge University Press, Cambridge

Mason CF, Macdonald S (2004) Growth in otter (*Lutra lutra*) populations in the UK as shown by long-term monitoring. AMBIO: J Human Environ 33:148–152

Mucci N, Pertoldi C, Madsen AB et al (1999) Extremely low mitochondrial DNA control-region sequence variation in the otter *Lutra lutra* population of Denmark. Hereditas 130:331–336. doi:10.1111/j.1601-5223.1999.00331.x

Murk AJ, Leonards PEG, van Hattum B et al (1998) Application of biomarkers for exposure and effect of polychlorinated aromatic hydrocarbons in naturally exposed European otters (*Lutra lutra*). Environ Toxicol Pharmacol 6:91–102. doi:10.1016/S1382-6689(98)00026-X

Perez-Haro M, Vinas J, Manas F et al (2005) Genetic variability in the complete mitochondrial control region of the Eurasia otter (*Lutra lutra*) in the Iberian Peninsula. Biol J Linn Soc 86:397–403. doi:10.1111/j.1095-8312.2005.00536.x

Randi E, Davoli F, Pierpaoli M et al (2003) Genetic structure in otter (*Lutra lutra*) populations in Europe: implications for conservation. Anim Conserv 6:93–100. doi:10.1017/S1367943003003123

Reuther C (1994) European otter habitat network. In: Seminar on the conservation of the European otter (*Lutra lutra*), Environmental Encounters 24, Strasbourg, Leeuwarden Council of Europe

Rozas J, Sánchez-DelBarrio JC, Messeguer X (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497. doi:10.1093/bioinformatics/btg359

Xia X, Xie Z (2001) DAMBE: data analysis in molecular biology and evolution. J Hered 92:371–373. doi:10.1093/jhered/92.4.371