Natural Environment Research Council

Dynamics of demographic expansion and population structure in the otter (*Lutra lutra*).

A thesis submitted to Cardiff University for the degree of

Doctor of Philosophy

By

Nia Evelyn Thomas

May 2021

Table of Contents

List of Figures

List of tables

Statement of contributions

Where the generation of genetic data used in this thesis was conducted by others, their contributions are explicitly detailed below:

Chapter 2: Country-wide genetic monitoring over 20 years reveals spatial connectivity with persistent population structure in a recovering and expanding carnivore population – Previous research students Dr Geoff I. Hobbs and Dr Dave W. G. Stanton generated the genetic data for 207 and 7 of the samples used in this chapter used in the publications Hobbs et al. 2011 and Stanton et al. 2014 respectively. The remaining 193 samples were genotyped by myself along with the calibration of datasets through the re-analysis of 14 samples from Hobbs et al. 2011. Contributions are specified by study in more detail in supplementary 2.2 (p37-40).

Chapter 3: Evaluating environmental factors as facilitators and barriers to gene flow in the Eurasian otter (Lutra lutra) using landscape genetics – The 192 genotyped individuals used in this analysis were genotyped by Dr Geoff I. Hobbs as part of a previous publication Hobbs et al. 2011 as acknowledged in the relevant methods section of this chapter (3.2.2 sample collection and genotyping, p49).

Chapter 4: Detecting changes in effective population size in an expanding population with known recent history – As in Chapter 2 previous research students Dr Geoff I. Hobbs and Dr Dave W. G. Stanton generated the genetic data for 207 and 7 of the samples used in this chapter used in the publications Hobbs et al. 2011 and Stanton et al. 2014 respectively. The remaining 193 samples were genotyped by myself along with the calibration of datasets through the re-analysis of 14 samples from Hobbs et al. 2011.

Acknowledgements

This thesis is dedicated to Andrew Thomas, thanks for sharing with me your sense of humour, love of rugby, excellent taste in music and for being an all-round fantastic uncle. Despite everything life threw at you there was always a smile on your face and a positive attitude, thinking of you and your outlook kept me going through some of the harder times in this process. You were always excited about the idea of having a Dr in the family, and I'm sad you won't be here to see it, but so glad for all the great times (and beers) we did share together.

More people have helped get this thesis to completion than I could possibly name – thanks to all of you, I'm so grateful to have met each and every one of you.

Mike, Liz and Frank – Thank you for all the time, patience and wisdom you've put into this project over the past five and a half years. It's been a rough old ride at times, and I'll be forever grateful for your belief in me, particularly when I didn't believe in myself. I've learnt so much along the way, both academic and otherwise, and I'm grateful for your support, approachability and enthusiasm across both areas equally – thanks for being colleagues as well as supervisors. Mike, thanks for the extra encouragement over these past couple of months, I don't think I would have got to the end without it, and when I say 'nice jumper' I really do mean it!

The MolEcol lab+ – The pandemic (and home working) has highlighted more than ever what a fantastic bunch of people you all are. I've missed everything about C5.15 during this time away; the friendship, the laughs, the willingness to pile round a computer screen or take a tea break and problem solve. Many people have passed through the office during my time, but its essence remains the same and I hope that's something that continues forever. Thank you all for your support both big and small it really has made all the difference, please continue to have fun, make jokes and generally be silly as well as doing science.

Cardiff University Otter Project – Thanks to everyone who's been involved in the project since its inception, from the people reporting dead otters, to the organisations collecting and delivering them, to everyone who's spent time in the otter PM lab. The archive is what it is because of each and every one of you and this thesis wouldn't have been possible without you.

Josie – I always said, if nothing else, then this experience was worth it for meeting you. Thanks for seeing me through thick and thin, always making me laugh and saying yes to all kinds of plans and schemes (I realise that many of them were more type 2 fun…). You really are one in a million and I'm looking forward to many more adventures, and *road surfaces* with you :)

Sophie – I'm so glad that a friendship hatched from having the same pj bottoms for the same reason all those years ago. Thanks for sharing so many highlights and pivotal experiences with me, bear crew and Alaska would have been very different without you there too, as would have the last few years and I hope we continue to share adventures in the future.

Amy – You are a fabulous human and a fountain of natural history knowledge, thank you for all the times you've made me laugh, the walks round the park, G&T's and for deciphering my bird call impressions.

Cat – Thanks for getting me through the past year and a half. I know its not always been easy but I can't imagine having done it without you. I'm very much looking forward to spending more time and adventures with you.

Family – Mum and Dad, thanks for always encouraging me to do what I enjoy, for letting me forge my own path, even when it's a bit odd and for listening when things have been hard. Carys and Rhod, thanks for the support and distractions in equal measure, I can't wait to actually spend some time with you both. The wider Thomas clan, thanks for being such a big, boisterous and loving family.

The Girls – Thanks for being the best bunch of friends I could ask for. Your support, love and friendship over the past three decades (as well as the past five years) has helped both keep me going and given me perspective. I love you all and can't wait for a proper catch up.

Songbirds Choir – Thank you all for the weekly singing, the copious giggles and new friendships. Being part of something completely separate from my PhD, with a diverse group of people that leaves me with a smile on my face has been a joy and a real lifeline at times. You're all fantastic human beings and I can't wait until we can sing properly together again.

Jess, Mat and the ERS team – Firstly, thank you for giving me the chance to be a part of some really interesting work while paying the bills. Secondly, and more importantly, thank you for being such awesome colleagues; the laughs, the friendship and the care you've shown me over the last year has made all the difference.

Somerset Otter Group – Thank you to all the volunteers who helped collect otter spraint from across the Tone and Brue catchments in 2016 – 2018. I'm sad that the pandemic has meant that this work couldn't be included within my thesis anymore but I look forward to sharing the results with you all at some point in the future.

Thank you to the NERC GW4+ DTP studentship and CASE partner Somerset Otter Group for funding this work.

Summary

Genetic diversity is considered one of the three main pillars of biodiversity, yet to date conservation policy has largely overlooked monitoring and protecting it in favour of species and ecosystem-based targets. Populations may experience considerable genetic erosion without extinction, and this loss can take considerably longer to recover from than population size alone. Low genetic diversity can have detrimental effects on the evolutionary potential of species and the viability of populations, yet population recovery at the genetic level remains poorly understood in the wild.

In this thesis, I use large scale spatio-temporal sampling to explore the recovery of the otter population across the UK at a genetic level, after a well-documented decline in the latter half of the 20th century. Successive national surveys have documented increased occupancy rates, from small fragmented stronghold areas to a near continuous distribution in the 21st century. Firstly, I show that despite increased gene flow between sub-populations over the last 20 years, significant population structure persists, and intra-regional genetic diversity has not increased. Secondly, I use a landscape genetics approach, finding that favourable habitat is the most significant environmental factor affecting functional connectivity of otters across the Wales and Borders region, with implications for conservation management. Thirdly, I show that despite a near contiguous distribution and assumed population recovery, effective population sizes remain significantly below those required for long-term population viability in regional otter populations, and in some cases even raise concerns over short-term viability.

This thesis serves as an important demonstration of the practical value of genetic data, even for populations where a significant (non-genetic) monitoring program has existed for several decades. The differing conclusions regarding population status, when based on genetic data rather than occupancy data, highlight the inherent dangers in assuming that spatial connectivity is an indicator of genetic connectivity.

1

General Introduction

1.1 Global loss of species and genetic diversity

Wildlife populations have experienced unprecedented rates of change under the influence of humans, as illustrated by contemporary global species extinction rate estimates that are 1000 – 10,000 times greater than estimated background rates (Pimm et al. 2014; De Vos et al. 2018). Aside from extinctions, anthropogenic pressures are resulting in a multitude of other detrimental changes, such as functional declines, population declines and fragmentation (Diaz et al. 2019). This global trend of wild population decline has recently been shown to reflect a parallel global decline in genetic variation (Leigh et al. 2019). Given that genetic variation is one of the pillars of biodiversity and provides the building blocks for evolution and adaptation it is crucial that its widescale decline is halted (Diaz et al 2020). Against this backdrop of global decline, some populations and species have been identified as vulnerable, and action taken to promote population recoveries (Marsh and Trenham 2008). However, data and milestones on what constitutes a successful recovery, especially at the genetic level, in natural populations are sparse and much more is needed to inform management practises going forward.

1.2 Expanding populations

The distribution of organisms changes over both space and time making population size fluctuations a relatively common occurrence. These fluctuations have occurred historically, i.e. due to the contraction into and subsequent expansion out of glacial refugia (Hewitt 2000) and more recently, due to anthropogenic pressures (such as hunting) and subsequent protection (Amos & Harwood 1998; Cardillo et al. 2005; Nussey et al. 2005). While ancient events such as post-glacial population expansions have been relatively well documented using genetic methods (Lessa et al. 2003; Arenas et al. 2012), more contemporary processes have received much less attention (Hagen et al. 2015).

Theoretical population genetic studies have shown that the genetic patterns created during a spatial population advance are both qualitatively and quantitatively different from those expected based on demographic increase without spatial expansion (Ibrahim et al. 1996; Excoffier & Ray 2008; Excoffier et al. 2009). Purely demographic expansions should: i) maintain stable allele frequencies due to lower rates of genetic drift when compared to stable populations (Kimura and Crow 1963); ii) have both an excess of rare alleles and homozygosity relative to the number of

alleles due to the majority of mutations accumulating over more recent timescales where population size is larger (Maruyama & Fuerst 1984, 1985; Watterson 1986; Tajima 1989); iii) have low levels of linkage disequilibrium even between closely linked loci (Slatkin 1994; Pritchard and Przeworski 2001). However, detection of these patterns can be influenced by the specific demographic histories of natural populations including the levels of gene flow (both past and recent) between any sub-populations (Ray et al. 2003; Excoffier 2004). Conversely, spatial expansion of a population is expected to result in a different set of genetic patterns; i) a decline of genetic diversity from the core population toward the expansion front due to consecutive bottleneck events (Austerlitz et al. 1997), this reduction will be more extreme in species with a low number of long-range dispersers due to the reduced number of founders making up the wave front (Wade and Mccauley 1988; Deshpande et al. 2009; Whitlock and Mccauley 1990); ii) these processes also promote higher genetic differentiation at the expansion front and increased differentiation between sub-populations (Ibrahim et al. 1996); iii) allele surfing, where otherwise rare alleles reach high frequency rates, such allele surfing effects have been shown to rapidly change population genetic structure both in experimental studies using bacteria (Hallatschek, et al. 2007) and in silico using simulations (Excoffier & Ray, 2008). In these instances it is the situation (i.e. being at the wave front) rather than the genetic makeup of individuals that is the advantage; alleles proliferate due to being in the right place at the right time (Moreau et al. 2011) and not (as may otherwise be inferred) due to selection and rapid adaptive change (Currat et al. 2006). In sub-divided populations, as expansion continues, connectivity between sub-populations will be re-established and spatial population structure is expected to decrease due to gene flow, resulting in increased genetic diversity of the super-population (Ibrahim et al. 1996; Excoffier et al. 2009). Likewise, range expansions and their genetic signatures may be effected by gene flow, if there is a large amount of migration between demes during the process of range expansion then this will preserve genetic diversity and the genetic signals will look more like that of a large demographic expansion (Ray et al. 2003; Currat et al. 2004).

The recovery of large carnivore populations across areas of Europe provides a natural experimental design to test these predictions (Chapron et al. 2014; Hagen et al. 2015). Hagen et al. (2015) showed that the population expansion of brown bears in Finland roughly followed the genetic predictions made from simulations, a decline in genetic differentiation and an increase in genetic variability over time as gene flow between sub-populations increased, but that the speed at which genetic variation changed temporally was much faster than predicted by these simulations (1.5 generations instead of 10). Further empirical studies on natural populations that are currently undergoing spatial expansions are needed to determine whether these more rapid genetic changes are common. In addition to the parameters included in simulations to date, natural populations are likely to be undergoing multiple demographic and stochastic processes at

once and as such it is also important to explore how population recoveries proceed under the more complex conditions experienced by natural populations expanding during the Anthropocene.

1.3 Contemporary barriers to gene flow

Spatial patterns of genetic diversity can be influenced by both physical and non-physical barriers. The presence of barriers to gene flow can lead to cryptic genetic structure, even in continuous populations of mobile species (Rueness et al. 2003). In populations experiencing range expansions or recoveries after decline, such barriers could impede both spatial advancement of the expansion through physical barriers (e.g. mountain ranges and/or anthropogenically mediated infrastructure such as roads, Fedorca et al. 2019), and admixture of previously fragmented populations through non-physical barriers (e.g. mating behaviour or communication, Clark et al. 2008).

The field of landscape genetics has been developing over the last 15 years as an intersection between population genetics and landscape ecology (Storfer et al. 2010). Landscape genetics aims to relate to spatial patterns in genetic variation observed in a population to landscape features and determine the impact that these features have on functional connectivity and gene flow (Manel et al. 2003; Storfer et al. 2007). These studies can illuminate differences between assumed barriers to gene flow based on species ecology and actual barriers to dispersal which may be different (Hohnen et al. 2016). However, time is an important factor in the interpretation of landscape genetic results as depending on the equilibrium state of the population, analyses may reflect past rather than contemporary gene flow due to the time-lag problem (Bolliger et al. 2014). Knowledge of the demographic history of the population and factors affecting the likely length of time-lag to equilibrium are therefore important, and sample selection along with the scale over which the analyses are performed should be informed by these factors (Epps et al. 2015).

1.4 Ecological and genetic studies at different scales

The problem of scale and its relation to the patterns observed is a central one in ecology (Levin 1992). Studies have shown that the scale at which observations are made can heavily influence its outcome (Levin 1992, Schneider 1994, Peterson and Parker 1998), and is reflected by the fact that different biological processes happen at differing scales in time and space, such that there is often no single scale at which a system should be studied. This also applies to population genetic patterns, which result from ecological processes acting at different spatial scales and therefore scalar questions also apply to these studies and strongly influence how we infer demographic processes from genetic data.

Fine-scale genetic studies can be used to determine numbers, movements, home range size and dispersal of individuals at a local scale, e.g. within a river catchment (Pagacz 2016). Such data can help us to understand the ecology of a species, especially for more cryptic taxa that are difficult to observe directly (Waits and Paetkau 2005). These ecological traits will then influence patterns, processes and landscape interactions at a population level by contributing to the spatial distribution of genetic variation (Manel et al. 2003; Holderegger and Wagner 2008; Storfer et al. 2007). However, sampling regimes which facilitate the collection of multiple samples from individuals are necessary to achieve sufficient resolution (genetic mark-capture-recapture) and constraints on both time and budget mean that such studies are usually only feasible over relatively small spatial scales (Zhan et al. 2006).

Over larger spatial scales, genetic differences within or among populations can be used to infer population history and highlight areas of genetic differentiation (Pritchard et al. 2000). They may also be used to monitor population recovery post-decline and determine levels of gene flow and how these processes differ temporally and spatially (Hagen et al. 2015; Schregel et al. 2017). Even low levels of gene flow between two populations (e.g. one migrant per generation) are predicted to prevent the accumulation of genetic differentiation (Wright 1931), however the difference between migration (i.e. dispersal) and effective migration (i.e. gene flow) is stark, and thus in natural populations higher levels of migration may be necessary to ensure an acceptable level of effective migration (Mills and Allendorf 1996). Studies taking advantage of current population expansions in large carnivores across Europe are exploring both the relationship between population demographic connectivity and population genetic connectivity (Schregel et al. 2017), and the rate at which sub-populations are becoming admixed, comparing these empirical data with predictions made by population genetic theory using simulation (Hagen et al. 2015). Understanding how these expansions affect genetic variation and how quickly changes are likely to be detected is important for wildlife monitoring as it will provide a guide with which informed management goals can be set (Hagen et al. 2015).

Landscape genetic studies can bridge the two spatial extremes and are usually conducted at an intermediary, medium scale. These population differences in genetic variation and their pattern across the landscape are ultimately mediated by individual movement and species ecology, and as such the rate of spatial expansion as well as any breakdown of structuring over time through contact and admixture of previously isolated populations will also be mediated, at least in part, by these factors (Hagen et al. 2015). Therefore, an understanding of species ecology and in particular dispersal, as well as the effect of various landscape factors on dispersal is key to putting changes in genetic variation in a large scale in context.

1.5 Census and effective population sizes

Historically wildlife conservation and management programs have been developed to count and monitor changes in the number of individuals in a given population or study area over time. This estimation of census population size (N_C) is commonly defined as the number of adults or mature individuals in a population or study area (Luikart et al. 2010). This goal has led to the development of many methods of estimation of census population sizes (Verner 1985), including most recently the application of machine learning and large-scale citizen science projects (Parham et al. 2017; Norouzzadeh 2021). Advances in molecular methods have facilitated the estimation and inference of N_C without the need to directly count individuals (Kendall et al. 2009) which can be especially useful when dealing with elusive or low-density species, large populations or remote locations.

Arguably more important than N_C is the estimation of contemporary effective population size (N_e) which quantifies the number of individuals which effectively contribute to the next generation through the magnitude of inbreeding and genetic drift in a population (Wang et al. 2016). This estimation is crucial in determining the viability of populations as low N_e indicates higher levels of genetic erosion in a population through increased inbreeding and subsequent loss of genetic diversity. The importance of N_e as a key parameter in measuring the maintenance of genetic diversity is exemplified by the call for its inclusion in the Convention of Biodiversity 2020 targets (Hoban et al. 2020).

England et al. (2011) showed that changes in N_e over time can provide an early warning signal of population fragmentation, a key contemporary driver of genetic diversity loss (Allendorf and Luikart 2006). Conversely N_e estimates may also be used to monitor population recoveries, although work in this area is less substantive (Tallmon et al. 2012).

1.6 Eurasian otter ecology

The Eurasian otter (*Lutra* lutra L. Linnaeus 1978 – from here on referred to as 'otter') is the most widespread of the 13 species of otter in the family Lutrinae (Corbet 1966; Figure 1.1). However, despite being widely distributed, otters (like many carnivores), are difficult to study as their elusive and predominantly nocturnal nature means that some aspects of their ecology, particularly their socio-biology, remain largely unknown (Chanin 2003; Kruuk 2006). Otters occupy linear home ranges, along freshwater habitats such as rivers and are highly mobile, capable of traversing large distances (Jenkins 1980; Sjoasen 1997; Durbin 1998; Pagacz 2016). While earlier work suggested large home range sizes of 20-30 km (Erlinge 1967), a more recently study in Ireland found adult female ranges covering just 7.5 km ± 1.5 km (Néill et al. 2009). Studies of otter home-range and dispersal mechanisms have been largely limited to radio-telemetry and surveying for otter signs (spraint, footprints, holts). Radio-telemetry studies are limited to a small number of

individuals due to the costs involved and the difficulty and ethics of trapping otters (Néill et al. 2009; Quaglietta et al. 2014), while surveys using otter sign can only determine presence-absence not the number of individuals or specific individual movements (Conroy and Chanin 2000).

Figure 1.1: Range map of *Lutra lutra* (modified from the IUCN red list; http://maps.iucnredlist.org/map.html?id=12419). The orange areas of the map show current presence of the species.

1.7 Population history in the UK and current population structure

Otters experienced a dramatic population decline in the UK during the 1950's-1970's (Coxon et al. 1999), replicated range-wide (MacDonald and Mason 1983). This decline was attributed largely to the use and bioaccumulation of pesticides such as dieldrin and polychlorinated biphenyls (PCBs), as well as habitat destruction and direct persecution by humans (Conroy and Chanin 2000; Mason and Macdonald 2004). As a result, the population in Great Britain declined both in numbers and range, and by the 1970's was predominantly confined to three strongholds in Scotland, Mid and West Wales and South-West England, although very small remnant populations in East-Anglia and North Yorkshire also existed (Jones & Jones 2004). Subsequently, legislative protection for both otters (Wildlife and Countryside Act, 1981) and their freshwater habitats (EU Habitats Directive, 1992) in the UK resulted in population expansion, confirmed by an increased number of sites with positive otter spraint signs in successive national surveys, indicating recolonisation of the historic range (Crawford 2003; Jones & Jones 2004; Strachan 2007; Crawford 2010). The majority of this recolonisation has been natural, through breeding and dispersal, however in East Anglia; where only a small remnant population remained; a reintroduction program released 117 otters between 1983-1999 (Jessop & Cheyne 1992) aiding the recovery in this area, additionally there were also reinforcement releases in North Yorkshire. In general, the otters released into Southern parts of England were captive bred by the Otter Trust, and those released in Northern England

were individuals rehabilitated by the Vincent Wildlife Trust (Green 1997). While the rehabilitated individuals would have been of UK origin, the history of the captive bred individuals is more dubious and likely involved otters of non-UK origin (Hájková et al. 2007).

Studies on the genetic structure of natural otter populations in Great Britain have shown that the variation observed matches the population history of isolated sub-populations which became genetically distinct from one another (Hobbs et al. 2011; Stanton et al. 2014). The five genetic clusters observed broadly match the locations of stronghold populations, with high F_{ST} values observed between them (Stanton et al. 2014). The scent profiles of otters from these distinct genetic clusters also vary similarly, creating 'odour dialects' across the UK (Kean et al. 2017). Genetic monitoring across time should indicate how population expansion is being contributed to by these different genetic clusters and whether F_{ST} between sub-populations is, as would be expected theoretically, degrading over time.

1.8 Advancing knowledge through studying recoveries in complex populations

The increase in numbers of otters in the UK offers a great opportunity to advance our understanding of dynamic population genetic changes in the wild. Previous work on brown bears in Finland (Hagen et al. 2015) provided some insight beyond theoretical modelling, however this population also had a relatively simple geographic structure as there were only two sub-populations spanning a North – South range. There are several ways in which the UK otter population adds more complexity to this, an important advance in determining similarities and differences of the process between geographies and species. Firstly, there are a greater number of sub-populations across the UK as identified by previous work (Hobbs et al. 2011; Stanton et al. 2014) and thus the effective migration between them and genetic differentiation between is necessarily more complex. Secondly, the UK is topographically complex with diverse land-use systems, comprising a collection of peninsulas with uncrossable (to otters) tracts of sea in between, making movement more complex than in any study (theoretical or empirical) to date since the species needs to use a diversity of anthropogenically altered landscapes to navigate the regions. Thirdly, otters are semi-aquatic and are therefore more closely associated to the vast network of freshwater rivers and lakes and are unlikely to move freely across land in the same way as terrestrial species such as brown bears. These added complexities offer opportunities for fundamental advance in knowledge of expanding populations that goes well beyond the more simplistic theoretical models and previous empirical studies. Greater understanding of these complexities will aid the utility of genetic monitoring to on the ground conservation efforts in these instances.

The uneven distribution of otters across the landscape in England and Wales also provides the opportunity to investigate how areas with higher otter density (i.e. Wales and South-West

England) might differ from areas of lower otter density (i.e. Northern, Central and Eastern England). The findings from successive national surveys indicate that the presence of spraint is higher in some areas than others but that all areas have a generally increasing incidence of otter sign over time (Crawford 2010; Stachan 2015). If the higher occurrence of otter signs can be used as a proxy for higher otter density (as is done currently) then it is possible to examine differences in effective migration between high density and lower density sub-populations as well as exploring any affects density has on dispersal behaviour. A study on otters across a major drainage divide between Poland and Slovakia (Pagacz 2016) detected individuals crossing this mountain range in an area with high population density. Although it is known that otters can live at high altitudes i.e. up to 4,120 meters in Tibet (Ruiz-Olmo 1998), in general the literature suggests that otters avoid these areas in Europe due to lower food abundance (Ruiz-Olmo et al. 2001) or the presence steep, dry slopes (Janssens et al. 2008). However, it may be that once population density becomes high enough, these usual habitat preferences and landscape barriers to movement break down as territories become scarce. These too are important questions that can be addressed through genetic samples that will fundamentally advance our knowledge of how genetic monitoring can inform conservation.

1.9 Roadkill as a resource for research

Obtaining high quality DNA samples (tissue or blood) from elusive species, such as otters, can be difficult. However, every year many animals are killed on our roads, and in the UK several hundred of these are otters (Chadwick 2006). The proliferation of citizen science over the last decade has allowed data and sample collection on scale previously impossible (Devictor et al. 2010). One such area of study which has greatly benefited from citizen science is road-kill monitoring (Vercayie and Herremans 2015), but while there are many initiatives globally very few collect and archive samples. In the UK, the Cardiff University Otter Project uses public reports of dead otters to run a nation-wide monitoring and archiving scheme (Chadwick 2006).

As the population density of otters has increased, so has the number of deaths due to road traffic accidents (RTAs) which are now one of the leading causes of mortality for the otter in Europe (Philcox et al. 1999; Hauer et al. 2002). These mortalities, however, are a valuable resource for science and study of the species. The Cardiff University Otter Project has been collecting and performing post-mortems on road killed otters from across the UK since the early 1990's and thus has built up an archive of over 30 years' worth of samples. Tissue samples taken from the leg muscle during post-mortem provide a high-quality source of DNA from otters across the UK that is both temporally and spatially distributed. Many additional observations are recorded, and genetic samples are taken during *post-mortem*, meaning that each individual has in depth metadata including age class, sex and location found. As such, this archive provides a unique opportunity to

study any changes in population structure and genetic diversity as a native species recolonises an historic range.

1.10 Thesis Aims

This thesis represents the first true incidence of genetic monitoring i.e. repeated sampling over time, as opposed to genetic assessment i.e. a one-time sampling event, that has been conducted on the Eurasian otter. Despite the term 'genetic monitoring' having been used for decades, examples of true genetic monitoring remain rare (Schwartz et al. 2007). Even though previous genetic work conducted on otters in the UK has spanned wide time-periods (Hobbs et al. 2006, Hobbs et al. 2011, Stanton et al. 2014), these previous studies would be considered genetic assessments due to the pooling of data across the time frame. The novelty of the work presented in this thesis resides in the use of separate temporal datasets allowing the assessment of dynamic changes in the genetic variability and structure over time. This temporal element is of critical importance to conservation and management as it is the only way to determine, in this case genetically, if the population in question is moving in a favourable or unfavourable direction. This monitoring takes on extra importance for species such as otters in the UK, that are undergoing rapid changes such as population expansions.

Despite the long-standing knowledge that genetic diversity is a pillar of overall biodiversity and its importance to both the evolutionary potential of species and the viability of populations (Lande 1988), population recovery at the genetic level in wild populations remains poorly studied and understood (Hagen et al. 2015). Theoretical studies using simulations have shown characteristic changes to genetic diversity in expanding populations, however empirical data are rare. Data from a number of empirical studies across multiple taxa on rates of genetic changes and timescales over which they occur is required to understand how these processes manifest themselves in wild populations, whether characteristics including patterns of genetic differentiation are still detectable and how this information might help inform management and biodiversity targets. The objective of this thesis is to address some of these outstanding questions by investigating the characteristics and constraints on population recovery in a wild species using otters in the UK as a case-study.

More specifically I address the following aims in each data chapter:

Chapter 2

i) Quantify population structure and regional genetic diversity of otters in the UK using a large-scale spatial and temporal dataset and determine changes in genetic diversity and differentiation over space and time.

ii) Use these data to assess the degree of genetic recovery that has occurred in the UK otter population over the last 20 years and contextualise this alongside the well documented spatial and demographic recovery of otters from national survey data.

Chapter 3

i) Explore landscape factors affecting this recovery, specifically focusing on contemporary functional connectivity in otters across the Wales and Borders region using the most recent developments in landscape genetic methods.

ii) Assess the influence of scale on landscape predictors through conducting analyses at both a fine (100 m^2) and coarse scale (1 km^2) lens.

Chapter 4

i) Assess contemporary changes in population size across Wales and England using genetic methods.

ii) Estimate the effective population size of otters across Wales and England and compare this to population sizes estimated from national survey data.

iii) Use temporally replicated sampling to estimate changes in effective population size over time, a key determinant of long-term population viability.

2

Country-wide genetic monitoring over 20 years reveals spatial connectivity with persistent population structure in a recovering and expanding carnivore population

Abstract

Population declines are common for many terrestrial mammals during the 20th and 21st centuries, due largely to anthropogenic pressures. For some species, including the Eurasian otter *Lutra lutra,* environmental and legal protection has since led to population growth and recolonisation of previously occupied distribution ranges. While heralded as conservation successes, few case studies have specifically addressed the extent of the achieved population recovery from a genetic perspective, i.e. whether genetic variability and connectivity have been restored to previous norms. We here use large-scale genetic monitoring data from UK otters, whose population underwent a well-documented population decline between the 1950's to 1970's, to explore the dynamics of a population re-expansion over a 20-year period. Otters from across Wales and England were genotyped at five time points between 1994 and 2014 using 15 microsatellite loci. We used this combination of long-term temporal and large-scale spatial sampling to evaluate 3 genetic hypotheses on range expansions; that genetic structuring would weaken over time, geneflow between sub-populations would increase over time, and that the genetic diversity of previously isolated populations would increase. Our results show that, over time, geneflow increased between regions across the study area, but genetic structuring remained strong and genetic diversity did not increase. These findings highlight an underappreciated aspect of population recovery of endangered species, that genetic recovery may often lag behind the processes of spatial and demographic recovery, and that physical connectivity of populations does not always mean functional connectivity.

2.1 Introduction

2.1.1 Population expansions

Many large mammal species experienced population declines in the last century, largely due to anthropogenic causes (Cardillo et al. 2005). Some of these declines are now being reversed due to protective legislation, and population recoveries are being observed (Chapron et al. 2014). These declines and subsequent re-expansions into historically occupied ranges provide natural experiments of how genetic diversity and structure change both temporally and spatially during

population growth and allow for the testing of theoretical predictions from simulation studies (Hagen et al. 2015).

Theoretical studies have shown that population expansions are likely to be accompanied by changes in genetic diversity (Excoffier and Ray 2008), which may differ from the changes caused by demographic growth alone (Excoffier, Foll and Petit 2009). During population expansions sequential founder events can cause unusual phenomena, including reduced genetic diversity at the wave front and 'allele surfing' where specific alleles may reach much higher frequencies than expected, due to random genetic drift combined with the relative isolation of the founder individuals from the rest of the population (Excoffier and Ray 2008). Such phenomena can be difficult to distinguish from signals of selection and have only rarely been shown in empirical studies (Moreau et al. 2011, Graciá et al. 2013). As the population continues to expand, connectivity may be established between previously isolated groups or individuals. If this connectivity results in effective migration, decreases in spatial population structure via establishment of gene flow between differentiated groups should be observed, resulting in increased genetic diversity within groups (Ibrahim, Nichols and Godfrey 1996; Excoffier, Foll and Petit 2009).

Wildlife populations that have experienced large declines are often fragmented in the process, leading to differentiation through genetic drift (Rueness et al. 2003, Manel et al. 2004). Subsequent recovery of these populations through expansion is therefore predicted to reduce spatial genetic structure through gene flow by re-establishing connectivity between previously isolated genetically differentiated units (Hagen et al. 2015). However, detecting this connectivity requires temporal sampling of the population to determine changes in geneflow, genetic differentiation and genetic diversity between and within sub-populations over time. Such information quantifying the dynamics of range expansions is vital in order to make predictions about the nature of population recoveries, and to develop and monitor progress towards realistic management goals.

2.1.2 Otter populations in the UK

In the UK, otters are undergoing a population expansion, recovering from a well-documented large-scale decline in the second half of the 20th century (Crawford 2010; Strachan 2015). This decline was largely attributed to the use and bioaccumulation of pesticides such as dieldrin and polychlorinated biphenyls (PCBs), although habitat loss and direct persecution through hunting may also have contributed (Conroy & Chanin 2000; Mason & MacDonald 2004). Resulting population declines since the 1950s led to extinction of otters across large areas of England (Figure 2.1), and survival in geographically isolated and sparsely populated strongholds. Previous work has shown these stronghold populations, located in Western Wales, South-West England

and Scotland, to have relatively low genetic variability, and are moderately to strongly differentiated from each other (Hobbs et al. 2011; Stanton et al. 2014). Subsequently, legislative protection for both otters (Wildlife and Countryside Act, 1981) and their freshwater habitats (EU Habitats Directive, 1992) in the UK halted the population decline, and national survey data (Matthews et al. 2018) has indicated that the population has been expanding both in range and in population size (Figure 2.1) since the 1980s. This apparent spatial and demographic recovery has seemed to re-connect previously isolated sub-populations, which are now contiguous once again.

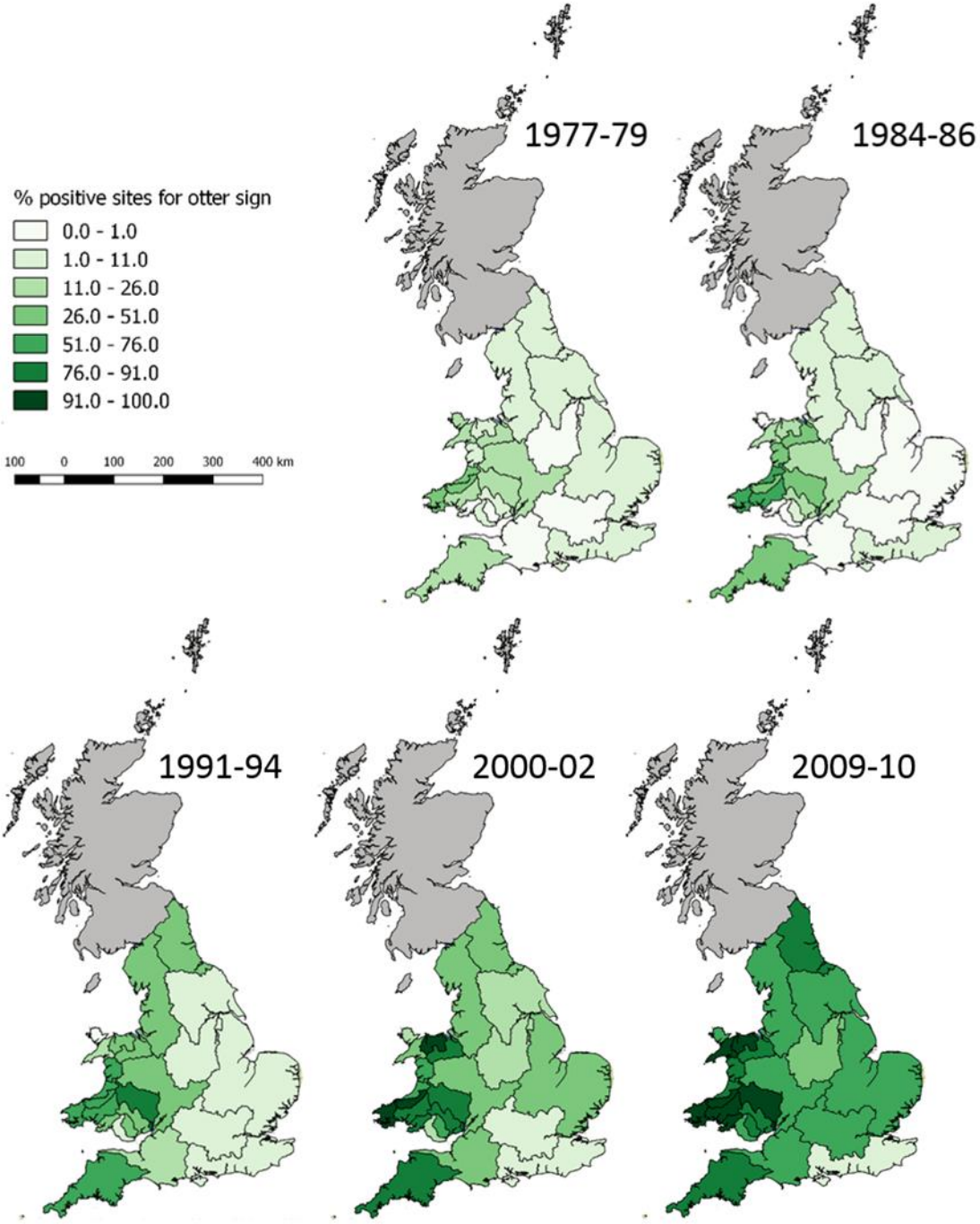


Figure 2.1: Collated national survey data for otters in Wales and England from 1977 to 2010. Maps show increased frequency of otter sign and increased distribution of otters over time. (Data for Wales from: Crawford et al. 1979, Andrews & Crawford 1986, Andrews et al. 1993, Jones & Jones 2004, Strachan 2015. Data for England from: Lenton et al. 1980, Strachan et al. 1990, Strachan & Jefferies 1996, Crawford 2003, 2010).

Previous work on the population structure of otters in the UK showed genetic differentiation linked to geographic location. This sub-structuring reflected the locations of previous population strongholds (e.g. in Southwest England and Wales) and reinforcement programs (e.g. northeast England and southeast England), indicating strong north-south genetic differentiation, as well as five genetically distinct sub-populations (Hobbs et al. 2011, Stanton et al. 2014 – see Supplementary 2.1).

The otter therefore provides an ideal case study with which to test the following hypotheses:

1. Genetic structuring across the study area will weaken over time as population recovery and expansion proceed, reconnecting previously isolated sub-populations.
2. Geneflow between sub-populations/regions will increase over time with increased contact due to range expansions.
3. Genetic diversity of previously isolated sub-populations will increase as geneflow between neighbouring sub-populations allows the influx of new alleles.

2.2 Methods

2.2.1 Sampling

Road-killed otter carcasses from across Wales and England were collected and their locations recorded as part of the national Cardiff University Otter Project program (www.cardiff.ac.uk/otter-project). Muscle samples were taken from the hind leg and stored in ethanol at -20 °C. Two hundred of these samples were selected from 2009 & 2014 (104 and 96 from each respective year, including 7 samples in 2009 from Stanton et al. 2014). Samples were selected by splitting the study area into 20km squares and randomly selecting one sample from every 20km grid square in which at least one dead otter was collected that year, thus avoiding potential pseudoreplication that would arise from true random sampling. An additional 207 previously genotyped samples from a prior study (Hobbs et al. 2011) were selected at time points 1993-1995, 1998-2000 and 2004 (see Supplementary 2.2 for data allocation to each study).

Each sample was mapped as a point location in ESRI ArcGIS 10.3, and each point allocated to a river basin district (as specified in the Water Framework Directive Cycle 2, Environment Agency 2015; Natural Resources Wales 2015). Regional aggregations of river basin districts (RBD regions) were used to ensure adequate sample size for certain analyses (Table 2.1). These RBD regions were combined with the temporal data (sampling year) to yield 23 spatial-temporal groupings (STGs) for further analysis.

Table 2.1: Geographic regions used in genetic diversity and differentiation analysis of UK Eurasian otters (*Lutra lutra*). RBD Regions are based on amalgamations of River Basin Districts (RBDs) as defined in the Water Framework Directive Cycle 2 (Environment Agency 2015; Natural Resources Wales 2015), N is number of otters genotyped and Land Area is the total km^2 of land within an RBD region (no Land Area is provided for the 'other' RBD region as it covers a vast and unconnected area of land with very few samples).

RBD Region	Included river basin districts (RBD's)	N	Land Area (km^2)
Eastern	Anglian, South East, Thames	74	50,226
Northern	Humber, North West, Northumbria, Solway Tweed	59	61,601
Severn	Severn	84	21,056
South West	South West	77	18,191
Western Wales	Dee & Western Wales	102	14,715
Other	Ireland & Scotland	11	NA

2.2.2 DNA extraction and microsatellite genotyping

DNA was extracted from muscle tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's guidelines. Samples were genotyped at 15 microsatellite loci using three multiplexes as in Hobbs et al. (2006). The 15 loci were: Lut 435, 457, 604, 615, 701, 715, 717, 733, 782, 818, 832, 833, and 902 (Dallas and Piertney 1998; Dallas et al. 1999), and 040T05 and 040T22 (Huang et al. 2005). Polymerase chain reactions were conducted in 10 µl reactions with 1x (5 µl) QIAGEN Multiplex PCR Master Mix (Qiagen), 0.2 µM of each primer, sterile water and 2 µl of template DNA. The cycling conditions used were as follows: 95 °C for 15 mins; 29 cycles of 94 °C for 30 s, 58 °C for 90 s, 72 °C for 60s; and a final extension at 60 °C for 30 mins. Samples selected from the previous studies (Hobbs et al. 2011; Stanton et al. 2014) had been genotyped at the same 15 loci using the same conditions. Fluorescently labelled PCR product was sent to DNA Sequencing Services (Dundee, Scotland) for fragment analysis using an Applied Biosystems 3730XL DNA analyser and visualised using Genemapper 4.0 (Applied Biosystems 2006). A sub-set of 14 previously genotyped samples (Hobbs et al. 2011) were also re-analysed to allow calibration of fragment size scoring between the three studies (Stanton et al. 2014 calibrated their data to Hobbs et al. 2011, therefore calibration using the 14 samples from Hobbs et al. 2011 was sufficient to bring all three sets of data together). Altogether, this procedure enabled adjustment of allele sizes to account for variation between sequencing platforms and scorers, allowing the dataset to be analysed as a whole (see 'Results' for details).

2.2.3 Genetic variability by locus, within spatio-temporal groupings (STGs) of River Basin District (RBD) regions

Genotyping errors and null allele frequencies were estimated using Microchecker V2.2.3 (Van Oosterhout et al. 2004). Each spatial-temporal grouping (STG) was run independently, and loci were only removed if they were identified as having null alleles in the majority of regions at a specific time point (D'Urban Jackson et al. 2017). Null alleles are a potential source of bias during

the estimation of population differentiation (F_{ST}) therefore it is important to identify them if present. An exact test of Hardy Weinberg Equilibrium (HWE), and a test for linkage disequilibrium (LD) between all pairs of loci were conducted in ARLEQUIN 3.5 (Excoffier & Lischer 2010). Deviation from HWE was estimated using 1,000,000 Markov chain steps and 100,000 dememorisation steps. Linkage disequilibrium between loci was estimated using 10,000 permutations with the number of random initial conditions set to 2 and a significance level of 0.05. These analyses were run on the whole dataset (i.e. for all STG's combined) initially and then repeated on the samples in each STG individually, for LD only time points 2009 & 2014 were included in the analysis as a representative sample as for true linkage disequilibrium you would expect the same result at any given time point and these data had full spatial coverage over the study area and the largest sample size.

Genetic diversity per locus was estimated using MICROSATELLITE TOOLKIT (Park 2001). Unbiased expected heterozygosity (Nei 1987), observed heterozygosity and number of alleles per locus were calculated along with the mean of each of these statistics. A paired 2-sample t-test was used to test for a significant difference between the expected and observed heterozygosity's using R studio (R version 3.4.3, R Core Team 2017).

Genetic diversity was estimated for each STG using multiple diversity statistics. We used MICROSATELLITE TOOLKIT (Park 2001) to estimate an unbiased estimator of expected heterozygosity, based on Nei's unbiased gene diversity (Nei 1987) and observed heterozygosity including standard deviations. Expected heterozygosity and observed heterozygosity are more robust to sample size changes, when sample size is greater than 5-10 individuals (a threshold which was met in the majority, 21/23 for N>5 and 19/23 for N>10) of our STGs), however it is particularly important that standard deviations are reported and considered alongside estimates for smaller sample sizes (Pruett and Winker 2008). HP-RARE v1.1 (Kalinowski 2005) was used to estimate allelic richness and private allelic richness for each STG; initially this was calculated using the smallest sample size at each individual time point respectively. As samples size fluctuated both between regions and time points, we considered the bias that these fluctuations might introduce into each analysis. For analyses that were sensitive to such data fluctuations (e.g., allelic richness), we used a resampling approach where the resample size was set to the smallest sample size in any RBD region for that year (Pruett and Winker 2008) to allow comparisons to be made across all RBD regions at a particular time point. A second analysis was run where each dataset was resampled at to the smallest sample size across all time points (N=6) to account for sample size bias in the estimations and allow comparisons to be made across all STGs. Additional statistical analyses of genetic diversity estimates were carried out using R; correlations between both allelic richness and private allelic richness under both resampling regimes were assessed using Kendall's Tau (non-parametric rank correlation method that allows for tied values), as was

the relationship between observed heterozygosity and RBD region land area. A one-way ANOVA was performed to determine if there was any significant association between RBD region and observed heterozygosity, with a Tukey *post hoc* test to test pairwise differences.

The inbreeding coefficient (F_{IS}) was initially estimated based on all individuals from each STG, using FSTAT version 2.9.3.2 (Goudet 1995), and tested for significant deviation from 0 using a 1-sample T-test in R with correction for false discovery rate (FDR) using the Benjamini-Hochberg method to account for multiple testing (Benjamini and Hochberg 1995). Population structure and admixture can affect F_{IS} estimates, however. For example, if individuals from more than one genetic cluster are analysed as a single population, a deficiency of heterozygotes is likely to be observed due to violation of the Hardy-Weinberg assumption of a single, randomly mating population (Hardy 1908; Weinberg 1908; Waples 2015). This deficiency produces a positive F_{IS} value, and thus a false indication of inbreeding, and is termed the 'Wahlund effect' (Wahlund 1928). In order to distinguish between the Wahlund effect and actual recent population-level inbreeding, we therefore recalculated F_{IS} using only individuals identified as belonging to the dominant genetic cluster for the STG, defined as follows. First, cluster assignments and admixture proportions were calculated for each STG using STRUCTURE 2.3.4 (Pritchard, Stephens and Donnelly 2000), at the smallest value of K which maximised global likelihood (Kalinowski 2011). For each STG we then identified individuals as either belonging to the dominant cluster, or to one of the two minor clusters, based on their proportional allocation to each. Individuals not belonging to the major cluster were excluded from the recalculated value of F_{IS}, and comparison of the two F_{IS} estimates allowed us to highlight any results potentially explained by Wahlund effects. Individuals were considered "admixed" if their assignment proportion was less than 0.8 to a single cluster (Rutledge et al. 2010 & Heppenheimer et al. 2018), remaining individuals were considered "non-admixed".

2.2.4 Population structure and geneflow

Population differentiation was estimated between STGs in ARLEQUIN 3.5 (Excoffier & Lischer 2010), using the pairwise F_{ST} estimator by Weir & Cockerham (1984) with the 'number of different alleles (F_{ST}-like)' option and 10,000 permutations and significance level set at 0.05. Further, an exact test of population differentiation (Raymond and Rousset 1995) was conducted, with 100,000 Markov chain Monte Carlo (MCMC) steps, 10,000 dememorisation steps and significance level set at 0.05. Analysis of Molecular Variance (AMOVA) was also performed in ARLEQUIN 3.5, with 10,000 permutations for significance and 1000 permutations for mantel test, using the 'number of different alleles (F_{ST}-like) option for molecular distance.

We tested for isolation by distance and spatial autocorrelation using pairwise matrices of land distance and genetic distance estimates for each time point. Land distance was calculated using the GDISTANCE package in R to determine the least-cost pathway between each pair of otters at each respective time point from a rasterised map of Great Britain with cell size set to 1 km^2. On an irregularly shaped island like Great Britain, with multiple peninsulas separated by sea, land distance was deemed a more realistic measure of physical space between otters than Euclidean distance. Each land cell was given a resistance of 1 (sea cells were classified as 'NoData'), such that the least-cost resistance estimates calculated reflected the land distance between each pair of otters. For genetic distance, the proportion of shared alleles was calculated between pairs of otters using GenAlEx 6.5 (Peakall and Smouse 2012). Mantel tests and Mantel correlograms were performed using the package VEGAN in R using Pearson's correlations and 10,000 permutations with an $\alpha = 0.05$. The Holm correction for testing multiple p-values was used for the Mantel correlograms, and break points for distance classes were set to every 50 km from 0 km to 800 km. Potential sex differences in isolation by distance and spatial autocorrelation were additionally evaluated by applying the same testing to subsets of the data including only adult females, and only adult males.

Geneflow between STGs was estimated using two different methods, for four time points spanning 1999-2014 (the first time point, 1994, was omitted due to small sample size and lack of geographic coverage across the whole study region). Firstly, GENEPOP v4.6 (Rousset 2008) was used to estimate the effective number of migrants (Nm), corrected for sample size, using the private alleles method developed by Barton & Slatkin (1986) which should be most sensitive to recent migration due to the rare nature of private alleles (Yamamichi & Innan 2012). Nm was estimated across the whole dataset, and pairwise between all STGs. Secondly, BayesAss v3.0 (Wilson and Rannala 2003) was used to estimate pairwise directional gene flow between regions. This program uses a Bayesian algorithm to estimate recent migration (last 2-3 generations) between specified populations. Initially the program was run with the default values (of 0.1) for the three continuous parameters (migration rates (Δ_M), allele frequencies (Δ_A) and inbreeding coefficients (Δ_F)). Subsequently, these three parameters were adjusted until acceptance rates were within the recommended bounds of 20-60% (Rannala 2007), resulting in selection of $\Delta_M = 0.3$, $\Delta_A = 1.0$ and $\Delta_F = 1.0$ for all time points. Three runs were performed per time point using different random seeds (starting points) with 10,000,000 MCMC iterations following a burn-in of 1,000,000 MCMC iterations and a sample interval of 5000. Trace output files were recorded and used to monitor for mixing and convergence using TRACER v1.7.1 (Rambaut and Drummond 2003) and the three runs were averaged to obtain the percentage migrants between each region in a pairwise fashion. Migration rates between the Western Wales and Severn regions and the Northern and Eastern regions respectively were unlikely to provide reliable results as the pairwise

F_{ST} between these regions was <0.05 and as such the results for these pairwise estimates were discarded (Faubet, Waples, and Gaggiotti, 2007).

To determine the extent of population structure within UK otters we used two complementary approaches, one parametric and one non-parametric. The first was a Bayesian clustering algorithm implemented in STRUCTURE 2.3.4 (Pritchard, Stephens and Donnelly 2000), with no location prior, using the admixture model with correlated allele frequencies. All samples, regardless of collection year, were run together as one dataset for K=1 to K=13, with a burn-in of 100,000 followed by 1,000,000 MCMC steps, running 10 replicates for each value of K. We chose K=13 as the maximum K value based on the 11 river basin districts in Wales and England, plus Ireland and Scotland. The results were summarised using STRUCTURE HARVESTER v 0.6.94 (Earl & vonHoldt 2012), and we used the method by Evanno et al. (2005) and the likelihood of K (Pritchard & Wen 2003) to explore the most likely number of clusters present in the data. Biologically sensible values of K were investigated further (Janes et al. 2017). Individual admixture proportions were averages across the 10 runs for each K using CLUMPAK using default parameters (Kopleman et al. 2015). Based on a cut-off of 0.8 for the proportion of assignment to a specific cluster, we determined the number of 'admixed' individuals (Rutledge et al. 2010, Heppenheimer et al. 2018) at each time point for each value of K (i.e. any individual with less than 0.8 assignment to a single cluster was considered admixed). We used these data to quantify the percentage of admixed individuals across all clusters in the dataset at each time point.

The second approach used was a *Discriminant Analysis of Principal Components* (DAPC), a multivariate approach (Jombart, Pontier, & Dufour, 2009) that avoids making strong assumptions about the underlying genetic model (such as populations being in Hardy-Weinberg and linkage equilibrium). We used this approach as implemented in the R package *adegenet* (Jombart, Devillard & Balloux 2011). Firstly the number of *de novo* clusters was estimated using *find.clusters*, and for each model a Bayesian Information Criterion (BIC) was computed. An optimal K or a range of K values was then selected based on the lowest BIC value or the steepness of the gradient between K's on the graph (Jombart et al. 2010; Miller et al. 2020). Subsequently a DAPC was conducted using these pre-defined groups (K).

Progressive partitioning (Hobbs et al. 2011) based on Bayesian clustering results from STRUCTURE was conducted independently for the three time periods (2004, 2009 and 2014, i.e. excluding 1994 and 1999 for which samples were not available from the South West), based on data for all regions. As before, a burn-in of 100,000 followed by 1,000,000 MCMC steps was used, restricting K to K=2 for 5 replicate runs. For each run, individuals were assigned to one of the two clusters based on >50% assignment and each cluster went through another round of partitioning at K=2 until assignment of individuals was ~50% to each cluster.

2.3 Results

2.3.1 Genetic variability by locus and River Basin District (RBD) region

Full genotypes (for 15 microsatellite loci) were obtained for all 193 samples analysed in this study (100% genotyping success rate), and re-analysis of the 14 calibration samples allowed an additional 214 genotypes selected from previous studies (Hobbs et al. 2011; Stanton et al. 2014) to be adjusted such that the datasets could be merged and analysed as one. Across years, none of the loci showed significant evidence ($p<0.05$) of null alleles at >2 of the 5 geographic regions, and therefore all 15 loci were retained for further analysis. All 15 loci were polymorphic, with the number of alleles per locus ranging from 6 to 11 and observed heterozygosity from 0.40 – 0.70 (Table 2.2).

Table 2.2: Genetic variability and information on loci. Multiplex indicates which of the 3 primer multiplex mixes each locus belongs to, dye refers to the fluorescent dye used to label the PCR product, N_A: the number of alleles detected at each locus, size range: states the size range of the alleles at that locus in number of base pairs (bp), H_e: unbiased expected heterozygosity and H_o: observed heterozygosity. Mean values across all loci for N_A, H_e and H_o are given. A single asterisk* indicates that the locus was out of HWE in at least one RBD region, two asterisks** indicate that the locus was out of HWE in the majority (>3) of RBD regions.

Locus	Multiplex	Dye	N_A	Size range (bp)	H_e	H_o
Lut435*	1	Fam	11	117-145	0.63	0.50
Lut453*	1	Hex	10	117-135	0.69	0.53
04OT05**	1	Hex	7	171-191	0.75	0.63
Lut717**	1	Ned	7	175-203	0.59	0.44
04OT22*	1	Fam	8	138-164	0.75	0.59
Lut604	2	Fam	7	127-137	0.72	0.62
Lut733	2	Fam	7	156-182	0.70	0.50
Lut615**	2	Fam	11	214-231	0.77	0.59
Lut902**	2	Hex	11	145-182	0.74	0.61
Lut782*	2	Ned	6	161-196	0.46	0.40
Lut818*	3	Fam	8	158-188	0.74	0.64
Lut701	3	Fam	9	193-248	0.66	0.49
Lut833*	3	Hex	7	154-176	0.75	0.70
Lut715	3	Hex	6	187-216	0.62	0.52
Lut832**	3	Ned	8	177-197	0.67	0.47
Mean	-	-	*8.2*	-	*0.68*	*0.55*

Observed heterozygosity was consistently and significantly smaller than expected heterozygosity across all loci (paired two-sample $t_{14} = 11.219$, $p<0.001$; Table 2.2), potentially indicating population sub-structuring (Jin and Chakraborty 1995). When the loci were tested for Hardy-Weinberg equilibrium (HWE) across all sub-populations and time points, all loci departed significantly from HWE expectations, and significant linkage disequilibrium (LD) was also observed across all loci. When analysing each RBD region independently, different loci were found to deviate significantly from HWE in different RBD regions: 11 of the 15 loci deviated significantly

from HWE in at least one RBD region, but only 5 of those loci deviated significantly in the majority (>3) of RBD regions.

Observed heterozygosity was also consistently smaller than expected heterozygosity across all STG's (Table 2.3). Across all regions (using data from 2009 & 2014 only) many pairs of loci showed linkage disequilibrium over one or more regions, but none had significant linkage disequilibrium over all five. This difference in the results between the pooled, and the RBD-based analysis indicates Wahlund-type geographic population structure in the dataset.

All diversity indices suggested greater genetic diversity in the Eastern RBD regions (Eastern and Northern) than the Western RBD regions (Severn, South West, Western Wales) (Table 2.3). Genetic diversity within STG's as measured by H_e ranged from 0.48 (Western Wales, 1994) to 0.73 (Eastern, 2009) with an average of 0.61, while H_o ranged from 0.44 (Western Wales, 1994) to 0.70 (Northern, 1994) (although the latter estimate was based only on two individuals and therefore should be treated with caution, see Table 2.3). The highest value for H_o from a region with sample size >5 was 0.69 (Eastern, 1999) with an average of 0.56. Higher genetic diversity, as estimated by observed heterozygosity (H_o) was generally found in larger RBD regions (Northern and Eastern) while the lowest estimates were consistently found in the smallest RBD region (Western Wales) (Kendall's Tau Correlation: Tau = 0.49, p = 0.004). This was not reflective of sample size (N) (Kendall's Tau Correlation between Ho and N: Tau = -0.07, p = 0.6484). There was a significant difference in H_o between RBD regions (one-way ANOVA: $F_{4,16} = 28.25$, $p<0.001$), with significantly lower Ho in the Severn, South-West and Western Wales regions when compared with Eastern and Northern regions (Tukey post hoc test: $p<0.001$ for all six comparisons). There was no significant difference in H_o between the Eastern and Northern regions, or between the Severn, South West and Western Wales regions (for full results see Supplementary 2.3) indicating a significant West-East divide in this genetic diversity estimate.

Other measures of diversity followed a similar pattern, of higher genetic diversity in the Eastern than the Western RBD regions (Table 2.3). This was true for the average number of alleles per locus (N_A) allelic richness (A_R) and private allelic richness (pA_R). A_R and pA_R were calculated using the smallest sample size across the regions for that time point as the resampling number, and recalculated using the smallest (reasonable) sample size of any region at any time point (N=6) as the resampling number, to allow comparisons unbiased by sample size. Although estimates were lower when resampling size was smaller, as expected, there was a highly significant correlation between Ar and pAr estimates using both resampling techniques (Kendall's Tau = 0.78, $p<0.001$), indicating that uneven sampling across time does not impact overall conclusions about regional genetic diversity.

The F_{IS} estimates suggested significant inbreeding in the majority of RBD regions at certain time points, but many incidents of apparent inbreeding were likely due to the Wahlund effect (annotated with *i* and *w* respectively in Table 2.3, with additional detail in Supplementary 2.4).

Table 2.3: Genetic diversity statistics for all River Basin District regions (RBD regions) by year. N: number of individuals, N loci: number of loci retained for analysis, H_e: Nei's unbiased expected heterozygosity, H_o: observed heterozygosity, N_A: average number of alleles per locus, A_r: allelic richness, pA_r: private allelic richness, and F_{IS}: coefficient of inbreeding (Weir & Cockerham, 1984). Asterisks indicate significant deviation from 0 after FDR correction (* = P<0.05, **=P<0.01); w indicates likely Wahlund effect whereas i indicates inbreeding (based on comparison with F_{IS2}, see SI4 for detail). A_r and pA_r are reported twice, once with rarefaction based on the smallest sample size at each time point (excluding sample sizes <5), and once using n=6 (12 genes). Data in italics derive from small sample sizes and should hence be treated with caution, na's indicate that a particular statistic could not be calculated for that sample size. Underlined years indicate that the time point includes the year stated ± 1 year in order to increase sample size where this was low. These analyses do not include samples from Ireland or Scotland due to lack of temporal coverage.

RBD region	Year	N	N loci	He	Ho	N_A	A_r	pA_r	A_r (n=6)	pA_r (n=6)	F_{IS}
Eastern	*1994*	*1*	*15*	*0.53 ± 0.13*	*0.53 ± 0.13*	*1.53 ± 0.52*	*na*	*na*	*1.53*	*0.67*	*Na*
Eastern	1999	17	15	0.71 ± 0.03	0.69 ± 0.03	5.27 ± 1.53	5.22	1.12	4.23	1.00	0.027
Eastern	2004	7	15	0.72 ± 0.05	0.60 ± 0.05	4.67 ± 1.18	4.67	0.52	4.51	0.53	0.178
Eastern	2009	28	15	0.73 ± 0.03	0.65 ± 0.02	6.07 ± 1.39	5.41	0.85	4.43	0.78	0.107*w
Eastern	2014	21	15	0.71 ± 0.04	0.65 ± 0.03	5.47 ± 0.83	5.13	0.66	4.24	0.53	0.084
Northern	*1994*	*2*	*15*	*0.66 ± 0.04*	*0.70 ± 0.08*	*2.47 ± 0.52*	*na*	*na*	*2.47*	*0.36*	*Na*
Northern	1999	16	15	0.70 ± 0.02	0.62 ± 0.03	4.87 ± 1.19	4.87	0.58	3.97	0.68	0.110*w
Northern	2004	11	15	0.71 ± 0.02	0.65 ± 0.04	4.80 ± 1.08	4.34	0.42	4.16	0.44	0.079
Northern	2009	16	15	0.72 ± 0.02	0.64 ± 0.03	5.73 ± 1.83	5.49	0.79	4.43	0.71	0.115*w
Northern	2014	14	15	0.70 ± 0.02	0.67 ± 0.03	5.00 ± 0.85	5.00	0.42	4.08	0.38	0.045
Severn	1994	6	15	0.54 ± 0.04	0.49 ± 0.05	3.00 ± 0.53	3.00	0.47	3.00	0.11	0.109
Severn	1999	31	15	0.54 ± 0.03	0.48 ± 0.02	4.53 ± 0.99	3.93	0.02	3.10	0.08	0.112*w
Severn	2004	20	15	0.58 ± 0.03	0.51 ± 0.03	4.53 ± 0.64	3.50	0.12	3.35	0.12	0.137*w
Severn	2009	13	15	0.54 ± 0.04	0.51 ± 0.04	3.53 ± 0.92	3.53	0.08	3.01	0.07	0.054
Severn	2014	14	15	0.57 ± 0.03	0.45 ± 0.03	4.00 ± 1.07	4.00	0.16	3.28	0.15	0.223**i
South-West	2004	23	15	0.55 ± 0.05	0.52 ± 0.03	4.00 ± 1.46	3.20	0.06	3.09	0.08	0.061
South-West	2009	30	15	0.56 ± 0.05	0.49 ± 0.02	4.73 ± 1.16	4.09	0.06	3.32	0.15	0.124*i
South-West	2014	24	15	0.60 ± 0.04	0.55 ± 0.03	5.13 ± 0.83	4.67	0.15	3.65	0.17	0.095*w
Western Wales	1994	19	15	0.48 ± 0.05	0.44 ± 0.03	3.27 ± 0.80	2.77	0.23	2.77	0.2	0.083
Western Wales	1999	34	15	0.54 ± 0.04	0.51 ± 0.02	4.00 ± 0.76	3.71	0.08	3.13	0.07	0.061
Western Wales	2004	15	15	0.52 ± 0.04	0.46 ± 0.03	3.60 ± 1.12	3.08	0.01	2.97	0.02	0.122
Western Wales	2009	17	15	0.56 ± 0.04	0.52 ± 0.03	3.93 ± 0.80	3.77	0.05	3.20	0.07	0.075
Western Wales	2014	17	15	0.55 ± 0.03	0.53 ± 0.03	3.73 ± 0.96	3.61	0.06	3.04	0.06	0.028

2.3.2 Population structure and geneflow

Exact tests of population differentiation showed that all STG's were significantly differentiated from each other apart from Severn-West Wales in 2009 (Table 2.4). The degree of differentiation between RBD regions varied spatially. The significant pairwise F_{ST} values ranged from 0.24 (South West-Western Wales, 2004) to 0.02 (Severn-Western Wales, 1999 & 2004). All pairwise comparisons showed high F_{ST} values between RBD regions, apart from Eastern-Northern and Severn-Western Wales comparisons. By 2014, pairwise F_{ST} estimates remained high between the majority of RBD regions, with significant differentiation remaining between these sub-populations. Global estimates of F_{ST} at each time point estimated from AMOVA slightly decreased over time where all 5 RBD regions were included.

Table 2.4: Pairwise F_{ST} estimates (Weir & Cockerham 1984) between RBD regions between 1999 and 2014. n/a: not enough data to estimate F_{ST} at that RBD-year combination. All values were significant ($p<0.05$) based on exact tests of population differentiation (Raymond and Rousset 1995), except where noted (n.s.). Bottom rows: global differentiation estimated from AMOVA, along with the number of RBD regions (# RBD regions) included in the analysis.

Pairwise RBD regions	1999	2004	2009	2014
Eastern-Northern	0.07	0.04	0.06	0.04
Eastern-Severn	0.18	0.11	0.15	0.12
Eastern-South West	na	0.09	0.13	0.11
Eastern-Western Wales	0.16	0.14	0.15	0.13
Northern-Severn	0.13	0.10	0.10	0.10
Northern-South West	na	0.17	0.13	0.11
Northern-Western Wales	0.13	0.12	0.11	0.11
Severn-South West	na	0.19	0.19	0.18
Severn-Western Wales	0.02	0.02	$0.01^{n.s.}$	0.03
South West-Western Wales	na	0.24	0.20	0.17
Global F_{ST}	0.11	0.14	0.13	0.12
# RBD regions	4	5	5	5

Mantel tests for signals of isolation by distance (IBD) at each time point were not significant, indicating that there is no consistent overall IBD pattern across the UK. However, when we tested for spatial autocorrelation in the data using mantel correlograms it was possible to observe both negative and positive autocorrelation in different distance classes. Across all time points the relationship between Mantel's correlation and distance class was similar (Figure 2.2 - A) with negative spatial autocorrelation over the shorter distance classes (<150 km) and positive spatial autocorrelation at the larger distance classes (>150 km). Additionally, both adult females (Figure 2.2 – B) and adult males (Figure 2.2 – C) exhibited significant negative spatial autocorrelation over the first three distance classes (< 150 km) at most time points suggesting that dispersal is not sex-biased but in fact undertaken by both sexes.

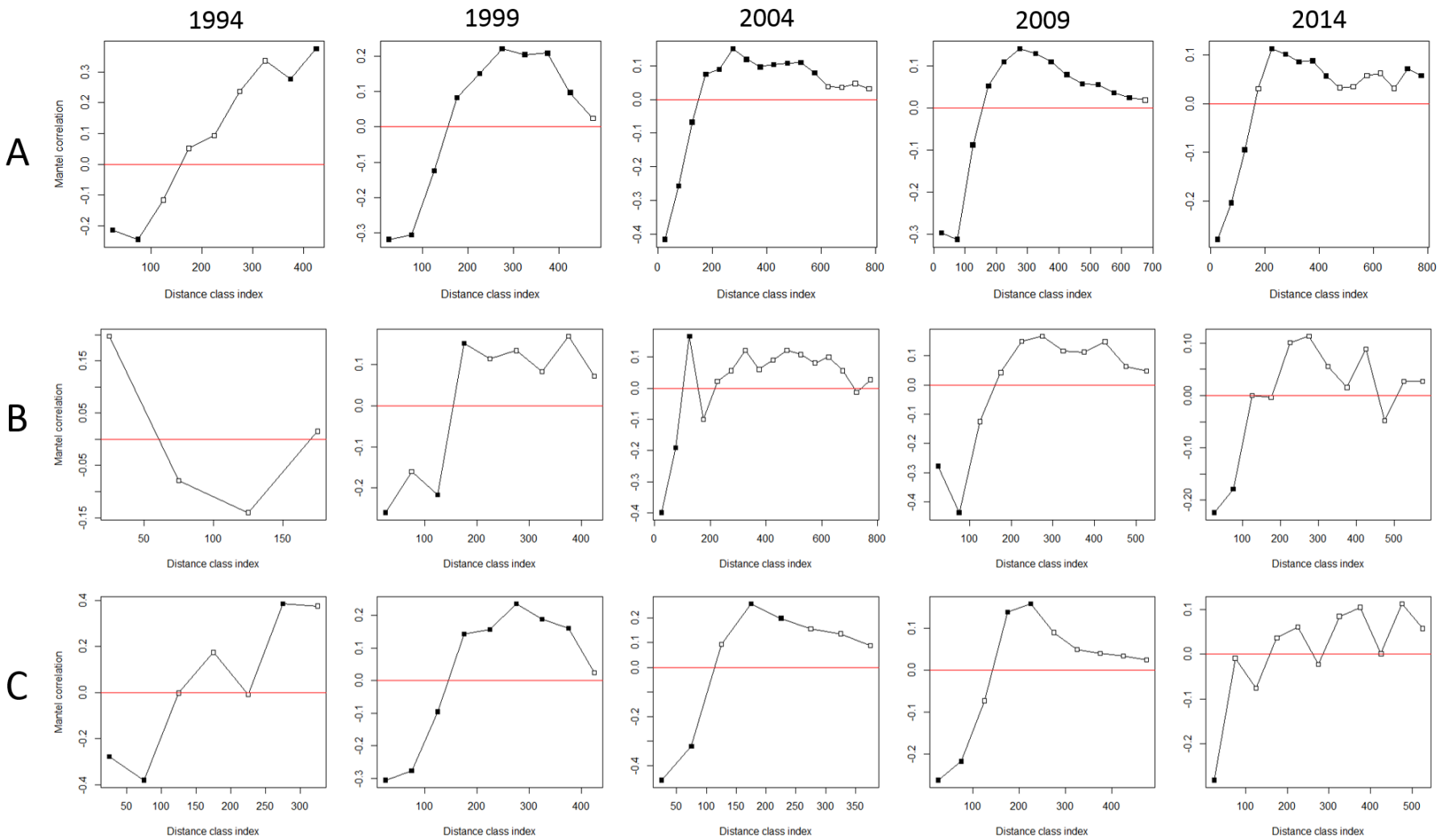


Figure 2.2: Mantel correlogram showing isolation by distance correlations across distance classes. A = All data. B = Adult females only. C = Adult males only. Black squares represent statistically significant spatial autocorrelation in that distance class, white squares represent non-significant results. Mantel correlation values >0 indicate positive spatial autocorrelation and mantel correlation values <0 show negative spatial autocorrelation.

Estimates of geneflow between RBD regions suggest an increase over time, whether using the non-directional private alleles method (Barton and Slatkin 1986 - Figure 2.3A) or the Bayesian approach in BayesAss (directional migration - Figure 2.3B). The number of migrants estimated by the private alleles method approximately doubled at each time point between 2004 and 2014, indicating steadily increased geneflow over time (0.70, 0.73, 1.61 and 2.95 in 1999, 2004, 2009 and 2014 respectively). From the directional migration estimates from BayesAss, it is also possible to observe that in earlier time periods there was asymmetric geneflow with higher effective migration rate from Wales and the South West into Eastern and Northern parts of England than in the opposite direction. By the later time points, there was a more even exchange across the Western and Eastern regions.

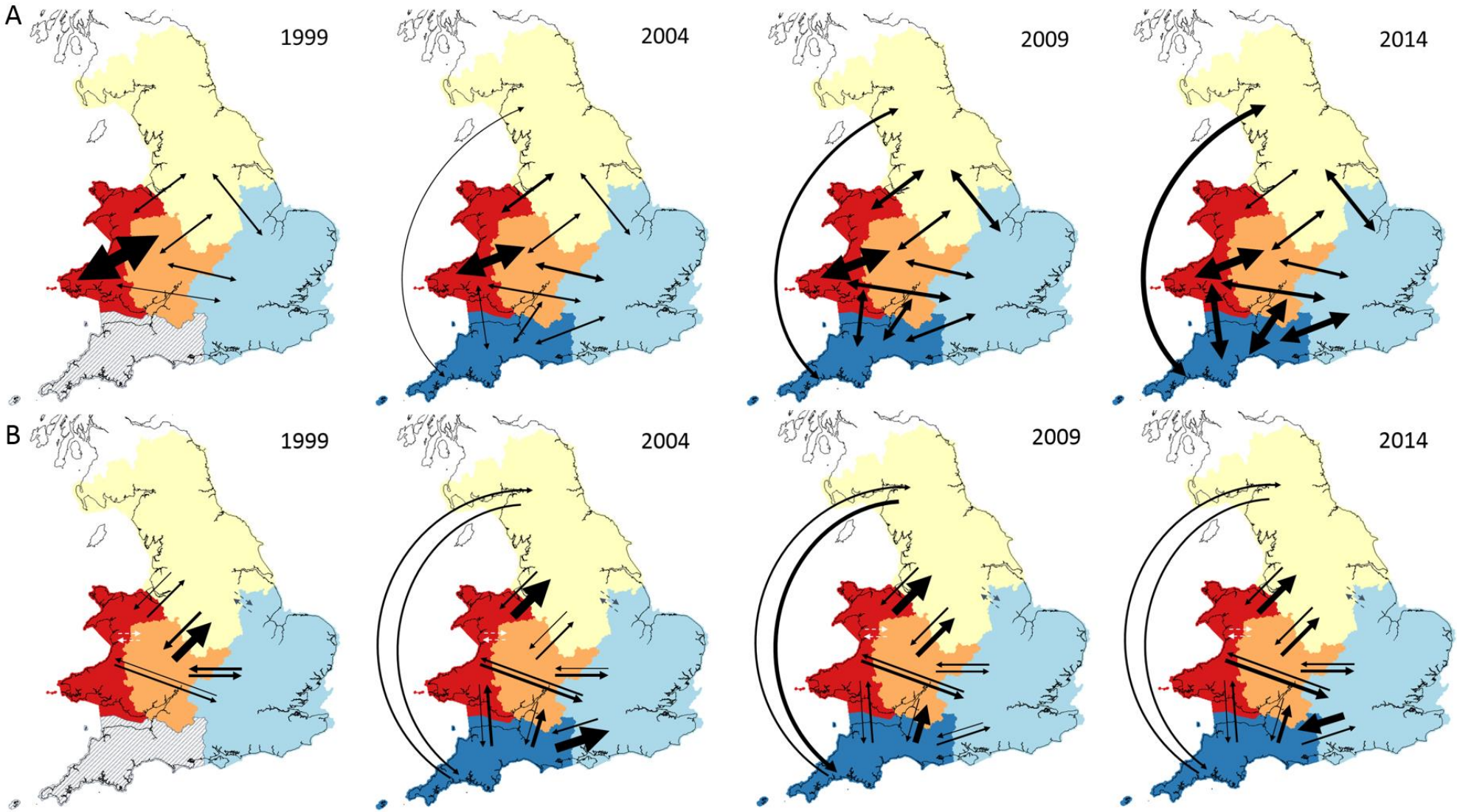


Figure 2.3: Geneflow between RBD regions at each of four time points (1999 - 2014). A. Geneflow estimates from private alleles method with arrows showing the migration between pairs of RBD regions and the arrow weight being directly proportional to the number of effectively migrating individuals. B. Directional geneflow estimates from BayesAss with arrows showing direction of effective migration and where arrow weight is directly proportional to percentage migration between populations. White and grey dashed arrows show regions where pairwise migration rates could not be reliably inferred due to low genetic differentiation. Regions are coloured as follows; Western Wales - red, Severn - orange, Northern - yellow, Eastern – light blue, South West – dark blue. Hatched lines indicate that a region was not included in the analysis at that time point.

Otters across the study area showed clear evidence of geographic substructuring. The value of K identified from the STRUCTURE runs using the Evanno ΔK method (Evanno et al. 2005) was K=2. However, this method is biased towards K=2 (Janes et al. 2017) and therefore other values of K were also explored (**Figure 2.4**) based on model likelihood, magnitude of ΔK and biological relevance (Cullingham et al. 2020). The three values of K shown here all exhibit strong geographic clustering, with Wales and the South West being the most genetically distinct areas. At the values of K shown, the percentage of individuals considered genetically admixed between clusters (Q<0.8) increased over time (i.e. suggesting increased geneflow between areas in the UK) from 4% in 1994 (at K=2, K=3 and K=5) to 16%, 17% and 23% in 2014 (at K=2, K=3 and K=5 respectively; for detail see Supplementary 2.5).

The number of genetic clusters in the data as inferred by Discriminant Analysis of Principal Components (DAPC) using the lowest BIC score as the model evaluation criterion was K=8, however, these clusters were very difficult to rationalise biologically or spatially (data not shown) and thus we explored other values of K, using the gradient of the decline between BIC values as a guide. Comparison of the clusters inferred by Discriminant Analysis of Principal Components (DAPC) with those inferred from STRUCTURE (**Figure 2.4**) show close agreement for K=2 and K=3, but at K=4 (not shown, due to low support) and K=5 (shown), DAPC and STRUCTURE inferred differing geographic clusters. This indicates that while parametric and non-parametric methods agree for the strongest genetic divisions within the overall population (here, up to K=3) assignments differ as genetic divisions weaken (i.e. at higher values of K).

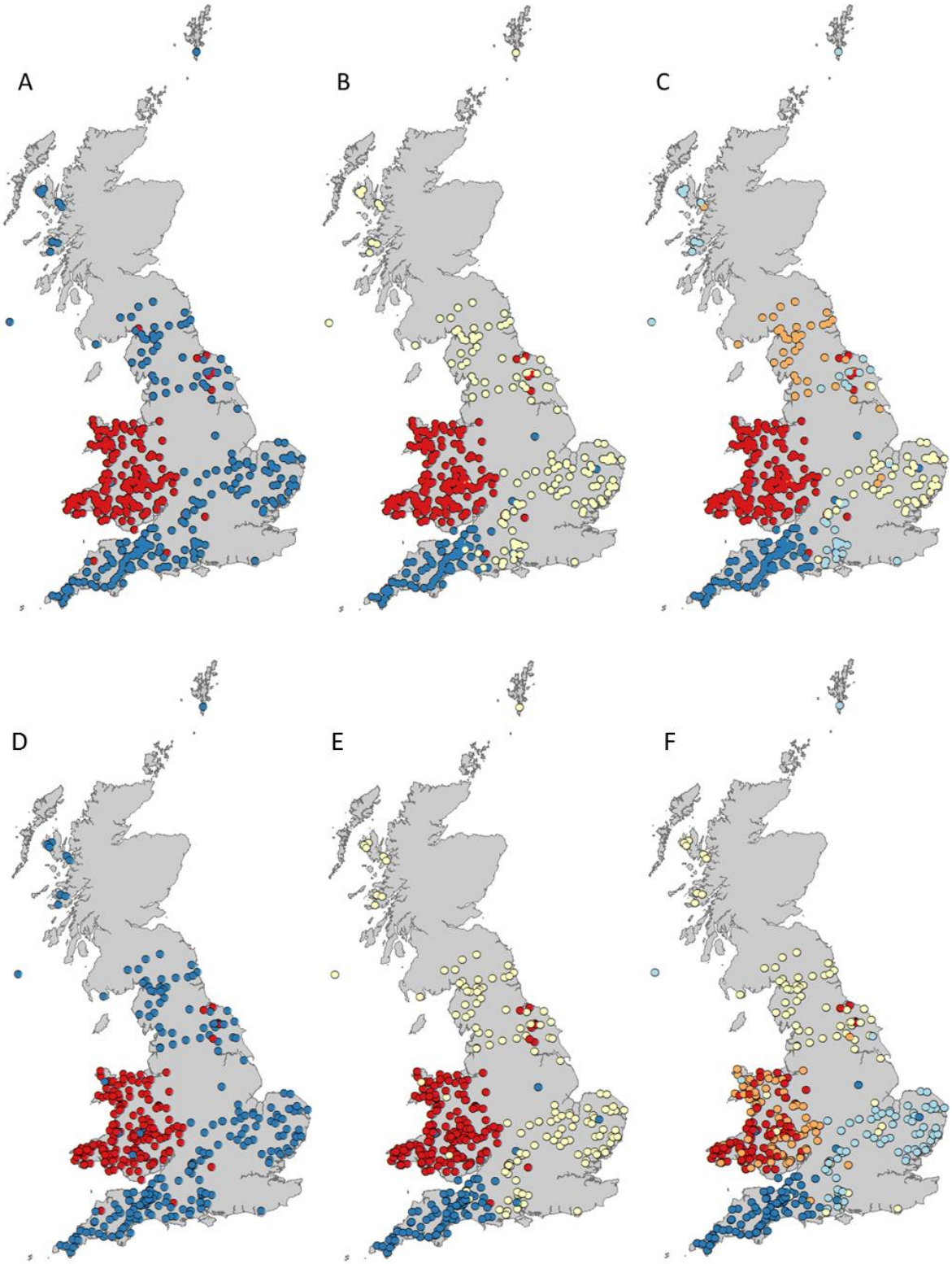


Figure 2.4: Genetic clusters identified in UK otters using a Bayesian approach in STRUCTURE (A-C) and the non-parametric approach Discriminant Analysis of Principal Components (D-F). Circles show the location of each otter in the dataset and the colours indicate which genetic cluster each belongs to. A & D. Show K = 2, B & E. show K = 3 and C & F. show K = 5

Progressive partitioning over selected time points where data for the whole study area was available (2004-2014) showed that the strength of regional differentiation differed with time (Supplementary 2.6). Despite this, the same four main sub-populations were identified in 2009 and 2014. In 2004

only three of the four main sub-populations were identified, however sparse sampling in most of England at this time may have resulted in a lack of power to realise the Eastern-Northern partition. The 2004 data also showed a significant partition between Cornwall and the rest of the South West (pairwise F_{ST} = 0.13, p<0.001), more pronounced than at any subsequent time point. The major partitions exhibited significant genetic differentiation, as measured by pairwise F_{ST} across all three time points (p<0.001 for all pairwise comparisons, for full data see Supplementary 2.6).

2.4 Discussion

Relatively few studies to date have explicitly quantified changes in genetic structure over time during contemporary population expansions, allowing empirical testing of theoretical predictions. Using the Eurasian otter as our case study, we predicted a weakening of genetic structure as anthropogenically fragmented populations reconnected as part of population recovery. However, we here show that the overarching population genetic structure showed surprisingly little change over time, with no strong increase in genetic mixing or genetic variability within regions despite increased geneflow. These results suggest that the UK otter population is less functionally connected than previously presumed, perhaps due to landscape or other barriers impeding effective migration. Understanding such limitations in this and other recovering populations is important with respect to managing and quantifying conservation successes.

2.4.1 Genetic diversity

The genetic diversity estimates in this study showed an increase over time in some metrics (i.e. allelic richness), especially in regions with very low initial diversity, but no trend in others (i.e. observed heterozygosity). While genetic diversity estimates did not show clear temporal trends, spatial differences of higher genetic variability in Eastern and Northern than in Western regions (Severn, South West and Western Wales) remained detectable across the 20-year study period. Before the timescale of this study both the Eastern and Northern regions underwent otter reinforcement projects (1983-1999, and 1990-1993 respectively), during which otters from other regions of the UK (and potentially of non-UK origin – Hájková et al. 2007) were released to bolster populations (Wayre 1985; Jefferies, Wayre and Shuter 2000; White et al. 2003; Hobbs et al. 2011; Bonesi et al. 2013). The arrival of new alleles in Eastern and Northern regions is likely to have contributed to their higher genetic diversity (evidenced across all metrics), while their sparsely populated landscapes prior to reinforcement would also have allowed more natural immigration of neighbouring otter populations. Currently this higher level of genetic diversity has not spread even

to the adjacent RBD region (Severn) indicating a persistence of (similar levels) of genetic differentiation in the population despite the putative re-establishment of demographic connectivity implied by recent national otter surveys.

Differences in the number of loci deviating from Hardy-Weinberg Equilibrium (HWE) and in linkage disequilibrium (LD) when calculated for the whole dataset versus by RBD region are likely due to the Wahlund effect, i.e. an excess of homozygotes is observed due to sampling from a structured population (treating the whole dataset as one when subdivisions exist) even though locally populations are randomly mating. An important consequence of this Wahlund effect is allelic associations between loci across the total population which can lead to signals of LD (Garnier-Gere and Chikhi 2003). If the signal reflected physical linkage between pairs of loci then it would be expected to occur across all regions, whereas the patterns observed suggest other processes are occurring which emulate true physical linkage. One alternative explanation could be genetic admixture between regions. Admixture linkage disequilibrium (ALD) has been described in unlinked loci due to differing allele frequencies in parental populations when there is geneflow between genetically distinct demes (Pfaff et al. 2001; Rybicki et al. 2002). Our two stage F_{IS} estimate added further clarity to this via establishing that many RBD regions are exhibiting Wahlund effects at specific time points. The F_{IS} values estimated for this dataset are higher than those previously estimated by Hobbs et al. (2011), suggesting either more sub-population level inbreeding than previously thought, or pronounced and sustained population structuring of UK otters (i.e. Wahlund effects). The latter is consistent with the relative lack of admixture of the sub-populations, which may be an indication that the UK otter population is demographically reconnected but still in an early genetic recovery phase, with individuals of differing genetic descent moving into RBD regions but admixture of the genotypes yet to occur. Being able to disentangle inbreeding from Wahlund effect is important in a conservation context and emphasizes the importance of understanding the population structure present as well as the distribution of a species relative to the sampling area. Without this understanding, accidental sampling of individuals from genetically distinct demes as 'one population' even at low levels can have wide ranging effects on HWE, LD and F_{IS} estimates (Waples 2015). Misinterpretation of significant results for any of the above could result in false assumption of inbreeding or natural selection within a population and may result in unnecessary interventions or designation of management units.

Several other studies across Europe have used similar methods to estimate genetic diversity and population structure of otters at various spatial scales. In our study, across all samples, the average number of alleles (N_A) was high compared with other European studies, whereas observed heterozygosity was relatively low (Lanszki et al. 2008; Mucci et al. 2010). This difference could be

due to the significant sub-structuring and differentiation between regions in the UK, with the past population bottleneck and subsequent genetic drift leading to different suites of alleles being present in the different regions, and resulting in a high N_A when looking at all the data as a whole. The only other study in Europe with a N_A higher than ours is Arrendal et al. (2004) which similarly sampled across the whole of Sweden (and part of Norway), including populations that were highly genetically distinct from one another and that had received introductions. When split by spatial-temporal grouping (STG) the N_A's were found to be within the range of values for countries across Europe (Mucci et al. 2010) - but with eastern regions being more diverse than western ones.

2.4.2 Population structure and geneflow

Both measures of population structure (STRUCTURE and DAPC) showed that otters in the UK have maintained a highly genetically structured population despite increased connectivity between sub-populations through range expansion. At higher values of K the two methods deviated slightly in their spatial-clustering pattern, this could be due to the assumptions made of Hardy-Weinberg equilibrium and/or linkage disequilibrium in the Bayesian clustering model implemented in STRUCTURE which do not exist in the DAPC model (Jombart et al. 2009). Under both methods the genetic clusters inferred reflect known otter stronghold populations and agreed with previous studies on genetic structure of the Eurasian otter in the UK (Hobbs et al. 2011; Stanton et al. 2014). This maintenance of population structure is also reflected in the predominantly high and consistent pairwise F_{ST} estimates between most regions across all time points, in contrast to global F_{ST} which decreased over time. This could be due to difference in the way that the two estimates are calculated (with pairwise F_{ST} using allele frequencies and AMOVA taking into account mutational distances between alleles; Excoffier et al. 1992), and also the larger sample size used in a global AMOVA. However, the global F_{ST} values only decrease marginally and a global F_{ST} of 0.12 in 2014 still indicates a substantial level of inter-region differentiation. All methods of estimating gene flow between regions showed increased levels of effective migration over time, however further generations of exchange of individuals may be needed to effect real change to population structure.

There are very few empirical studies that explicitly quantify changes in genetic structure over time during contemporary population expansions and therefore the body of work with which to compare our results is limited. However, the consistency of overarching genetic structure seen in the UK otter population is somewhat unexpected given the rapid disintegration of genetic structuring found in brown bears (*Ursus arctos*) in Finland by Hagen et al. (2015) over a similar time period. In their study, brown bears in Finland exhibited a rapid loss of population structure between 1996-2010 (16 years), as shown by the significantly decreasing pairwise F_{ST} values over time between the identified

northern and southern genetic clusters. These changes to the genetic structure of the population were estimated to have occurred over only 1.5 generations; in our study we observed much less dramatic changes to population structure over a similar sampling time period and a smaller study area (Finland is 338, 424 km^2; Great Britain is 209,31 km^2). Commonly used generation times in brown bears and otters are 11 years (Nilsson 2013) and 3 years (Randi et al. 2003) respectively. Even using the more conservative standardised generation length estimates developed by Pacifici et al. (2013) the average generation length of the Eurasian otter is 7.6 years, whereas for the brown bear it is 16.4 years, indicating that despite the study lengths covering a similar time frame, in otters this should equate to more generation times and therefore more opportunities for genetic mixing. In another study of brown bears, Schregel et al. (2017) found that high genetic differentiation remained despite a continuum of individuals across their study area (Sweden and Norway, between 2006 and 2013). As in our study of otters, this illustrates that demographic connectivity is not always a reliable indicator of genetic connectivity.

There are many reasons that could contribute to the lack of population structure dissolution in the UK otter population. Firstly, the landscape in the UK has changed significantly over the 40 years that otters were absent from large areas; increased urbanisation, human population size, number of roads, volume of traffic and alterations to rivers could all affect the realised connectivity between sub-populations. However, there is also a lag time for the genetic signatures left by old barriers to disappear and those created by new barriers to become detectable (Landguth et al. 2010). Number of generations since barrier removal or creation needs to be taken into account as well as the dispersal ability of the species when interpreting barriers to geneflow as both of these factors have been shown to affect the rate at which such genetic signatures decay or evolve respectively (Landguth et al. 2010). It is assumed that otters can disperse over relatively large distances, as several tracking studies have recorded individuals moving tens of kilometres in one night (Jenkins 1980; Green et al. 1984; Quaglietta et al. 2013) therefore this should be less of a problem than with other less mobile species. Isolation in fragmented populations may have also affected other aspects of otter ecology through genetic drift – for example, Kean et al. (2017) found regionally differentiated scent odour profiles in the UK which reflected genetic structure within the population. The discovery of these dialects is important given that otters communicate predominantly by scent material left in their spraint (Trowbridge 1983). Scent differences apparently signalling age, reproductive status, sex and even individual identity have previously been described (Kean, Müller and Chadwick 2011). These differing regional scent 'dialects' could be restricting gene flow, if otters are preferentially mating with otters of similar scent.

Despite changes in otter population size, the results from our spatial autocorrelation analysis showed similar patterns at each time point (negative spatial autocorrelation in distance classes at small spatial scales, suggesting local dispersal). Given the territorial nature of otters (Erlinge 1968) it would be expected that as a local population grows and reaches carrying capacity, individuals would need to move further to establish their own territory (Sjoasen 1997), leading to a gradual increase in the distance over which we observe negative spatial autocorrelation. Instead, our results show no change with time, suggesting that even though populations are assumed to be approaching carrying capacity in some areas, dispersal distance remains largely unchanged. A lack of density-dependent dispersal in otters might explain this result, but seems unlikely given that this phenomenon has been observed in over 70% of the mammal species studied (review: Matthysen 2005) and was also proposed as a mechanism for the short dispersal distances detected by Quaglietta et al. (2013) in otters in Portugal. An alternative explanation could be that improvements in river quality have increased local carrying capacities through recovery of prey populations (e.g. brown trout, Monteith et al. 2005). At higher prey densities, otter range size may decrease (Sidorovich 1991; Néill et al. 2009), as has been observed for some other carnivore species such as the Eurasian lynx (Herfindal et al. 2005), countering increased dispersal distances associated with expanding populations. Our analysis of adult females and adult males suggests that both sexes disperse as they exhibit similar patterns of negative spatial autocorrelation in the smaller distance classes (<150 km). This is contrary to previous evidence from radio tracking which observed male-biased dispersal in Eurasian otters (Quaglietta et al. 2013) but is similar to the limit of geneflow detected across both sexes in Scotland by Dallas et al. (2002).

During the last 30 years the otter population in Great Britain has expanded from small stronghold populations and near extinction in some areas (i.e. East Anglia) to an almost continuous distribution (Crawford 2010; Strachan 2015). This large-scale population recovery has been proclaimed as a success for policy and practise, where changes in pollution control and the improvement of river and riparian habitat (Crawford 2010) have supported a largely natural population expansion. The current study shows that this spatial connectivity has not translated into genetic connectivity in the manner or speed expected, with the population having retained the strong spatial genetic structuring observed earlier on in the recovery process (Hobbs et al. 2011; Stanton et al. 2014). The appearance of a spatially continuous population may therefore give a false sense of security with respect to genetic robustness, since in fact the population remains vulnerable, comprised of genetically fragmented sub-populations (Reed 2004). Given the overall increasing levels of gene flow seen in

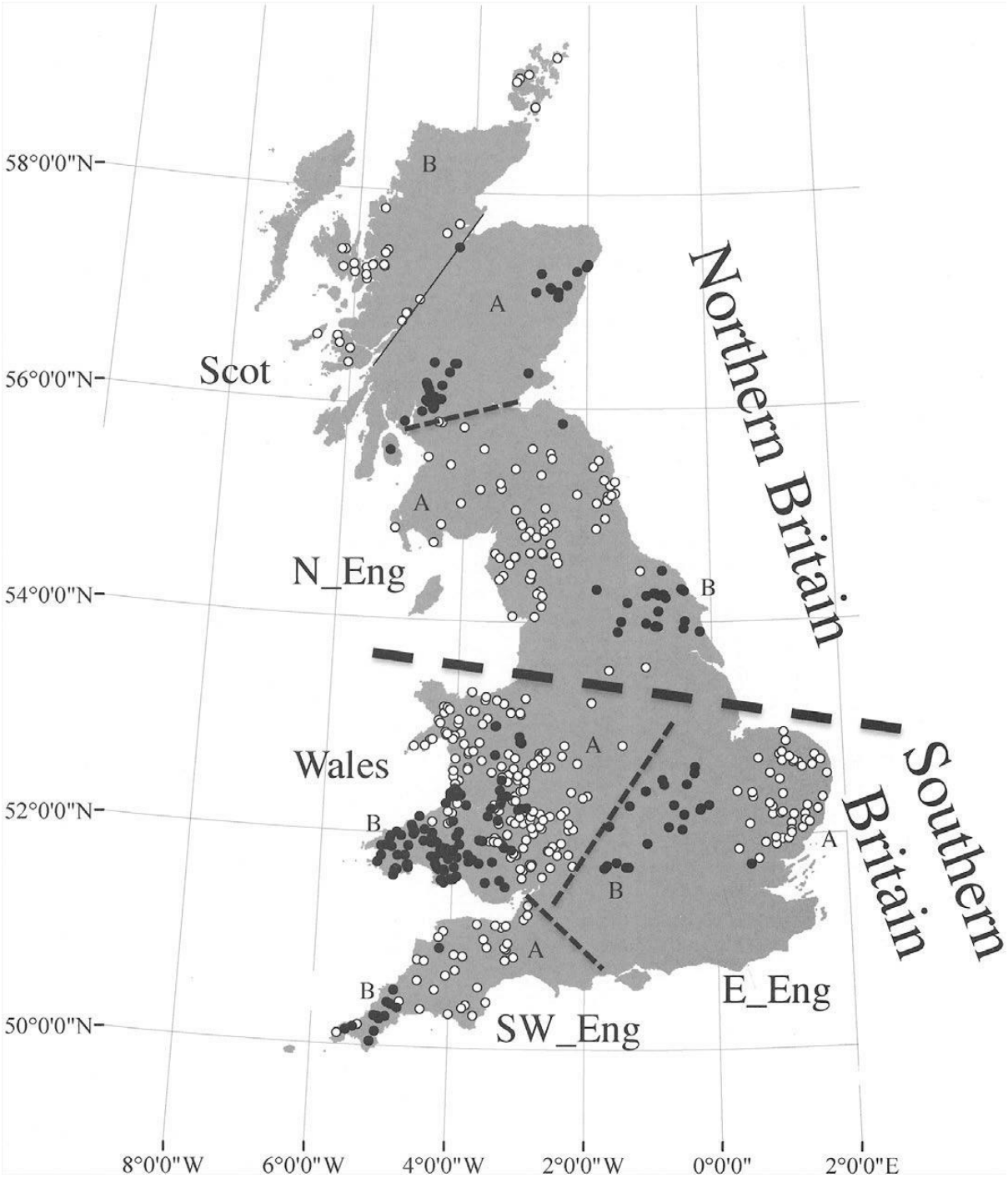
this study, it may be that more time is all that is needed for the achieved spatial connectivity of the otter population to translate into genetic mixing. Future analysis of the data using landscape genetic techniques (Manel et al. 2003) or estimated effective migration surfaces (Petkova et al. 2016) may help identify any barriers to geneflow, whether extrinsic (e.g. landscape variables) or intrinsic (e.g. differences in scent profiles; Kean et al. 2017).

2.5 Conclusions

Spatial-temporal studies of how genetic diversity and population structure changes during contemporary population expansions are still rare, perhaps due to the difficulties and costs of sampling over such large ranges, both spatial and temporally. However, such empirical studies are urgently needed to make predictions on population recovery progress, and to set realistic goals for management and monitoring activities. Our findings illustrate that achieved spatial recovery of formerly endangered species may not necessarily imply that genetic recovery has occurred as well. Demographic, behavioural and spatial barriers to genetic mixing of sub-populations can lead to a time lag, where genetic recovery in terms of variability and connectivity requires much more time than a recovery in re-occupied distribution range. Our study therefore highlights that population recovery after past bottlenecks can require much longer than apparent from spatial data alone, reinforcing the value of population genetic monitoring programs for endangered species.

Supplementary Information

S2.1 Map of mainland Britain showing *Lutra lutra* subpopulations (black and while circles, with each subpopulation referenced as either A or B) and groups (marked by thin dashed lines) defined by GENELAND Bayesian clustering and used as discrete populations for analysis. Group labels refer to the following locations: Scotland (Scot), northern England (N_Eng), Wales (Wales), southwestern England (SW Eng), and eastern England (E_Eng). Approximate location of the Great Glen is shown by the solid line, between subpopulations A and B in Scotland. (From Stanton *et al.* 2014)



S2.2 Data allocation across the three studies

Ind ID	Year	Study	RBD Region
7	1993	Hobbs et al 2011	Severn
34	1993	Hobbs et al 2011	W Wales
47	1993	Hobbs et al 2011	W Wales
48	1993	Hobbs et al 2011	W Wales
8	1994	Hobbs et al 2011	Severn
11	1994	Hobbs et al 2011	W Wales
21	1994	Hobbs et al 2011	W Wales
25	1994	Hobbs et al 2011	W Wales
26	1994	Hobbs et al 2011	W Wales
31	1994	Hobbs et al 2011	Severn
37	1994	Hobbs et al 2011	W Wales
38	1994	Hobbs et al 2011	W Wales
42	1994	Hobbs et al 2011	W Wales
244	1994	Hobbs et al 2011	W Wales
24	1995	Hobbs et al 2011	W Wales
27	1995	Hobbs et al 2011	Severn
29	1995	Hobbs et al 2011	Northern
32	1995	Hobbs et al 2011	W Wales
35	1995	Hobbs et al 2011	W Wales
36	1995	Hobbs et al 2011	W Wales
49	1995	Hobbs et al 2011	W Wales
50	1995	Hobbs et al 2011	W Wales
51	1995	Hobbs et al 2011	W Wales
53	1995	Hobbs et al 2011	Eastern
56	1995	Hobbs et al 2011	Severn
60	1995	Hobbs et al 2011	W Wales
246	1995	Hobbs et al 2011	Northern
259	1995	Hobbs et al 2011	Severn
143	1998	Hobbs et al 2011	W Wales
147	1998	Hobbs et al 2011	W Wales
148	1998	Hobbs et al 2011	W Wales
149	1998	Hobbs et al 2011	W Wales
151	1998	Hobbs et al 2011	W Wales
152	1998	Hobbs et al 2011	Severn
155	1998	Hobbs et al 2011	W Wales
163	1998	Hobbs et al 2011	W Wales
166	1998	Hobbs et al 2011	W Wales
169	1998	Hobbs et al 2011	Severn
174	1998	Hobbs et al 2011	Eastern
176	1998	Hobbs et al 2011	Eastern
177	1998	Hobbs et al 2011	Northern
179	1998	Hobbs et al 2011	Northern
184	1998	Hobbs et al 2011	Eastern
185	1998	Hobbs et al 2011	W Wales
187	1998	Hobbs et al 2011	Severn
188	1998	Hobbs et al 2011	Severn
190	1998	Hobbs et al 2011	W Wales
191	1998	Hobbs et al 2011	W Wales
192	1998	Hobbs et al 2011	Severn
193	1998	Hobbs et al 2011	W Wales
194	1998	Hobbs et al 2011	W Wales
195	1998	Hobbs et al 2011	W Wales
214	1998	Hobbs et al 2011	W Wales
216	1998	Hobbs et al 2011	W Wales
217	1998	Hobbs et al 2011	W Wales
262	1998	Hobbs et al 2011	Northern
264	1998	Hobbs et al 2011	Northern
274	1998	Hobbs et al 2011	W Wales
278	1998	Hobbs et al 2011	Northern
281	1998	Hobbs et al 2011	Northern
201	1999	Hobbs et al 2011	Eastern
206	1999	Hobbs et al 2011	W Wales
208	1999	Hobbs et al 2011	Severn
210	1999	Hobbs et al 2011	W Wales
211	1999	Hobbs et al 2011	W Wales
215	1999	Hobbs et al 2011	W Wales
218	1999	Hobbs et al 2011	W Wales
225	1999	Hobbs et al 2011	Severn
226	1999	Hobbs et al 2011	Severn
227	1999	Hobbs et al 2011	Northern
230	1999	Hobbs et al 2011	Eastern
231	1999	Hobbs et al 2011	Eastern
232	1999	Hobbs et al 2011	W Wales
235	1999	Hobbs et al 2011	Northern
238	1999	Hobbs et al 2011	Northern
241	1999	Hobbs et al 2011	Severn
242	1999	Hobbs et al 2011	W Wales
243	1999	Hobbs et al 2011	W Wales
248	1999	Hobbs et al 2011	Severn
255	1999	Hobbs et al 2011	Eastern
256	1999	Hobbs et al 2011	W Wales
267	1999	Hobbs et al 2011	Other
268	1999	Hobbs et al 2011	Northern
386	1999	Hobbs et al 2011	Eastern
254	2000	Hobbs et al 2011	Severn
257	2000	Hobbs et al 2011	Eastern
258	2000	Hobbs et al 2011	Severn
260	2000	Hobbs et al 2011	Severn
285	2000	Hobbs et al 2011	Northern
286	2000	Hobbs et al 2011	Northern
289	2000	Hobbs et al 2011	Severn
290	2000	Hobbs et al 2011	Severn
295	2000	Hobbs et al 2011	Severn
299	2000	Hobbs et al 2011	Northern
301	2000	Hobbs et al 2011	Severn
304	2000	Hobbs et al 2011	Severn
306	2000	Hobbs et al 2011	Northern
307	2000	Hobbs et al 2011	W Wales
314	2000	Hobbs et al 2011	Severn
315	2000	Hobbs et al 2011	Severn
317	2000	Hobbs et al 2011	Severn
319	2000	Hobbs et al 2011	W Wales
320	2000	Hobbs et al 2011	Severn
321	2000	Hobbs et al 2011	Northern
323	2000	Hobbs et al 2011	Eastern
324	2000	Hobbs et al 2011	W Wales
325	2000	Hobbs et al 2011	Eastern
327	2000	Hobbs et al 2011	Severn
328	2000	Hobbs et al 2011	Eastern
331	2000	Hobbs et al 2011	Eastern
333	2000	Hobbs et al 2011	W Wales
335	2000	Hobbs et al 2011	Severn
342	2000	Hobbs et al 2011	Severn
346	2000	Hobbs et al 2011	Eastern
350	2000	Hobbs et al 2011	Severn
351	2000	Hobbs et al 2011	Severn
354	2000	Hobbs et al 2011	Severn
356	2000	Hobbs et al 2011	W Wales
358	2000	Hobbs et al 2011	W Wales
361	2000	Hobbs et al 2011	Eastern
362	2000	Hobbs et al 2011	Severn
366	2000	Hobbs et al 2011	W Wales
367	2000	Hobbs et al 2011	Eastern
376	2000	Hobbs et al 2011	Northern
387	2000	Hobbs et al 2011	Eastern
396	2000	Hobbs et al 2011	Other
411	2000	Hobbs et al 2011	Severn
412	2000	Hobbs et al 2011	Severn

Ind ID	Year	Study	RBD Region
625	2004	Hobbs et al 2011	Severn
626	2004	Hobbs et al 2011	Eastern
627	2004	Hobbs et al 2011	Severn
628	2004	Hobbs et al 2011	Northern
630	2004	Hobbs et al 2011	W Wales
632	2004	Hobbs et al 2011	W Wales
634	2004	Hobbs et al 2011	W Wales
635	2004	Hobbs et al 2011	Severn
636	2004	Hobbs et al 2011	W Wales
641	2004	Hobbs et al 2011	Severn
642	2004	Hobbs et al 2011	Severn
650	2004	Hobbs et al 2011	W Wales
655	2004	Hobbs et al 2011	Northern
656	2004	Hobbs et al 2011	Severn
660	2004	Hobbs et al 2011	Severn
661	2004	Hobbs et al 2011	Severn
666	2004	Hobbs et al 2011	W Wales
667	2004	Hobbs et al 2011	Severn
669	2004	Hobbs et al 2011	Severn
670	2004	Hobbs et al 2011	W Wales
671	2004	Hobbs et al 2011	W Wales
672	2004	Hobbs et al 2011	W Wales
673	2004	Hobbs et al 2011	Severn
676	2004	Hobbs et al 2011	Severn
679	2004	Hobbs et al 2011	Severn
680	2004	Hobbs et al 2011	Severn
683	2004	Hobbs et al 2011	Northern
684	2004	Hobbs et al 2011	Severn
685	2004	Hobbs et al 2011	Severn
686	2004	Hobbs et al 2011	Severn
687	2004	Hobbs et al 2011	Northern
688	2004	Hobbs et al 2011	Northern
689	2004	Hobbs et al 2011	Northern

Ind ID	Year	Study	RBD Region
691	2004	Hobbs et al 2011	W Wales
702	2004	Hobbs et al 2011	Severn
706	2004	Hobbs et al 2011	W Wales
708	2004	Hobbs et al 2011	W Wales
710	2004	Hobbs et al 2011	W Wales
744	2004	Hobbs et al 2011	Northern
745	2004	Hobbs et al 2011	Northern
748	2004	Hobbs et al 2011	W Wales
787	2004	Hobbs et al 2011	Northern
856	2004	Hobbs et al 2011	Northern
957	2004	Hobbs et al 2011	W Wales
99634	2004	Hobbs et al 2011	South West
99647	2004	Hobbs et al 2011	South West
99650	2004	Hobbs et al 2011	Eastern
99654	2004	Hobbs et al 2011	South West
99658	2004	Hobbs et al 2011	South West
99659	2004	Hobbs et al 2011	South West
99667	2004	Hobbs et al 2011	South West
99685	2004	Hobbs et al 2011	Eastern
99686	2004	Hobbs et al 2011	South West
99687	2004	Hobbs et al 2011	South West
99692	2004	Hobbs et al 2011	Other
99695	2004	Hobbs et al 2011	South West
99701	2004	Hobbs et al 2011	South West
99714	2004	Hobbs et al 2011	South West
99722	2004	Hobbs et al 2011	Eastern
99732	2004	Hobbs et al 2011	South West
99739	2004	Hobbs et al 2011	South West
99745	2004	Hobbs et al 2011	South West
99766	2004	Hobbs et al 2011	South West
99778	2004	Hobbs et al 2011	South West
99795	2004	Hobbs et al 2011	South West
99796	2004	Hobbs et al 2011	South West

Ind ID	Year	Study	RBD Region
99800	2004	Hobbs et al 2011	South West
99811	2004	Hobbs et al 2011	Northern
99813	2004	Hobbs et al 2011	Eastern
99815	2004	Hobbs et al 2011	Eastern
99832	2004	Hobbs et al 2011	South West
99840	2004	Hobbs et al 2011	Other
99846	2004	Hobbs et al 2011	South West
99858	2004	Hobbs et al 2011	Other
99905	2004	Hobbs et al 2011	Severn
99911	2004	Hobbs et al 2011	South West
99945	2004	Hobbs et al 2011	South West
991037	2004	Hobbs et al 2011	Severn
991064	2004	Hobbs et al 2011	Eastern
1264	2009	Current Study	Eastern
1270	2009	Current Study	Northern
1271	2009	Current Study	Eastern
1272	2009	Current Study	Eastern
1274	2009	Current Study	Eastern
1276	2009	Stanton et al 2014	Northern
1281	2009	Current Study	W Wales
1282	2009	Current Study	Eastern
1283	2009	Stanton et al 2014	Northern
1293	2009	Current Study	South West
1296	2009	Current Study	South West
1302	2009	Stanton et al 2014	Severn
1305	2009	Current Study	South West
1306	2009	Current Study	South West
1309	2009	Current Study	South West
1310	2009	Current Study	South West
1313	2009	Current Study	South West
1315	2009	Current Study	South West
1317	2009	Current Study	South West
1318	2009	Current Study	Eastern

Ind ID	Year	Study	RBD Region
1319	2009	Stanton et al 2014	Northern
1320	2009	Stanton et al 2014	Northern
1321	2009	Current Study	Severn
1323	2009	Current Study	Severn
1329	2009	Current Study	Severn
1330	2009	Current Study	Eastern
1331	2009	Current Study	Eastern
1332	2009	Current Study	Eastern
1333	2009	Current Study	Eastern
1336	2009	Current Study	Severn
1346	2009	Current Study	Eastern
1353	2009	Current Study	Eastern
1354	2009	Current Study	South West
1355	2009	Current Study	South West
1356	2009	Current Study	South West
1357	2009	Current Study	South West
1358	2009	Current Study	South West
1362	2009	Current Study	Eastern
1363	2009	Current Study	Northern
1364	2009	Current Study	Eastern
1366	2009	Current Study	Northern
1367	2009	Current Study	South West
1374	2009	Current Study	Eastern
1376	2009	Current Study	Eastern
1377	2009	Current Study	Eastern
1378	2009	Current Study	Eastern
1384	2009	Current Study	South West
1386	2009	Stanton et al 2014	Northern
1393	2009	Current Study	Eastern
1398	2009	Stanton et al 2014	W Wales
1399	2009	Current Study	South West
1402	2009	Current Study	South West
1403	2009	Current Study	South West

Ind ID	Year	Study	RBD Region
1405	2009	Current Study	Northern
1412	2009	Current Study	South West
1418	2009	Current Study	Northern
1420	2009	Current Study	South West
1423	2009	Current Study	Eastern
1424	2009	Current Study	Eastern
1426	2009	Current Study	Eastern
1427	2009	Current Study	Eastern
1430	2009	Current Study	Severn
1437	2009	Current Study	W Wales
1440	2009	Current Study	Severn
1441	2009	Current Study	Eastern
1442	2009	Current Study	South West
1444	2009	Current Study	W Wales
1446	2009	Current Study	W Wales
1448	2009	Current Study	W Wales
1450	2009	Current Study	W Wales
1451	2009	Current Study	W Wales
1452	2009	Current Study	W Wales
1456	2009	Current Study	W Wales
1457	2009	Current Study	W Wales
1458	2009	Current Study	W Wales
1460	2009	Current Study	W Wales
1464	2009	Current Study	Eastern
1465	2009	Current Study	South West
1466	2009	Current Study	South West
1470	2009	Current Study	South West
1472	2009	Current Study	South West
1473	2009	Current Study	Northern
1474	2009	Current Study	Northern
1475	2009	Current Study	South West
1487	2009	Current Study	Eastern
1490	2009	Current Study	South West

Ind ID	Year	Study	RBD Region
1498	2009	Current Study	Severn
1500	2009	Current Study	Severn
1523	2009	Current Study	Severn
1531	2009	Current Study	Northern
1534	2009	Current Study	Northern
1535	2009	Current Study	Northern
1541	2009	Current Study	W Wales
1554	2009	Current Study	South West
1563	2009	Current Study	Eastern
1564	2009	Current Study	Eastern
1567	2009	Current Study	W Wales
1568	2009	Current Study	W Wales
1669	2009	Current Study	South West
1700	2009	Current Study	Northern
1735	2009	Current Study	W Wales
1873	2009	Current Study	Severn
1991	2009	Current Study	Severn
2066	2009	Current Study	Severn
2277	2014	Current Study	W Wales
2278	2014	Current Study	Severn
2292	2014	Current Study	Eastern
2294	2014	Current Study	Severn
2310	2014	Current Study	Northern
2311	2014	Current Study	Severn
2314	2014	Current Study	Northern
2319	2014	Current Study	Eastern
2321	2014	Current Study	W Wales
2323	2014	Current Study	W Wales
2325	2014	Current Study	Eastern
2330	2014	Current Study	South West
2331	2014	Current Study	Eastern
2332	2014	Current Study	Eastern
2336	2014	Current Study	South West

Ind ID	Year	Study	RBD Region
2340	2014	Current Study	South West
2341	2014	Current Study	South West
2342	2014	Current Study	South West
2344	2014	Current Study	Severn
2345	2014	Current Study	Severn
2347	2014	Current Study	Severn
2348	2014	Current Study	Severn
2351	2014	Current Study	Eastern
2355	2014	Current Study	South West
2356	2014	Current Study	South West
2358	2014	Current Study	South West
2362	2014	Current Study	Northern
2365	2014	Current Study	W Wales
2370	2014	Current Study	Severn
2377	2014	Current Study	South West
2380	2014	Current Study	Eastern
2383	2014	Current Study	Eastern
2386	2014	Current Study	W Wales
2388	2014	Current Study	W Wales
2391	2014	Current Study	W Wales
2398	2014	Current Study	South West
2401	2014	Current Study	South West
2406	2014	Current Study	Northern
2407	2014	Current Study	Northern
2409	2014	Current Study	Northern
2411	2014	Current Study	South West
2416	2014	Current Study	Northern
2420	2014	Current Study	South West
2422	2014	Current Study	Other
2424	2014	Current Study	Other
2425	2014	Current Study	Other
2434	2014	Current Study	Eastern
2435	2014	Current Study	Eastern

Ind ID	Year	Study	RBD Region
2438	2014	Current Study	Northern
2440	2014	Current Study	Eastern
2443	2014	Current Study	Eastern
2449	2014	Current Study	Severn
2453	2014	Current Study	Severn
2455	2014	Current Study	W Wales
2459	2014	Current Study	Northern
2460	2014	Current Study	Severn
2462	2014	Current Study	Eastern
2463	2014	Current Study	Eastern
2465	2014	Current Study	Eastern
2466	2014	Current Study	Northern
2467	2014	Current Study	Eastern
2468	2014	Current Study	Eastern
2471	2014	Current Study	Eastern
2474	2014	Current Study	W Wales
2477	2014	Current Study	W Wales
2478	2014	Current Study	W Wales
2479	2014	Current Study	W Wales
2480	2014	Current Study	Northern
2481	2014	Current Study	Northern
2482	2014	Current Study	Northern
2483	2014	Current Study	South West
2484	2014	Current Study	South West
2485	2014	Current Study	South West
2490	2014	Current Study	South West
2493	2014	Current Study	Eastern
2503	2014	Current Study	W Wales
2504	2014	Current Study	W Wales
2508	2014	Current Study	W Wales
2515	2014	Current Study	Severn
2519	2014	Current Study	W Wales
2521	2014	Current Study	W Wales

Ind ID	Year	Study	RBD Region
2527	2014	Current Study	Eastern
2529	2014	Current Study	South West
2533	2014	Current Study	Eastern
2544	2014	Current Study	South West

Ind ID	Year	Study	RBD Region
2545	2014	Current Study	South West
2547	2014	Current Study	Other
2548	2014	Current Study	Other
2560	2014	Current Study	Other

Ind ID	Year	Study	RBD Region
2585	2014	Current Study	Northern
2593	2014	Current Study	South West
2594	2014	Current Study	South West
2600	2014	Current Study	South West

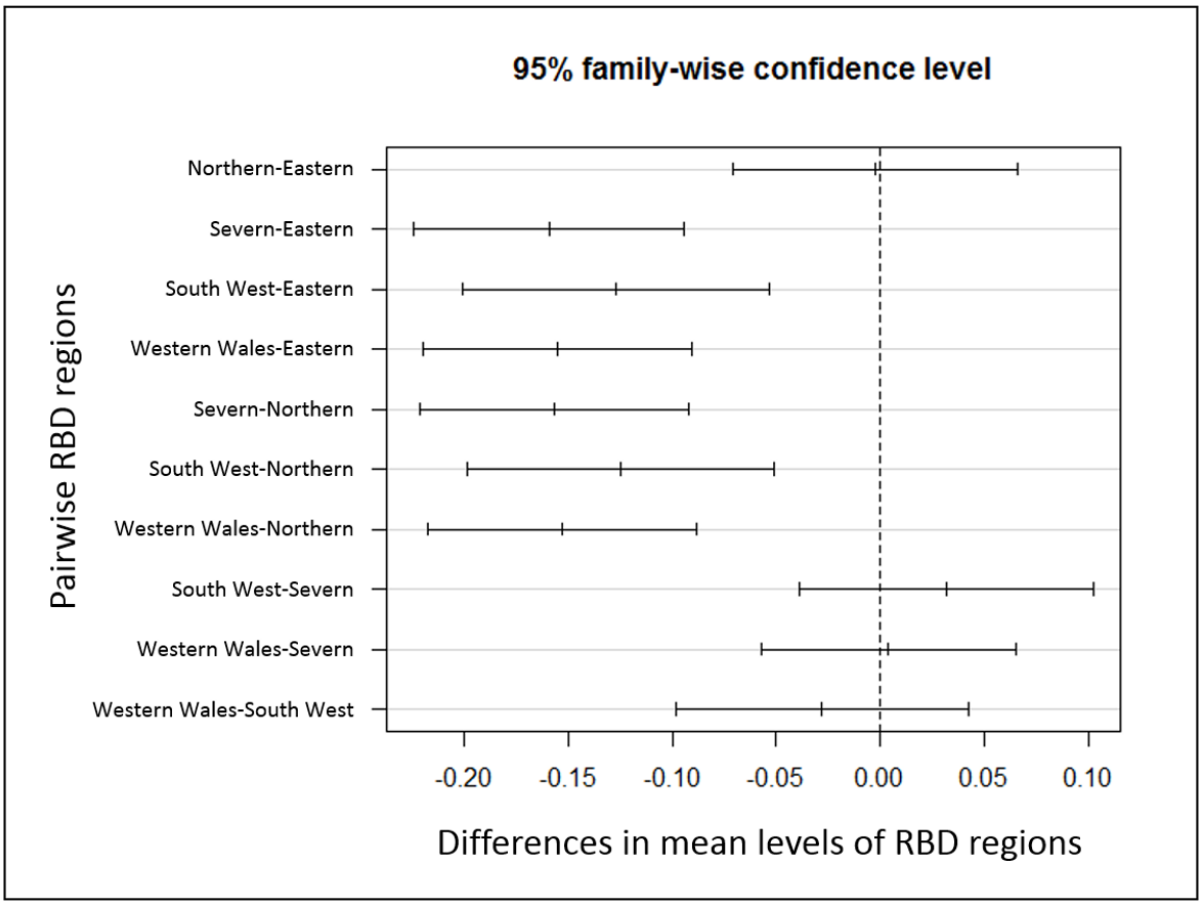
Ind ID	Year	Study	RBD Region
2605	2014	Current Study	South West
2620	2014	Current Study	Severn
2640	2014	Current Study	Severn

S2.3 One-way ANOVA analysis of RBD region and H_o (observed heterozygosity)

Table of Tukey Post Hoc test results (95% family-wise confidence level)

RBD region	diff	lwr	upr	p adj (3 dp)
Northern-Eastern	-0.003	-0.071	0.066	1.000
Severn-Eastern	-0.160	-0.224	-0.095	<0.001***
South West-Eastern	-0.128	-0.201	-0.054	<0.001***
Western Wales-Eastern	-0.156	-0.220	-0.091	<0.001***
Severn-Northern	-0.157	-0.222	-0.092	<0.001***
South West-Northern	-0.125	-0.199	-0.051	<0.001***
Western Wales-Northern	-0.153	-0.218	-0.088	<0.001***
South West-Severn	0.032	-0.039	0.103	0.643
Western Wales-Severn	0.004	-0.057	0.065	1.000
Western Wales-South West	-0.028	-0.097	0.043	0.743

Plot of Tukey Post Hoc test results (95% family-wise confidence level)



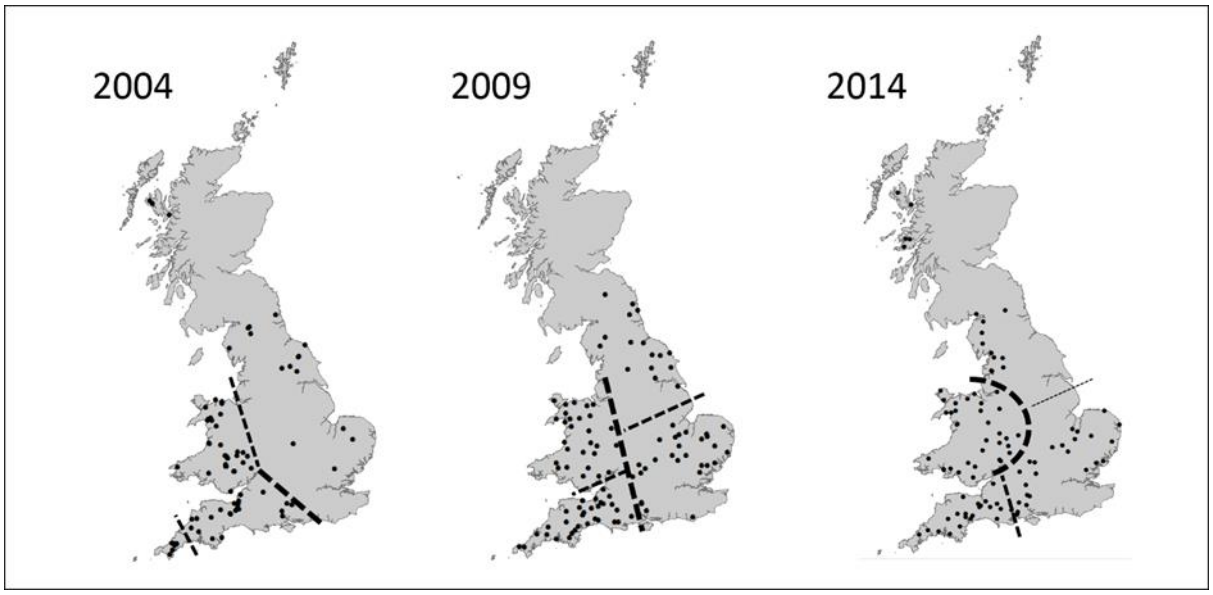
S2.4 F_{IS} values for each region at each time point with information on genetic assignment of the individuals within a region to major and minor geographic clusters. N = number of samples, Major geographic cluster = dominant genetic cluster for that region, Minor cluster = dominant genetic cluster for another region, % All = percentage of all individuals within a region assigned to that cluster (Q<0.5), % Non-admixed = percentage individuals assigned with Q>0.8 to that cluster. % Admixed = percentage of individuals assigned with Q<0.8 to that cluster. F_{IS} is the F_{IS} estimate including all individuals in the RBD region at that time point, $F_{IS}2$ is the F_{IS} estimate using only the individuals genetically assigned to the main geographic cluster for that region, Reason is the probable explanation for the significant F_{IS} values taking into account any differences between F_{IS} and $F_{IS}2$. Asterisks denote significance level after FDR correction with *<0.05 and **<0.01.

Region	Year	N	Major Geographic Cluster		Minor Cluster		F_{IS}	$F_{IS}2$	Reason
			% All	% non-admixed	% Admixed	% All	% non-admixed			
Eastern	1999	17	100%	88%	12%	0%	0%	0.027	0.027	-
	2004	7	71%	71%	0%	29%	29%	0.178	0.105	-
	2009	28	93%	86%	7%	7%	4%	0.107*	0.083	Wahlund
	2014	21	95%	86%	10%	5%	0%	0.084	0.077	-
Northern	1999	16	88%	75%	13%	13%	6%	0.11*	0.093	Wahlund
	2004	11	82%	82%	0%	18%	0%	0.079	0.09	-
	2009	16	94%	56%	38%	6%	6%	0.115*	0.108	Wahlund
	2014	14	86%	79%	7%	14%	14%	0.045	-0.004	-
Severn	1994	6	100%	100%	0%	0%	0%	0.109	0.109	-
	1999	31	90%	87%	3%	10%	6%	0.112*	0.111	Wahlund
	2004	20	85%	80%	5%	15%	15%	0.137*	0.15	Wahlund
	2009	13	85%	77%	8%	15%	8%	0.054	-0.06	-
	2014	14	93%	86%	7%	7%	0%	0.223**	0.241**	Inbreeding
South West	2004	23	96%	91%	4%	4%	0%	0.061	0.043	-
	2009	30	90%	77%	13%	10%	3%	0.124**	0.133*	Inbreeding
	2014	24	75%	58%	17%	25%	13%	0.095*	0.037	Wahlund
Western Wales	1994	19	100%	100%	0%	0%	0%	0.083	0.083	-
	1999	34	100%	91%	9%	0%	0%	0.061	0.061	-
	2004	15	93%	93%	0%	7%	0%	0.122	0.096	-
	2009	17	100%	88%	12%	0%	0%	0.075	0.075	-
	2014	17	100%	94%	6%	0%	0%	0.028	0.028	-

S2.5 Degree of admixture between genetic clusters at different values of K over time. Percentages show the number of individuals considered to be genetically admixed between clusters at a given value of K. Genetically admixed was defined as an individual being <80% assigned to a single genetic cluster. 1994 and 1999 also include samples from the years directly adjacent (i.e. 1994 = 1993, 1994 and 1995) due to low sample sizes. Note that although not excluded, there was low N from Anglian and Northern regions in 1994.

			% Admixture		
Year	**N**	**Area(s) excluded**	K=2	K=3	K=5
1994	28	South West	4%	4%	4%
1999	100	South West	9%	10%	15%
2004	79	None	10%	9%	14%
2009	104	None	14%	17%	21%
2014	96	None	16%	17%	23%

S2.6 Progressive partitioning analysis of three time points where samples were available for the entire study area.



Main population partitions of UK otters as determined by progressive partitioning across three time points. Line thickness denotes the level of the partition with the thickest black line indicating the 1st partition in the data at each time point.

Pairwise FST estimates between major partitions identified at each time point

2004	Wales	SW	Eng
Wales	-	***	***
SW	0.12	-	***
Eng	0.16	0.24	-

2009	Wales	SW	Northern	Eastern
Wales	-	***	***	***
SW	0.23	-	***	***
Northern	0.13	0.13	-	***
Eastern	0.22	0.20	0.07	-

2014	Wales	SW	Northern	Eastern
Wales	-	***	***	***
SW	0.21	-	***	***
Northern	0.12	0.12	-	***
Eastern	0.14	0.14	0.06	-

3

Evaluating environmental factors as facilitators and barriers to gene flow in the Eurasian otter (*Lutra lutra*) using landscape genetics.

Abstract

Eurasian otters have been undergoing a population recovery across much of Europe over the last 30 years. This recovery has led to a near contiguous population within Great Britain, however significant population structure persists. Little is known about the dispersal preferences of otters due to their solitary and nocturnal nature. Functional connectivity provided through effective dispersal of individuals across the landscape is necessary to avoid sustained population fragmentation. Using a large-scale genetic dataset from across the Wales and borders region, we investigate the influence of five environmental factors (habitat, elevation, rivers, slope and roads) at two different spatial scales (100 m^2 and 1 km^2) on otter functional connectivity using a landscape genetics framework. The results showed that habitat is the most important variable in driving the spatial distribution of genetic variation in otters, with broadleaf woodland and arable land facilitating geneflow and coniferous woodland, semi-natural grassland and urban areas impeding gene flow. These results were consistent across both spatial scales. Multi-surface analysis indicated that both elevation and water were also important explanatory variables for functional connectivity when combined with habitat. To our knowledge, these results represent the most comprehensive landscape genetics analysis conducted for Eurasian otters, providing evidence for conservation management on which environmental factors may help increase functional connectivity between previously fragmented otter populations.

3.1. Introduction

3.1.1 Landscape and gene flow

Landscape comprises the mosaic of habitats within which natural terrestrial populations reside. This mosaic is made up of a variety of features, some of which facilitate and others which impede movement. Landscape heterogeneity can be an important factor in influencing patterns of movement and therefore genetic variation across space (Storfer et al. 2007). Identifying landscape features which promote or impede gene flow is crucial in informing effective conservation and management of a species (Crandall et al. 2000; Banks et al. 2005). Barriers to gene flow may be natural, e.g. mountain ridges, or anthropogenically modified, e.g. urban areas, and their presence can mean that populations within or either side of such features become isolated when dispersal across these barriers is unlikely. A fragmented population trapped within 'islands' of suitable habitat and experiencing no gene flow can be vulnerable to genetic threats

such as genetic drift and inbreeding, that can occur in small, isolated populations (even when the overall population size is large) (Frankham 1995a; Reed 2004).

Landscape genetics brings together the disciplines of population genetics and landscape ecology to determine whether certain features significantly influence effective dispersal and therefore geneflow. The binomial categorisation of landscape features into either facilitators or barriers to gene flow is likely to be an oversimplification; landscape variables are more likely to act on a gradient of permeability to movement (Cushman et al. 2006), with potential synergistic or antagonistic interactions between habitat types also likely. The recolonisation of habitats relies on the ability of individuals to move across the landscape (Chapron et al. 2014), and the speed and scale of recolonisation will be mediated, in part, by the dispersal ability of a species. In order to understand the likely future distribution of species and the rate with which that distribution is likely to be achieved we must therefore understand the impact that landscape features have on the ability of individuals to move.

3.1.2 Inferring landscape resistance

The representation of environmental surfaces in landscape genetics are most commonly modelled as resistance surfaces, where values are assigned to each cell in a spatial layer predicting the degree to which that cell facilitates or limits geneflow, and thus functional connectivity, in a given species (Pérez-Espona et al. 2008, Hohnen et al. 2016). Traditionally these resistance values were determined based on expert opinion or empirical knowledge of the species biology and values based on competing cost ratios, where one variable is hypothesized to consistently facilitate geneflow (i.e. resistance = 1) and the other(s) to limit geneflow to varying degrees (i.e. resistance = 2, 5, 10, 100). By testing varying degrees of relative resistance, it is possible to determine the magnitude of the cost ratio (i.e. 1:2 or 1:100) between the variables (Russo et al. 2016). This approach has been applied widely in landscape genetics over the last decade using programs such as CIRCUITSCAPE (Dickson et al. 2019) which require *a priori* user defined resistance values as input (McRae et al. 2008).

However, given that much of the data informing these *a priori* values are patchy/limited for wild species, such predictions may be flawed from the outset. Additionally, environmental variables shown to be important determinants of species distributions or that are used preferentially in or as home ranges may not directly translate into facilitators to dispersal and geneflow (Hohnen et al. 2016). This can result in testing the wrong variables or hypothesis from the outset through a focus on factors important to a species territory rather than to its dispersal. Recently, some studies have tried to address these issues by categorising putative 'facilitating' landscape features with resistance values both higher, and lower, than those of 'limiting' features within the same surface (e.g. 100:1, 10:1, 5:1, 2:1, 1:2, 1:5, 1:10, 1:100) in order to explicitly to include and test the

inverse resistance relationship in addition to the hypothesised one (e.g. Northern Quolls, Hohnen et al. 2016).

Given these caveats in both identifying the role of landscape features accurately and only considering the limited resistance values to explain variation in genetic data, methods which do not rely on *a priori* resistance values are being increasingly used. ResistanceGA is an R package developed to optimise resistance surfaces using genetic algorithms (Peterman 2014, Peterman 2018). As such, the need to subjectively define the resistance values of the categorical or continuous landscape data prior to running the analysis is obviated. This approach has been demonstrated to accurately optimise resistance surfaces across a wide range of scenarios, including variation in sample size, degree of genetic differentiation and amongst alternative correlated surfaces (Winiarski et al. 2020).

3.1.3 Eurasian otters

The population decline of otters (*Lutra lutra*) in Europe in the latter half of the 20th century has been well documented, with a once continuous population reduced to small isolated sub-populations, and loss of the species from much of its original range (Kruuk 2006; Roos et al. 2015). Protective legislation has allowed population expansion and recolonisation of parts of this previous range, however the rate of expansion is dependent on the permeability of the current landscape to otter movement and dispersal. In the UK, recolonisation has been monitored since the 1970's using national surveys for otter sign (Crawford 2010; Findlay 2015; Strachan 2015). Published evidence suggests that features facilitating otter movement include water courses and broadleaf woodland (Mason & Macdonald 1986; Broadmeadow & Nisbit 2004) and features impeding otter movement include urban areas, upland areas and steep slopes (White et al. 2003; Janssens et al. 2008).

While the spatial genetic structure of otters across Europe has been relatively well studied on a variety of scales (Mucci et al. 2003; Janssens et al. 2008; Hobbs et al. 2011; Stanton et al. 2014), the effect of landscape on these spatial patterns has often been inferred indirectly. Janssens et al. (2008) used Bayesian clustering methods to infer that steep slopes and dry lands between river basins in France were positively correlated to the genetic distance between the groups of individuals residing there. Hobbs et al. (2011) identified fine spatial structure in the UK otter population and suggested dispersal might be limited despite the highly mobile nature of the species with future studies needed to address the role of landscape factors in this spatial structuring. Hobbs (2010) undertook some preliminary landscape genetics analysis on otters in the Wales and Borders region of the UK using least-cost pathways and found that both increased slope and upland habitat impeded otter gene flow. Stanton et al. (2014) also used cost-distance matrices produced by least cost pathways to explore the effect of seven landscape variables on

population genetic structure in Scotland, but although significant correlations between all of these variables and genetic distance were found, these relationships were lost once Euclidean distance was accounted for.

3.1.4 Wales and Borders sub-population

Spatial genetic structure of otters in the UK is well described (Hobbs et al. 2011; Stanton et al. 2014; Chapter 2) with five main sub-populations identified. These sub-populations match the geographic distribution of otter population strongholds during the nadir of the decline and reintroduction efforts in the 1980's and 1990's (Hobbs et al. 2011; Stanton et al. 2014), with the high degree of differentiation between these sub-populations likely due to bottleneck effects of this population decline. Given that landscape genetics aims to test the relative influence of landscape factors on gene flow and therefore patterns of spatial population structure (Storfer et al. 2007), where known past population demographics are likely to have heavily influenced that structure this pattern may be obscured. As such this study was undertaken at the sub-population scale, using the Wales and Borders region where otter sampling was most dense.

Here I used a large-scale genetic dataset combined with the latest landscape genetic analyses in order to identify the relative importance of landscape features in facilitating or impeding genetic connectivity of otter populations in the Wales and Borders region.

Specifically, I hypothesise that:

- Water (rivers and coastline) will facilitate otter functional connectivity
- Habitat will significantly explain functional connectivity across the study area with broadleaf woodland incurring low resistance, and urban areas incurring high resistance to otter functional connectivity
- Higher elevations and increased slopes will impede otter functional connectivity
- Different classes of roads will elicit different resistances to otter functional connectivity, based on a combination of their traffic volume and the scale of infrastructure crossing waterways

3.2 Methods

3.2.1 Study area

Previous work by Hobbs et al. (2011) used Bayesian clustering methods to establish the spatial genetic structure of the UK otter population. Four sub-populations were identified across Wales and England with high likelihood, which reflected known otter population strongholds during the nadir of the population decline. Wales and the bordering English counties (Cheshire, Shropshire,

Herefordshire and Gloucestershire) are together known to make up the geographic extent of otters belonging to the 'Wales and borders' sub-population within the UK (Hobbs et al. 2011). This region covers 31,895 km^2 with an elevation range from 0 to 1085 meters above sea level (masl). The climate is temperate, and the predominant land use is agricultural (80%), followed by forestry (15%) (Natural Resources Wales 2020). The area also includes urban conurbations mainly along the North coast and in the South East of Wales, and in the border's region of England. Major roads and motorways including the M4, M5, M6 and A55 link the more urbanised areas, with a network of smaller roads across the more rural areas of mid and west Wales.

3.2.2. Sample collection and genotyping

Otter carcasses from known locations were collected across the region, predominantly after road traffic accidents, and sent to the Cardiff University Otter Project. Muscle tissue samples were taken during routine *post-mortem* examinations and stored in 80% ethanol at -20 °C. Tissue samples from 192 otters collected between 1993 and 2007 and genotyped at 15 microsatellite loci (Dallas & Piertney 1998; Dallas et al. 1999; Huang et al. 2005) by Hobbs et al. (2011, see Chapter 2 methods for full lab detail) were selected for the current study (Figure 3.1).

Figure 3.1: Sample distribution map. Great Britain (grey) and Wales and Borders area (black box) with sample locations shown as black points.

3.2.3 Genetic distance estimates

Pairwise genetic relatedness was estimated for each pair of otters within the study by calculating proportion of shared alleles (POSA) in GenAlEx 6.5 (Peakall and Smouse 2006; Peakall and Smouse

2012). POSA was used as it does not rely on underlying biological assumptions, making it appropriate for use with any population at any level of inbreeding, while performing similarly well to PCA-based genetic distance metrics (Shirk et al. 2018). POSA was converted to a genetic distance measure by calculating 1 – POSA, the proportion of different alleles (D_{PS}) between two individuals (Bowcock et al. 1994).

3.2.4 Environmental data and constructing resistance surfaces

Environmental variables were selected based on their likelihood to facilitate or impede otter geneflow and were based on published evidence of otter preference, or influence on otter prey availability (Mason & Macdonald 1986; White et al. 2003; Broadmeadow & Nisbit 2004; Janssens et al. 2008). The landscape factors selected for analysis included: rivers and coastline, habitat, roads, elevation and slope (Figure 3.2). Following recommendations that grain size should be smaller than the home-range size of the study species (Anderson et al. 2010) a coarse scale of 1 km^2 was selected. However, as the landscape data contained two linear features (rivers and roads) a fine scale approach of 100 m^2 was also used to maintain linearity within the dataset once data were rasterised. Manipulation of landscape layers was undertaken in ArcGIS 3.10 (ESRI) and R (Core Team, 2017). All landscape data used in analysis, and their origins, are defined in Table 1.

Data for rivers across the study area was taken from the Waterways 1:50,000 dataset produced by the Centre for Ecology and Hydrology (CEH). These data were combined with the high mean water polygon (OS OpenData) of the UK to include the coastline. This rivers and coastline network was then rasterised to 100m^2 with all cells in the raster being labelled as ‘water’, ‘land’ or ‘open sea’.

Habitat data from across the study area was taken from the LandCover Map 2007 25 m raster V1.2 (LCM2007) produced by CEH. These data are derived using multispectral imagery data from the Landsat 8 Operational Land Imager to determine the land class of each 25 m^2 pixel (Morton et al. 2014). The data were resampled to a resolution of 100m^2 (using the nearest neighbour method) and grouped according to the aggregate classes which reduced the data down from 23 LCM2007 classes to 10 more broad classes, plus an 11th class indicating open sea, i.e. any area within the raster beyond the low mean water mark (see Table 3.1 for detail). The classes were used to provide broad habitat types as although broadleaf woodland is hypothesised to facilitate otter movement (Mason & Macdonald 1986; Broadmeadow & Nisbit 2004) other habitat types remain understudied. These classes are broadly equivalent to other habitat definitions used in ecological research such as Phase 1 habitat surveys and the EC Habitat Directive Annex 1 types (Jackson 2000).

Elevation across the study area was collated using the Ordnance Survey Digital Terrain Model (OS 50 DTM) which provides 50 m^2 resolution. The collated raster was resampled to 100 m^2 using the bilinear approach which is suitable for continuous datasets. This raster was then reclassified into 4 categories; 0-300 m, 300-600 m, 600 m, representing low, medium and high elevations across the study area, and open sea.

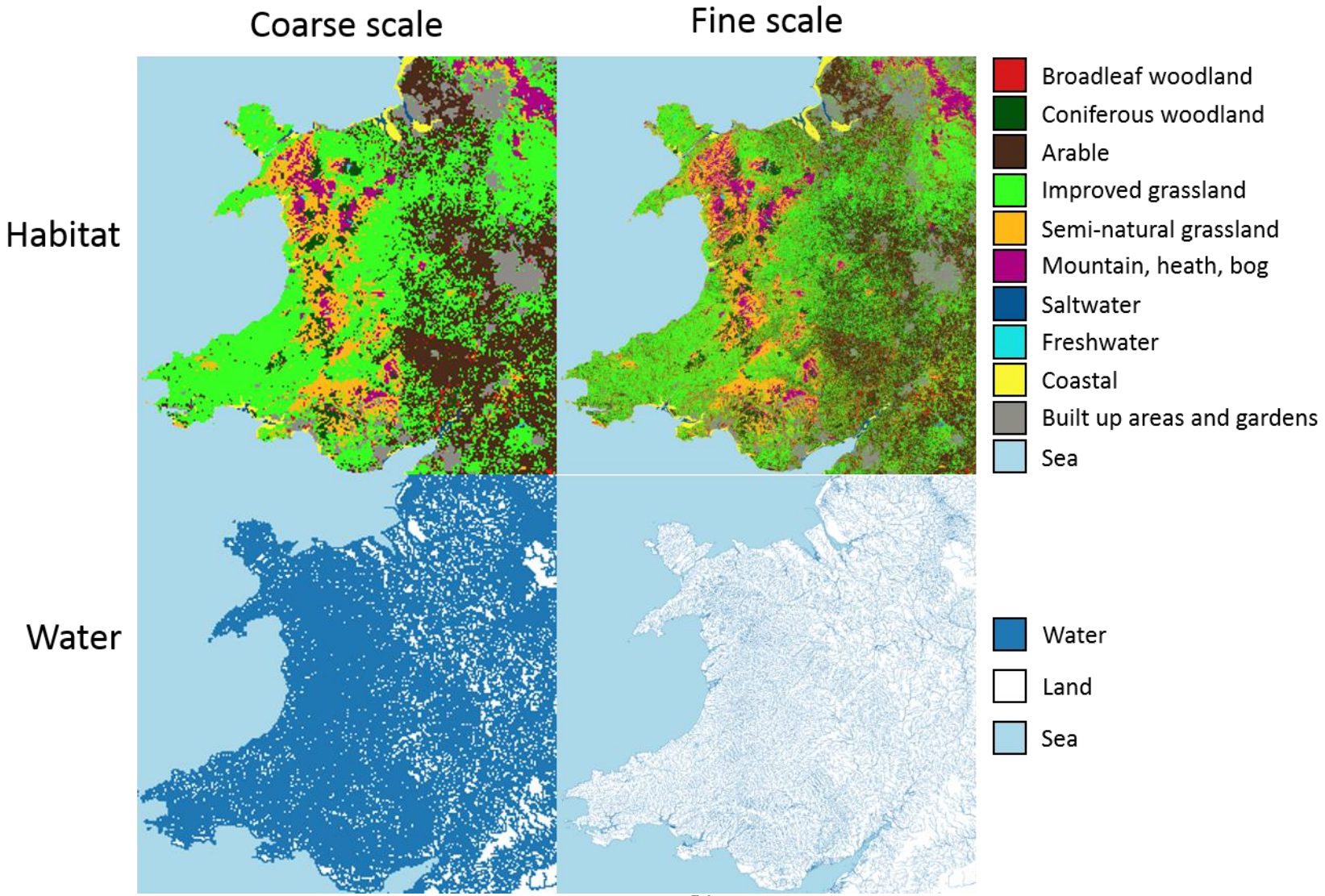
Slope was calculated using the slope tool in ArcGIS which identifies the maximum rate of change in Z value between cells in a raster in degrees, based on the 100m DTM raster (produced above) as the input. The continuous data was then reclassified into six classes; 0-10 degrees, 10-20 degrees, 20-30 degrees, 30-40 degrees, 40+ degrees, and open sea.

Polyline data for roads across the study area was taken from OS roads and rasterised to 100 m^2 with larger roads taking precedent over smaller roads during the rasterisation process. All types of minor roads (below B road class) were grouped together and A roads were split into those that were single versus dual carriageway (for full details on categorisation see Table 3.1).

For resampling of the landscape rasters from 100 m^2 (fine-scale) to 1 km^2 (coarse-scale) a priority rule was used for linear features (water and roads) where grid cells were categorised as containing a feature whatever the proportion present. For the three other landscape predictors (elevation, slope and habitat) a majority rule was applied, where the new cell was classified based on the class with the largest area within the cell (Table 2). The fine scale raster layers contained nearly 6.8 million pixels each while the coarse scale raster layers contained over 67,000 pixels each.

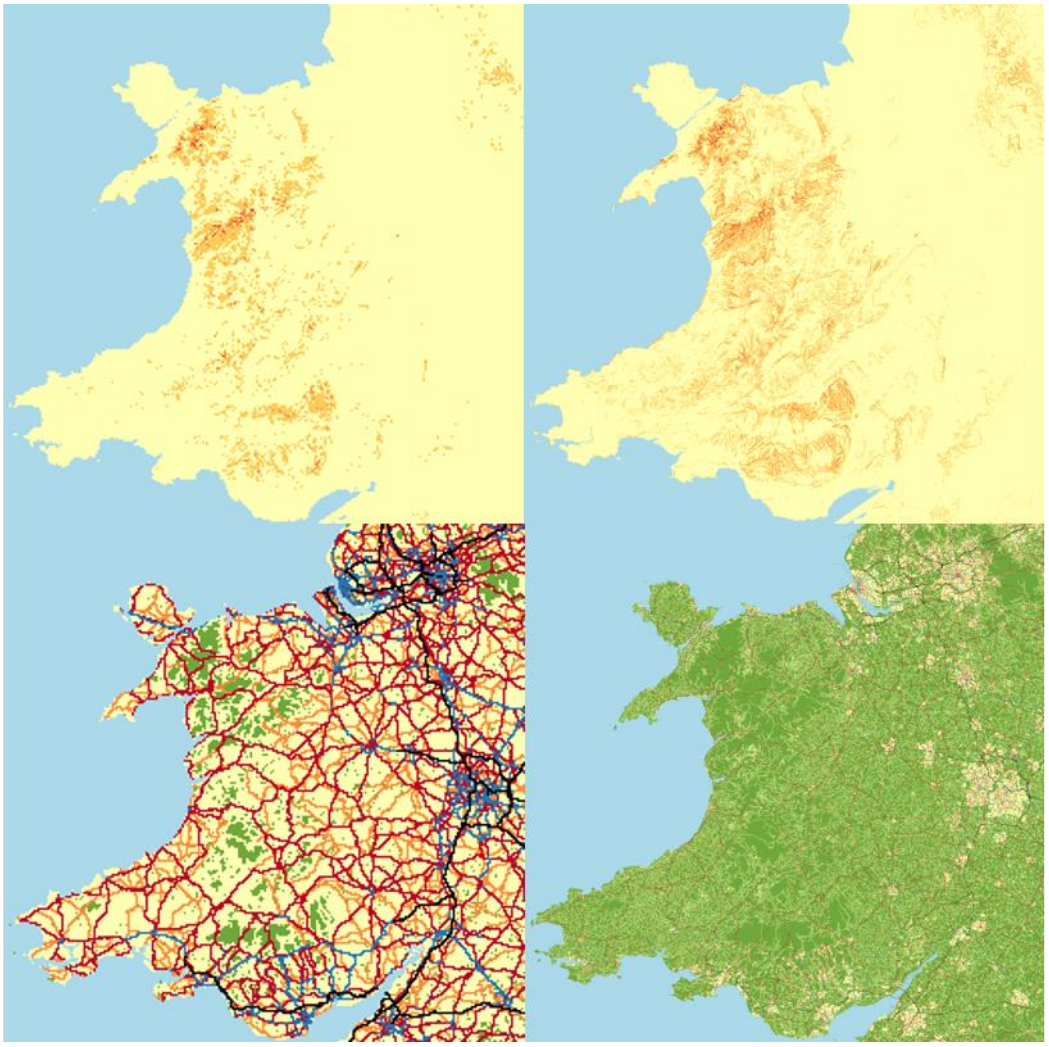
Table 3.1: Landscape features included in the analysis. Feature gives a description of the variable, No is number of classes within that feature, Classes gives details of the categories used in the current analysis, Categorisation of raw data' defines the original categories / numeric range on which the 'Classes' used in the current analysis were based. Note that in some cases these were the same while in others, the new 'Class' was based on the amalgamation of two or more original categories. Resample defined the method used to resample the 100 m^2 raster to 1 km^2. Data source provides information on the source of the spatial layers.

Feature	No.	Classes	Categorisation in raw data	Resample	Data Source
Water	3	Water Land Sea	Waterways and 100m buffer around coastline All other land Sea (apart from within 100m coastline buffer)	Min	Waterways 1:50,000K (CEH) UK HMW polygon
Habitat	11	Broadleaf woodland Coniferous woodland Arable Improved grassland Semi-natural grassland Mountain, heath, bog Saltwater Freshwater Coastal Urban Sea	Broadleaf woodland Coniferous woodland Arable & horticulture Improved Grassland Rough Grassland, Neutral Grassland, Calcareous Grassland, Acid Grassland, Fen marsh swamp Heather, heather grassland, bog, montane habitats, inland rock Saltwater Freshwater Littoral rock, sub littoral rock, littoral sediment, sub littoral sediment, saltwater Urban, suburban Unclassified (open sea)	Mode	LandCover Map 2007 (Edina Digimap)
Roads	7	Minor road B Road A Roads single carriageway A Roads dual carriageway Motorway Other Land Sea	C, unnamed and other minor road categories B roads A roads excluding dual carriageways A roads excluding single carriageways Motorway All other land (as described by the LCM2007) Unclassified (on LCM2007 i.e. open sea)	Max	OS roads
Elevation	4	Low Elevation Mid Elevation High Elevation Sea	-7.4 – 300 m.a.s.l. 300 – 600 m.a.s.l. 600 – 1085 m.a.s.l. Unclassified (on LCM 2007 i.e. open sea)	Mode	OS DMT 50 UK LMW polygon
Slope	6	0 – 10 degrees 10 – 20 degrees 20 – 30 degrees 30 – 40 degrees 40+ degrees Sea	0 – 10 degrees 10 -20 degrees 20 -30 degrees 30 – 40 degrees 40 - 52 degrees Not land (i.e. Sea)	Mode	OS DMT 50 UK LMW polygon


Coarse scale
Fine scale
Habitat
Water
Broadleaf woodland
Coniferous woodland
Arable
Improved grassland
Semi-natural grassland
Mountain, heath, bog
Saltwater
Freshwater
Coastal
Built up areas and gardens
Sea
Water
Land
Sea

Slope

Roads



0 – 10 degrees

10 – 20 degrees

20 – 30 degrees

30 – 40 degrees

40 + degrees

Sea

Land

Minor road

B road

A road (single carriageway)

A road (dual carriageway)

Motorway

Sea

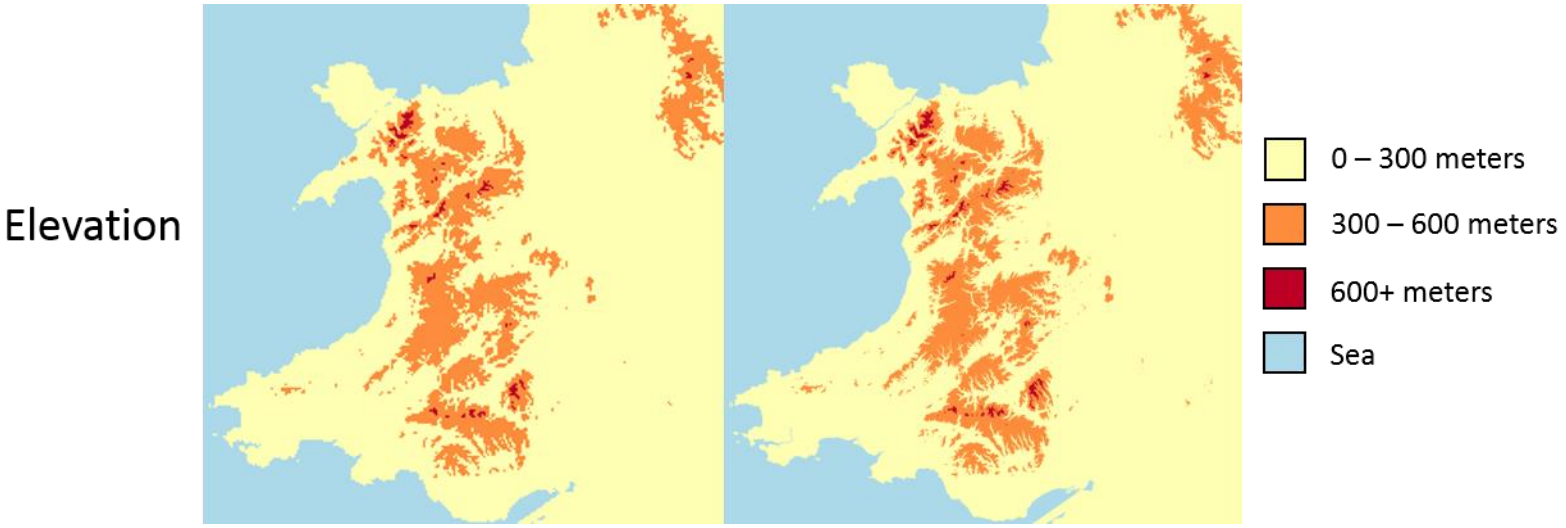


Figure 3.2: Input landscape variable raster data. Data are shown at both coarse (1 km^2) and fine (100 m^2) scale. Within figure legends describe the categories attributed to each landscape variable, while maps detail the distribution of these categories across the study area.

3.2.5 Landscape resistance modelling

The effects of the selected landscape features on functional connectivity were explored using the R package ResistanceGA 4.1 (Peterman 2014; Peterman 2018). Pairwise resistance values are calculated between all individuals but unlike earlier landscape genetic methods, ResistanceGA does not require *a priori* estimates of resistance values for the various landscape classes within a particular layer. Instead this method uses genetic algorithms, which mimic natural selection to explore the parameter space and find the optimal resistance value for each category (Scrucca 2013). To do this a random population of individuals, where each 'individual' is a parameter surface with a unique combination of resistance values is generated, pairwise resistance distances between each of the georeferenced samples are calculated across this surface. These distances are then evaluated against the pairwise genetic distances, while accounting for non-independence between the pairwise samples, using linear mixed effects models with maximum-likelihood population effects (MPLE) (Clark et al. 2002; Van Strien et al. 2012) as implemented in the R package lme4 (Bates et al. 2014). The 'fitness' of each simulated individual (i.e. model support) is determined using log-likelihood as the objective function and those individuals with the highest log-likelihood values are permitted to reproduce, generating the next generation of individuals. This new population inherits the traits (i.e. resistance value parameters) from those with the highest model support in the previous generation, with the addition of random mutation and recombination of the parameters. This process is repeated each generation until 25 generations pass without improvement to the log-likelihood of the MLPE model, at which point the resistance surface is considered optimal for the chosen landscape variable(s). In short, resistance estimates are calculated based on their ability to explain the variation in the genetic distance between individuals across the landscape. This removes potential biases that *a priori* resistance values derived from ecological data or expert opinion might introduce. The MLPE approach of linear-mixed modelling has been shown to result in a high degree of accuracy during model selection in a variety of simulated circumstances (Shirk et al. 2018) and thus provides a robust approach to determining the most appropriate explanatory landscape variables for the genetic dataset tested. After optimisation, models were compared using the sample size corrected Akaike Information Criterion (AICc). The model with the lowest AICc was considered the best model, models within 2 AICc units were considered equivalent, and models greater than 10 AICc units from the best model were considered uninformative (Burnham and Andersson 2002).

Both categorical and continuous surfaces may be optimised by ResistanceGA, either singularly where a set of optimised resistance values is produced for one landscape predictor at a time, or as multiple resistance surfaces where multiple landscape predictors are optimised simultaneously to provide a combined optimised resistance surface. The optimisations in this study are based on categorical data for each landscape predictor due to computational limitations and were

conducted using log-likelihood as the method of optimising resistance values and *commuteDistance* (Kivimäki et al. 2014) which has been shown to be equivalent to CIRCUITSCAPE (circuit theory-based estimation that considers all possible routes across a landscape simultaneously; McRae, Shah, & Edelman 2013) but with the added benefit that the process can be parallelised, reducing the computer processing time accordingly (Marrotte & Bowman 2017).

All optimisations were run twice, using different starting seeds, to check the convergence between independent runs of each landscape feature or combination of features. This is due to the potential for the algorithm to converge on an incorrect optimisation, which with real data, would be impossible to identify from a single run (Peterman 2017). Optimisations were run using 30 parallel central processing units (CPUs) each with 25 GB of random-access memory (RAM) providing a total of 750 GB RAM for each optimisation.

The XY co-ordinates of each otter were used as the input variable for the location of each individual during the *gdist.prep()* function to generate a pairwise distance matrix between all individuals to compare optimised pairwise resistance distances with.

3.2.6 Single surface optimisations

Individual single-surface analyses were performed at both fine (100 m^2) and coarse (1 km^2) spatial scales using the *SS_OPTIM* function. After the five surfaces had been optimised a pseudo-bootstrap analysis was run using the *RESIST.BOOT* function, to assess the robustness of each optimised surface using sub-sampling, using the default parameters of 75% sample proportion and 1000 iterations. During this process the dataset is sub-sampled without replacement at each iteration, the MLPE model is refitted for each optimised surface and AICc scores are calculated. If the analysis is robust, the relative ranking among optimised surfaces will remain stable, while any changes suggest that there are specific individuals in the dataset driving the resistance-genetic relationship.

Additionally, the *RESIST.BOOT* function was run on a dataset of combined coarse and fine scale single surface optimisations to determine which scale provided a better model fit for the genetic data for each landscape variable. This analysis was run using the optimised landscape surface with the lowest AICc for each landscape variable so that only one surface per landscape variable at each scale was analysed.

3.2.7 Multi-surface optimisations

Multi-surface optimisation, using the *ALL_COMB* function, was performed using the coarse-scale dataset only due to computational restrictions (30 CPUs with 25 GB RAM each). Combinations from one to all five landscape surfaces were tested. The *ALL_COMB* function first optimises the multi-surface resistance surfaces using MS_optim and then calls RESIST.BOOT to perform pseudo-

bootstrap analysis on the optimised resistance surfaces. The analysis was carried out using default parameter (as above with for the single surface models).

3.2.8 Visualising optimised surfaces

The top models from both the single and multi-surface optimisations at both scales were used as input resistance data for CIRCUITSCAPE V4.0 (McRae et al. 2013). CIRCUITSCAPE uses circuit theory to generate landscape conductivity surfaces that consider all possible pathways across a landscape simultaneously (McRae et al. 2008). The pairwise resistance estimates were generated between all the samples in the dataset and a current map for each input layer produced from these. The current maps generated were then visualised using ArcGIS 3.1.

3.3 Results

3.3.1 Effects of individual landscape features on functional connectivity

The ranking of single-surface optimisations remained consistent after bootstrapping and the top models showed consistent results across both spatial scales (100 m^2 and 1 km^2). Habitat was the best supported model using AIC_C (100 m^2: AIC_C = -24592.5, ΔAICc = 0; 1 km^2: AIC_C = -24629.4, ΔAICc = 0) and explained 19% and 17% of the genetic variation seen between individuals in the fine and coarse-scale models respectively (Table 3.2). The bootstrapping analyses indicate that both habitat and elevation better explained genetic variation in the sample than geographic distance alone. This pattern was consistent across both spatial scales, although landscape predicators that provided poorer model fits than geographic distance were inconsistent in their ranking across the two spatial scales. This may be due to both roads and water being linear features, which when modelled at a coarse scale would have covered a much higher proportion of the total study area than when modelled at a fine scale.

Table 3.2: Bootstrap results of single surface RESISTANCEGA analysis for both fine and coarse scale data. Results from the fine (a) and coarse (b) scale datasets were compared. Predictor, the landscape surface each model is based on; AICc, the average AICc values obtained for each model across 1000 bootstrap iterations; Iter, the iteration (with the highest AICc value) used as input for the analyses; k, the number of parameters in each model; ΔAICc, the difference in the average AICc value between the best supported model and the model in question, Weight, the average AICc weight across 1000 bootstrap iterations; R^{2m}, the average marginal R^2 across 1000 bootstrap iterations; Top model (%), the percentage of times that the given model was the top selected model across 1000 bootstrap iterations.

Predictor	AICc	Iter	k	ΔAICc	Weight	Rank	R^{2m}	Top Model (%)
(a) Fine scale – 100 m^2								
Habitat	-24592.5	2	12	0	0.916	1.08	0.19	92.3
Elevation	-24569.5	1	5	23.0	0.063	2.32	0.09	5.8
Geographic Distance	-24561.9	1	2	30.6	0.020	2.75	0.05	1.9
Water	-24542.7	1	4	49.8	<0.001	4.23	0.09	0
Slope	-24532.6	2	7	59.9	<0.001	4.63	0.14	0
Roads	-24294.8	2	8	297.7	<0.001	6.00	0.35	0
(b) Coarse scale – 1 km^2								
Habitat	-24629.4	2	12	0	0.996	1.00	0.17	99.7
Elevation	-24586.9	1	5	42.5	0.004	2.01	0.12	0.3
Geographic Distance	-24551.2	1	2	78.2	<0.001	3.07	0.05	0
Roads	-24521.1	1	8	108.3	<0.001	4.36	0.10	0
Water	-24516.8	1	4	112.6	<0.001	4.56	0.09	0
Slope	-24423.9	1	7	206.3	<0.001	6.00	0.27	0

In 92-99% of the bootstrap runs, habitat was selected as the top model, with the ΔAICc of the second ranked model (elevation) much greater than 10 (23.0 – 42.5; Table 3.2). As such only the optimised resistance results for the habitat surface are described below.

At the finer scale (100 m^2) the habitat class with lowest optimised resistance value (i.e. that best facilitated gene flow) was broadleaf woodland, which was followed by arable land and saltwater respectively. These three lowest resistance habitats were ranked consistently across both runs. The three habitat classes with the highest optimised resistance values (i.e. that most impeded gene flow) were also consistent across the two runs and were coniferous woodland, urban and semi-natural grassland in descending order. The other six habitat categories varied in their respective intermediary resistance rankings across the two runs (see Figure 3.3 for visualisation and Supplementary 3.1 for full details).

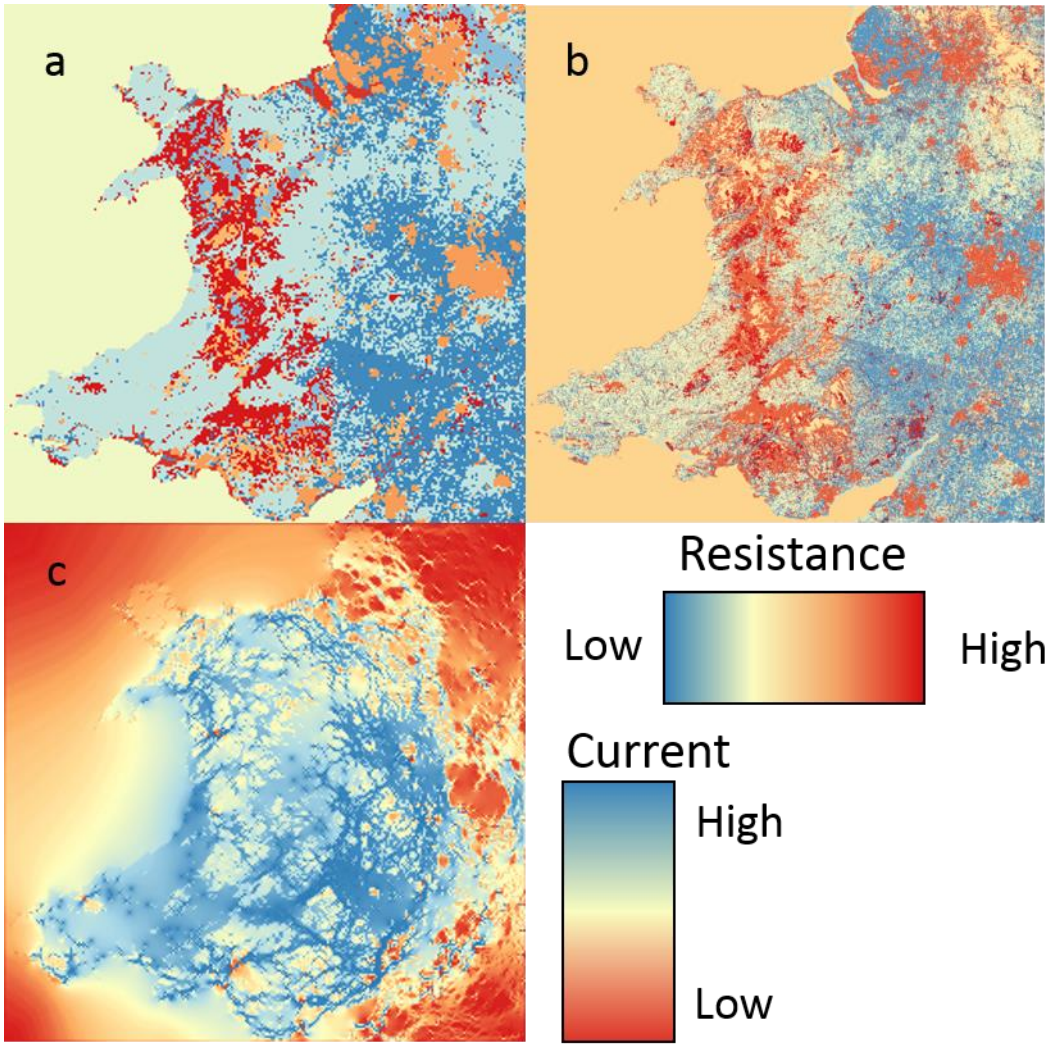


Figure 3.3: Optimised resistance surfaces and current map for single-surface optimisation of habitat at both a fine (100 m2) and coarse (1 km2) scale. A) optimised resistance surface at 1 km^2, b) optimised resistance surface at 100 m^2, c) current map generated using the 1 km^2 optimised resistance surface showing connectivity between the 192 otter sample locations. In both cases blue (i.e. low resistance or high current) represents areas that facilitate gene flow and red (i.e. high resistance and low current) represents areas that impede gene flow.

At the coarser scale (1 km^2) the habitat class with the lowest optimised resistance value was also broadleaf woodland, which was followed by arable land. Semi-natural grassland had the highest optimised resistance value. The remaining eight habitat classifications were different in their respective rankings between runs, but urban and coniferous were both at the high resistance value end of this set indicating some overall consistencies between habitat classes assigned the lowest resistance values (broadleaf woodland and arable) and highest resistance values (semi-natural grassland, urban and coniferous) at both spatial scales (see Figure 3.3 for visualisation and Supplementary 3.2 for full details).

3.3.2 Effects of grain scale on model fit

When optimised single surfaces were directly compared across the two spatial scales, the coarse scale (1 km^2) representation of habitat and elevation provided a better model fit than the fine scale (100 m^2) representations of these variables (Table 3.3). Both variables were better models of

functional connectivity than geographic distance regardless of scale. This suggests that 1 km^2 is a more appropriate scale for explaining functional connectivity in otters than 100 m^2. Although fine scale landscape variables provided a better fit for two of the three remaining predictors, overall, these ranked below geographic distance in their explanatory power.

Table 3.3: Bootstrap results of the comparison of single surface RESISTANCEGA analyses across both fine and coarse scale data. Predictor, the landscape surface each model is based on; Scale, the spatial scale of the predictor surface (either coarse or fine); Iter, the iteration (with the highest AICc value) used as input for the analyses; k, the number of parameters in each model; ΔAICc, the difference in the average AICc value between the best supported model and the model in question, Weight, the average AICc weight across 1000 bootstrap iterations; R^{2m}, the average marginal R^2 across 1000 bootstrap iterations; Top model (%), the percentage of times that the given model was the top selected model across 1000 bootstrap iterations.

Predictor	Scale	Iter	k	ΔAICc	Weight	Rank	R^{2m}	Top Model (%)
Habitat	Coarse	2	12	0	0.997	1.00	0.18	99.7
Elevation	Coarse	1	5	43.04	0.003	2.54	0.12	0.4
Habitat	Fine	2	12	43.59	<0.001	2.61	0.19	0
Elevation	Fine	1	5	66.75	<0.001	4.61	0.09	0
Geographic Distance	Fine	1	2	73.34	<0.001	4.73	0.05	0
Geographic Distance	Coarse	1	2	78.84	<0.001	5.91	0.05	0
Water	Fine	1	4	92.98	<0.001	7.33	0.09	0
Slope	Fine	1	4	103.77	<0.001	8.10	0.14	0
Roads	Coarse	1	8	109.11	<0.001	8.97	0.10	0
Water	Coarse	2	7	113.00	<0.001	9.19	0.09	0
Slope	Coarse	1	7	208.19	<0.001	11.00	0.27	0
Roads	Fine	2	8	342.10	<0.001	12.00	0.35	0

3.3.3 Effects of multiple coarse scale landscape features on functional connectivity

Testing all combinations of the five landscape variables produced 30 landscape resistance models (25 multi-surface models and 5 single-surface models) plus the geographic distance model. The three best predictive models for patterns of genetic distance in otters (at the coarse scale of 1 km^2, after bootstrapping) were a single surface model of habitat and two multi-surface models, one including habitat and elevation and the other including habitat and water (Table 3.4). Cumulatively, one of the best three models (as determined by bootstrapping) was selected as the top model in more than 96% of instances, indicating a strong likelihood that these three variables collectively explain a clear majority of the functional connectivity between otters at the 1 km^2 scale. Habitat featured in all three top models, and further breakdown of the percentage contribution by each surface showed that habitat explained 91% of the variation in the Habitat + Elevation model, and 63% of the variation in the Habitat + Water model, thus clearly demonstrating its high contribution in determining functional connectivity for otters (out of the tested landscape variables; see Figure 3.4 for visualisation and supplementary 3.2, 3.3, 3.4 and 3.5. for further details on resistance values).

The large majority (22/30) of the optimised landscape models performed better than geographic distance (after bootstrapping) indicating that the landscape variables selected for this study are relevant to otter functional connectivity across the study area (see Supplementary 3.6 and 3.7). As well as providing better model fit, all 30 landscape models describe more of the variation in genetic distance (based on R^{2m}) compared to geographic distance (between 9 – 32% and 5% respectively).

Table 3.4: Model selection for all multi-surface and single surface landscape variables at 1km² resolution for otter gene flow. Landscape surfaces include land surface elevation (Elevation), aggregated habitat classification from LCM 2007 (Habitat), degree of slope (Slope), road type (Roads) and rivers & coastline (Water). Results shown are a subset of the 31 surface models tested (full table can be found in the supplementary) and are the top 10 ranked models (see 'Rank') plus any model with a top model % >0 after bootstrapping, geographic distance was also included for comparison. Models are ranked based on average AICc after 1000 iterations of bootstrapping. Parameters included are: the model surface/s (Surface), number of model parameters (k), average Akaike Information Criterion for small sample sizes (AICc), the average Akaike weight (Weight), the average rank assigned to the model (Rank), the marginal R-squared (R^{2m}) and the percentage of times the model was ranked top (Top Model %). Bold font indicates the three top models. * this model had a ΔAICc < 2 in the repeat run and is therefore considered in the top three models.

Predictor	k	AICc	ΔAICc	Weight	Rank	R^{2m}	Top Model (%)
Habitat	12	-24631.8	0	0.437	2.5	0.176	34.5
Habitat + Elevation	16	-24631.2	0.6	0.369	1.869	0.178	49.5
Habitat + Water*	15	-24628.7	3.1	0.138	3.272	0.176	12.5
Habitat + Elevation + Water	19	-24623.2	8.6	0.016	4.698	0.179	0.6
Habitat + Slope	17	-24621.4	10.4	0.003	6.427	0.182	0
Habitat + Elevation + Slope	21	-24618	13.8	0.002	6.934	0.172	0
Habitat + Elevation + Roads	23	-24617.1	14.7	0.004	6.415	0.196	1
Habitat + Roads	19	-24616.4	15.4	0.002	8.688	0.190	0
Habitat + Slope + Water	20	-24613.4	18.4	<0.001	10.387	0.173	0
Habitat + Elevation + Roads + Water	26	-24608.6	23.2	<0.001	10.105	0.184	0
Elevation + Roads	12	-24606.6	25.2	0.013	13.474	0.170	1
Elevation + Slope	10	-24594.8	37	0.006	17.592	0.125	0.4
Elevation + Water	8	-24579.1	52.7	0.007	20.307	0.117	0.4
Elevation	5	-24575.1	56.7	0.002	21.767	0.137	0.1
Geographic Distance	2	-24571.3	60.5	0.001	22.339	0.0471	0

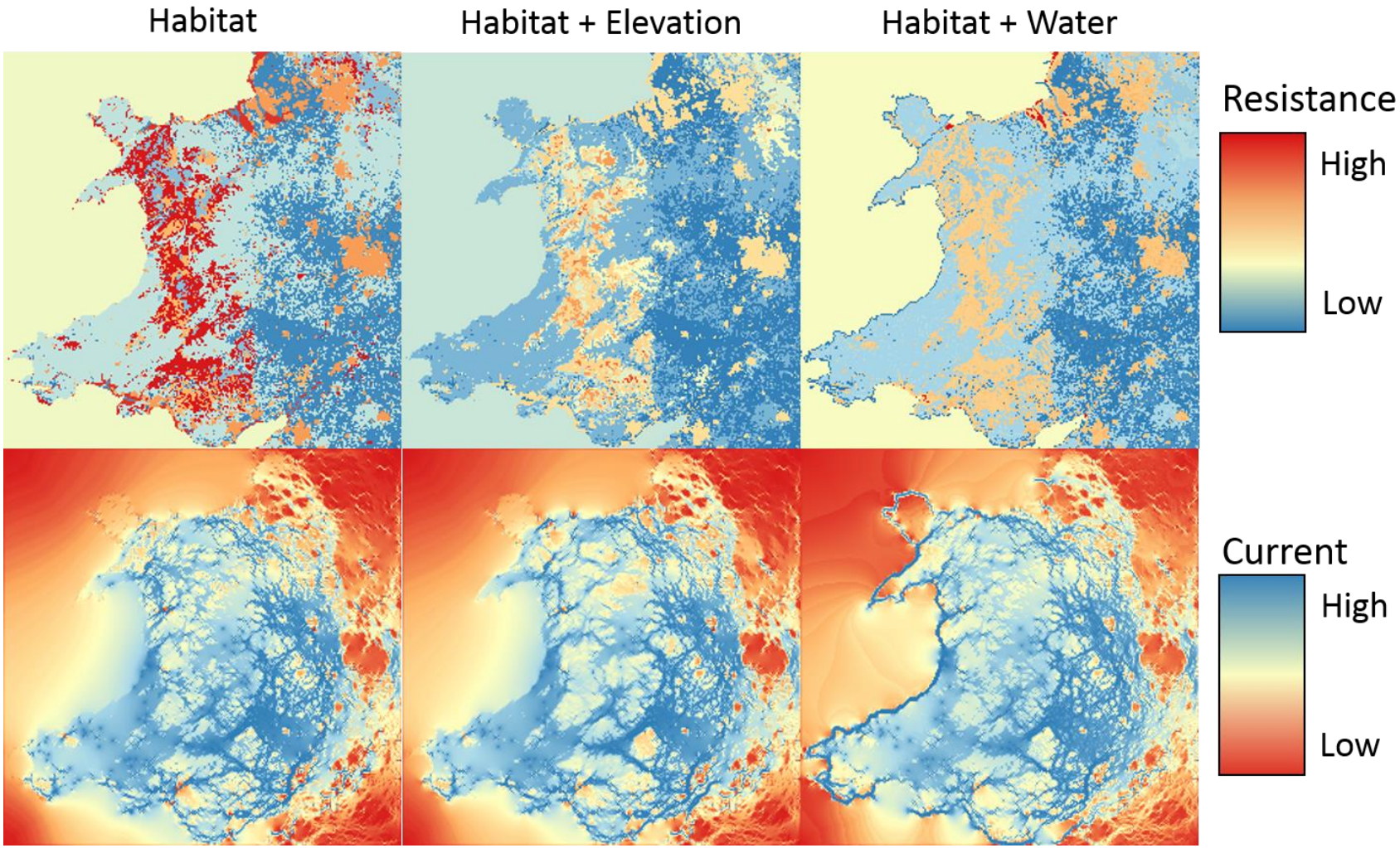


Figure 3.4: Optimised resistance and current maps for the top three multi-surface models. Top row shows the optimised resistance maps for each model with low resistance (blue) indicating areas facilitating gene flow and high resistance (red) indicating areas impeding gene flow. Second row shows the current maps between all individuals sampled with high current (blue) indicating facilitation of gene flow and low current (red) indicating impediment of gene flow.

3.4 Discussion

The present study aimed to establish the importance of five landscape variables, hypothesised to be important to otter movement, in facilitating or restricting geneflow across the study area. The analysis indicated that several of the landscape variables partially explain the degree of functional connectivity between the sampled otters. Habitat was the best supported landscape variable at both spatial scales and across both single and multi-surface optimisations, with results suggesting that broadleaf woodland and arable land facilitate gene flow, while semi-natural grassland, urban and coniferous woodland impede gene flow. In combination with habitat, elevation and water were also well supported models (ΔAICc < 2) with higher elevations impeding geneflow and the presence of a river or coastline facilitating geneflow in these scenarios. Slope and roads did not feature as variables in any of the best supported models and were ranked below geographic distance in the single surface optimisations at both grain scales.

These results indicate that of the five tested landscape variables habitat was the most important in shaping spatial genetic structure, as indicated both by its inclusion in all three top models, and by the high percentage contribution it provided to both multi-surface models. Elevation and water were also important landscape variables when considered in conjunction with habitat.

These results should inform conservation management of the species by indicating which habitat classes along with what combinations of landscape variables facilitate or impede geneflow.

3.4.1 Comparison with other areas of Europe

Previous work on distribution modelling for otters in Spain found that presence probability was negatively correlated with distance to urban centres and highways, strongly negatively correlated to altitude, and positively correlated to factors indicating higher abundance of freshwater (Barbosa et al. 2003). While that study could not consider how the factors affected gene flow, similarities exist between those variables found to be important in predicting otter presence in Spain and those positively related to gene flow in the current study, with higher altitudes and urban areas impeding gene flow and water facilitating gene flow (when modelled as a multi-surface with habitat). Similarly consistent with the present findings, a pan-European study of factors affecting otter occurrence also found human density and road density (degree of urbanisation) to be negatively associated with otter occurrence (Robitaille and Laurence 2002). That study did not find any relationship between forest cover and otter occurrence; however, the study did not distinguish between broadleaf and coniferous forests, two habitat types that were found to have a converse effect on gene flow in our study. Alternatively, the lack of relationship could also suggest that landscape factors linked to otter occurrence may be different from those linked to dispersal and gene flow.

Hobbs (2010) tested five landscape variables; broadleaf woodland, rivers, anthropogenic factors, upland habitats and slope; for effects on otter gene flow across the Wales and Borders region using least-cost pathways. The samples used were the same as in the current study, yet the results differed considerably with only slope and upland habitats having a significant (negative) effect on gene flow across the region in the previous study, however it is important to note that the environmental variables tested between the two studies also differ. In the current study slope did not provide a good model fit at either grain scale when tested as a single-surface, although it was included in multi-surface models that ranked above geographic distance (but did not feature in any of the three top models). Upland habitat was included as one of the classes in the habitat landscape variable in the current study and optimised resistance surfaces for the 11 habitat classes indicated that upland habitats conferred a medium resistance to gene flow. Unlike in the Hobbs (2010) study, broadleaf woodland was determined to facilitate gene flow in the current study. There are two main reasons that could explain these differences results; firstly, least-cost pathways only consider a single optimal route between each pair of individuals and are therefore likely to be overly simplistic through the assumption that individuals have prior knowledge of the landscape, whereas the random-walk commute time used in the current study uses all possible dispersal pathways in its prediction of gene flow (van Etten 2017). Secondly, unlike in the current study, the methods used by Stanton et al. (2014) required *a priori* classification of resistance values to the landscape data which would have limited the parameter space explored in their study and may have resulted in sub-optimal resistance values being tested.

Stanton et al. (2014) also used least-cost pathways to explore the effects of seven landscape variables on otter geneflow in Scotland but found no significant correlations between resistance distance and genetic distance once geographic distance was taken into account. This is somewhat surprising, given that arable and horticultural land, elevation and built-up land were three of the variables tested which overlap considerably with landscape variables found to be significant to otter gene flow in the present study. In addition to the two methodological reasons discussed above, geographic differences in spatial structuring may have also contributed to the differences in results. Otters in northern Britain (including the Scottish population) were found to have lower levels of genetic structuring and differentiation than those in southern Britain (including Wales) which may have limited the power for Stanton et al. (2014) to detect landscape effects on gene flow.

3.4.2 Single and multiple-surface models

Both the single and multi-surface models placed habitat in the top models, with a degree of consistency in the optimised resistance values attributed to the lowest and highest resistance classes but not for the middle resistance classes. This may be due to a relatively small percentage

of the study area falling into some of the habitat classes making it harder for the genetic algorithm to optimise them (W. Peterman, Pers. Comms.).

Our inclusion of all landscape variables into a multi-surface optimisation model showed that all variables featured in models which were ranked above geographic distance, despite three of those variables (water, slope and roads) having ranked below geographic distance when tested as single surfaces, at both grain scales. Not only this, but water was included in one of the top three models (Habitat + Water), emphasising that landscape variables are unlikely to be acting in isolation when shaping the functional connectivity of a population. This strongly emphasises the importance of considering multiple surfaces, as opposed to approaches which only perform single-surface optimisations (Kozakiewicz et al. 2020) or have proposed eliminating landscape variables after single-surface optimisation if they do not perform better than geographic distance alone (Kimming et al. 2020).

3.4.3 Grain scale of landscape variables

We have only just begun to understand how the spatial scales of all the data used in landscape genetic studies (i.e. sampling of individuals, grain size of landscape data, extent of landscape data etc) might affect the interpretation of how landscape features influence spatial genetic structure through their degree of resistance to geneflow (Anderson et al. 2010). In our single surface models, comparison of two grain sizes an order of magnitude apart (100 m^2 and 1 km^2) provided the same overall conclusion regarding which individual landscape variables best explained the patterns of spatial genetic structure found in otters in the Wales and Borders area. Importantly however, due to computational restrictions we were unable to establish whether the same was true for the multi-surface models. Given the linear characteristics of two of the landscape variables, these are more likely to differ between the two scales. For example, at the coarser scale (1 km^2), water covered much of the available land area across the Wales and Borders region due to the density of the river network, whereas the linearity of the river network remained much more intact at the fine scale (100 m^2) with a higher proportion of pixels classified as ‘other land’ (Figure 3.1).

The use of multiple grain scales to analyse the same genetic dataset in landscape genetics is still uncommon (Anderson et al. 2010) and thus the current study provides new information to the body of work on the influence of spatial scales.

3.4.4 Time-lags and landscape genetic analysis

Another important factor to consider in the interpretation of landscape genetics results are genetic time-lags i.e. the difference between the time at which a landscape change occurs and when a population genetic response to that change becomes detectable (Bollinger et al. 2014;

Epps and Keyghobadi 2015). It is important to consider how long landscape features have existed in their current states, in order to determine the likelihood that genetic variation has reached equilibrium, and thus disentangle historic versus contemporary effects (Anderson et al. 2010; Epps and Keyhobadi 2015). Such time lags were suggested by Hohnen et al. (2016) as a possible explanation for the inverse relationship found between terrain ruggedness and genetic distance in northern quoll compared with their hypothesis. As one of the many factors influencing landscape genetic lag-times, dispersal rates and distances have been shown to be the most important variables of those studied so far (Epps et al. 2015). Otters are readily able to disperse over large distances (Green et al. 1984) and thus should exhibit reduced lag-time.

Landscape effects on spatial genetic structure have been shown to be detectable after about five generations have passed (Murphy et al. 2008; Cushman and Landguth 2010), this would equate to around 15 years in otters using the stand quoted generation time of 3 years (Randi et al. 2003). In this study elevation, slope and most river courses are unlikely to have changed considerably over very long (geological) timescales and therefore estimates of their importance are unlikely to be significantly affected by genetic time-lags. However, the water quality of rivers within the study area has improved over a much more recent time period, especially in areas with higher proportions of arable land (Mason and MacDonald 2004) although the effects of water quality on otter gene flow were not explicitly tested in this study. Roads and habitat are likely to have changed more across the study area within the relevant (recent) timeframe due to anthropogenic factors such as increased traffic and changes to land use. By 2006 road traffic in Great Britain had increased 30% since 1993 and 100% since 1978 respectively (Department of Transport 2020), indicating a significant increase in traffic over both the study period and the potential lag-time. As such the significance of roads as a barrier to gene flow may increase in future and should be considered in future studies.

3.4.5 Future directions

Eurasian otters have the largest global distribution of any otter species, stretching from Western Europe to East Asia, across which they occupy a wide range of environments (Mason and MacDonald 1986). While there is some corroboration between our results and previous studies looking at how landscape affects genetic structure (Mucci et al. 2003; Janssens et al. 2008) and otter presence across Europe (Robitaille and Laurence 2002; Barbosa et al. 2003), given the different methodologies used, repeating the analysis across different areas, especially those with differing landscape would determine whether our findings are location specific or more robust indicators of general facilitators and impediments to gene flow across the otter range.

In addition to replicated studies in other areas of the range, further work on both spatial and temporal scales of analysis would help enhance understanding of their effects on the results of

landscape genetic studies more generally (Anderson et al. 2010). As would further studies on the behaviour and dispersal of otters to determine why certain habitat types pose higher resistances than others.

3.5 Conclusions

Landscape genetic methods can provide valuable insights into predictors of functional connectivity. For otters in the Wales and Borders region landscape variables significantly affecting gene flow included habitat as well as elevation and water in combination with habitat. The use of RESISTANCEGA, which requires no *a priori* assignment of resistance values to individual classes, to optimise resistance surfaces for each landscape variable meant that relative resistance of the classes could be varied in innumerable combinations until the best combination to match the genetic data was found. This represents a much deeper exploration of the parameter space than is afforded through *a priori* classification of resistance values. Our results show that higher elevations along with habitats often found at higher elevations in Wales (semi-natural grassland and coniferous forest) and urban areas impede gene flow in otters, while broadleaf woodland, arable areas, rivers and coastline facilitate otter gene flow. These results could help identify areas that otters are less likely to disperse across and which therefore present barriers to gene flow. This information can inform planning of conservation measures to facilitate further recolonisation of Europe by otters, as well as providing data on what habitats might facilitate gene flow between previously separated populations.

Supplementary

S3.1 – Fine scale resistance estimates for the 11 habitat categories for both RESISTANCEGA runs. Habitat class, the aggregate classes as defined by the Land Cover Map 2007 (see Table 1); Total area (%), the percentage of the total raster pixels assigned to each habitat class; Res 1 and 2, the optimised resistance values for each class from the first and second run respectively; Rank 1 and 2, relative rank (ascending) based on optimised resistance values from each run; ave rank, the average rank of each habitat class across both runs. Bold indicates habitat classes with consistent rank across both iterations.

Habitat Class	Total area (%)	Res 1	Res 2	Rank 1	Rank 2	Ave rank
Broadleaf woodland	4.1	1	1	1	1	1
Arable	15.3	157.7	186.8	2	2	2
Saltwater	0.4	347.9	636.9	3	3	3
Improved grassland	24.1	819.3	1008.0	5	4	4.5
Coastal	1.0	744.5	1317.6	4	6	5
Mountain, heath, bog	3.6	1238.3	1277.2	7	5	6
Freshwater	0.4	895.0	1777.0	6	8	7
Sea	32.8	1246.4	1388.5	8	7	7.5
Semi-natural grassland	10.1	1611.4	1926.0	9	9	9
Urban	5.6	1637.0	1964.4	10	10	10
Coniferous woodland	2.8	2051.8	2206.4	11	11	11

S3.2 – Coarse scale resistance estimates for the 11 habitat categories for both RESISTANCEGA runs. Habitat class, the aggregate classes as defined by the Land Cover Map 2007 (see Table 1); Total area (%), the percentage of the total raster pixels assigned to each habitat class; Res 1 and 2, the optimised resistance values for each class from the first and second run respectively; Rank 1 and 2, relative rank (ascending) based on optimised resistance values from each run; ave rank, the average rank of each habitat class across both runs. Bold indicates habitat classes with consistent rank across both iterations.

Habitat Class	Total area (%)	Res 1	Res 2	Rank 1	Rank 2	Ave rank
Broadleaf woodland	1.2	1	1	1	1	1
Arable	16.5	89.3	113.4	2	2	2
Mountain, heath, bog	3.0	399.4	677.3	4	3	3.5
Improved grassland	29.1	684.5	802.8	5	4	4.5
Saltwater	0.3	139.9	1038.2	3	7	5
Sea	32.9	982.6	948.0	6	5	5.5
Freshwater	0.2	1744.6	1022.9	9	6	7.5
Coniferous woodland	2.4	1542.3	2125.4	7	9	8
Coastal	1.0	2047.8	2089.2	10	8	9
Urban	6.0	1666.6	2286.7	8	10	9
Semi-natural grassland	7.5	2148.6	2316.0	11	11	11

S3.3 – Percentage contribution of each landscape variable to the top multi-surface models across both iterations (Habitat only model not included as single-surface models have no split contributions). Model, the combination of variables making up the multi-surface model in question; Variable, each individual landscape variable; Iteration, indicates whether the contributions shown are from the initial or repeat run. NA is used when the variable is not included in the specific model.

Model \ Variable	Habitat	Elevation	Water	Iteration
Habitat + Elevation	91%	9%	*NA*	1
Habitat + Elevation	86%	14%	*NA*	2
Habitat + Water	63%	*NA*	37%	1
Habitat + Water	64%	*NA*	36%	2

S3.4 – Optimised resistance values for the Habitat + Elevation multi-surface model across both iterations. Class, the individual landscape class for each of the landscape variables (see Table 1); Res 1 and 2, the optimised resistance values for each class from the first and second run respectively; Rank 1 and 2, relative rank (ascending) based on optimised resistance values from each run; Ave rank, the average rank of each habitat class across both runs. All ranks were consistent across both iterations. Note in the optimised layer resistances are combined across both landscape variables using the contribution weights shown in S3.3.

Class	**Res 1**	**Res 2**	**Rank 1**	**Rank 2**	**Ave rank**
(a) Elevation					
0 – 300 m	1	1	1	1	1
Sea	55.2	301.6	2	2	2
300 – 600 m	1257.2	1434.8	3	3	3
600+ m	2129.3	1530.7	4	4	4
(b) Habitat					
Broadleaf woodland	1	1	1	1	1
Saltwater	1.2	3.6	2	2	2
Arable	70.1	79.2	3	3	3
Mountain, heath, bog	343.6	481.4	4	4	4
Freshwater	459.6	595.6	5	5	5
Improved grassland	610.0	639.1	6	6	6
Semi-natural grassland	1148.9	1108.1	7	7	7
Sea	1260.6	1163.3	8	8	8
Coniferous woodland	1736.7	1575.3	9	9	9
Urban	2298.5	2273.6	10	10	10
Coastal	2372.4	2481.1	11	11	11

S3.5 – Optimised resistance values for the Habitat + Water multi-surface model across both iterations. Class, the individual landscape class for each of the landscape variables (see Table 1); Res 1 and 2, the optimised resistance values for each class from the first and second run respectively; Rank 1 and 2, relative rank (ascending) based on optimised resistance values from each run; Ave rank, the average rank of each habitat class across both runs. Bold indicates habitat classes with consistent rank across both iterations. Note in the optimised layer resistances are combined across both landscape variables using the contribution weights shown in S3.3.

Class	Res 1	Res 2	Rank 1	Rank 2	Ave rank
(a) Water					
Water	1	1	1	1	1
Land	144.9	70.8	2	2	2
Sea	1445.2	1494.2	3	3	3
(b) Habitat					
Broadleaf woodland	1	1	1	1	1
Sea	1.1	1.9	2	2	2
Arable	49.8	52.8	3	4	3.5
Saltwater	437.4	26.8	4	3	3.5
Mountain, heath, bog	640.3	656.0	5	5	5
Improved grassland	723.2	781.1	7	6	6.5
Freshwater	718.6	873.1	6	7	6.5
Semi-natural grassland	1926.5	2078.0	9	9	9
Coastal	1912.3	2246.2	8	11	9.5
Urban	2001.3	2001.4	11	8	9.5
Coniferous Woodland	1977.2	2125.0	10	10	10

S3.6 – Bootstrapping results for all multi-surface coarse-scale landscape variables, initial run.

Surface	ΔAICc	avg.weight	avg.rank	avg.R2m	Percent.top	k
Habitat	0.00	0.437	2.5	0.176	34.5	12
Elevation+Habitat	0.59	0.369	1.869	0.178	49.5	16
Habitat+Water	3.13	0.138	3.272	0.176	12.5	15
Elevation+Habitat+Water	8.66	0.016	4.698	0.179	0.6	19
Habitat+Slope	10.46	0.003	6.427	0.182	0	17
Elevation+Habitat+Slope	13.81	0.002	6.934	0.172	0	21
Elevation+Habitat+Roads	14.74	0.004	6.415	0.196	1	23
Habitat+Roads	15.37	0.002	8.688	0.190	0	19
Habitat+Slope+Water	18.46	<0.001	10.387	0.173	0	20
Elevation+Habitat+Slope+Water	23.20	<0.001	11.353	0.173	0	24
Elevation+Habitat+Roads+Water	23.22	<0.001	10.105	0.184	0	26
Habitat+Roads+Water	24.62	<0.001	13.166	0.202	0	22
Elevation+Roads	25.18	0.013	13.474	0.170	1	12
Habitat+Roads+Slope	28.78	<0.001	14.696	0.196	0	24
Elevation+Habitat+Roads+Slope	29.52	<0.001	12.996	0.186	0	28
Elevation+Roads+Slope	32.70	<0.001	16.004	0.164	0	17
Elevation+Roads+Water	32.87	<0.001	16.776	0.163	0	15
Habitat+Roads+Slope+Water	36.17	<0.001	16.921	0.197	0	27
Elevation+Slope	37.03	0.006	17.592	0.125	0.4	10
Elevation+Roads+Slope+Water	40.92	<0.001	18.989	0.166	0	20
Elevation+Water	52.71	0.007	20.307	0.117	0.4	8
Elevation	56.67	0.002	21.767	0.137	0.1	5
Distance	60.50	<0.001	22.339	0.047	0	2
Elevation+Slope+Water	65.38	<0.001	23.224	0.112	0	13
Roads+Slope	83.60	<0.001	24.804	0.141	0	13
Roads+Slope+Water	90.96	<0.001	26.135	0.145	0	16
Roads	102.98	<0.001	27.505	0.0908	0	8
Water	107.94	<0.001	28.406	0.090	0	4
Slope+Water	109.53	<0.001	28.375	0.110	0	9
Roads+Water	110.07	<0.001	28.888	0.092	0	11
Slope	237.54	<0.001	30.988	0.324	0	6

S3.7 - Bootstrapping results for all multi-surface coarse-scale landscape variables, repeat run.

Surface	ΔAICc	avg.weight	avg.rank	avg.R2m	Percent.top	k
Habitat	0	0.394	2.946	0.174	30.4	12
Elevation+Habitat	0.12	0.351	1.88	0.173	46.2	16
Habitat+Water	1.58	0.203	2.824	0.179	20.2	15
Elevation+Habitat+Water	8.37	0.008	4.522	0.177	0	19
Habitat+Slope	10.42	0.002	6.574	0.178	0	17
Elevation+Habitat+Roads	13.20	0.003	5.649	0.190	0.6	23
Elevation+Habitat+Slope	14.07	0.001	7.245	0.178	0	21
Habitat+Roads	18.27	<0.001	10.614	0.174	0	19
Habitat+Slope+Water	20.31	<0.001	11.81	0.174	0	20
Elevation+Habitat+Slope+Water	21.44	<0.001	10.573	0.174	0	24
Habitat+Roads+Water	22.24	<0.001	12.039	0.183	0	22
Elevation+Habitat+Roads+Water	22.83	<0.001	10.197	0.186	0	26
Elevation+Roads	23.94	0.013	13.793	0.168	0.8	12
Habitat+Roads+Slope	28.30	<0.001	15.229	0.186	0	24
Elevation+Habitat+Roads+Slope	28.72	<0.001	12.884	0.186	0	28
Elevation+Roads+Slope	31.78	<0.001	16.569	0.168	0	17
Elevation+Roads+Water	32.22	<0.001	17.54	0.163	0	15
Elevation+Slope	34.18	0.003	18.05	0.118	0.3	10
Habitat+Roads+Slope+Water	36.58	<0.001	17.991	0.190	0	27
Elevation+Habitat+Roads+Slope+Water	39.01	<0.001	16.773	0.190	0	31
Elevation+Roads+Slope+Water	39.65	<0.001	19.804	0.167	0	20
Elevation+Water	47.00	0.01	20.503	0.120	0.6	8
Elevation	53.07	0.01	22.213	0.122	0.8	5
Distance	57.91	0.002	23.309	0.047	0.1	2
Elevation+Slope+Water	64.48	<0.001	24.417	0.140	0	13
Roads+Slope	82.01	<0.001	25.932	0.141	0	13
Roads+Slope+Water	89.89	<0.001	27.373	0.142	0	16
Slope+Water	100.98	<0.001	28.814	0.105	0	9
Roads	101.12	<0.001	29.067	0.086	0	8
Water	102.56	<0.001	29.374	0.090	0	4
Roads+Water	103.47	<0.001	29.492	0.092	0	11
Slope	410.78	<0.001	32	0.335	0	6

4

Detecting changes in effective population size in an expanding population with known recent history

Abstract

Census (N_C) and effective population size (N_e) are important parameters that are used to assess the conservation status of endangered populations. Small N_C and N_e result in higher susceptibility to stochastic events and greater chance of extinction, while populations with a large N_e are assumed to possess increased evolutionary potential due to maintenance/accumulation of adaptive genetic variability. Past population dynamics, such as population bottlenecks or persistent fluctuations in population size, can have long lasting effects on N_e even when N_C is seemingly large or recovered. The otter population in the UK has undergone a well-documented population recovery over recent decades, with the presence of spraint increasing over successive national surveys. To determine whether this assumed increase in population size has led to recovery at the genetic level we used a large-scale genetic dataset (1993-2014, N = 407) to detect signals of recent changes in population size using BOTTLENECK and estimated N_e using the linkage-disequilibrium method with LDNe. We explored the biases in both analyses driven by unaccounted for genetic structure and the inclusion of admixed individuals and determined whether current population size estimates for otters in Wales and England, estimated from survey data, are supported by the genetic estimates of population size in this study. We found evidence of recent change in population size, that reflect changes in occupancy suggested by national surveys. The estimates of N_e across all regions were small and well below the sizes required for long-term population viability. The low N_e estimates reported in our study indicate the importance of including genetic monitoring of species in national monitoring plans: importantly, the conclusions drawn from successive national surveys using otter sign (e.g. spraint and footprints), namely of a robust population close to panmixia, are not supported by the genetic evidence.

4.1 Introduction

4.1.1 Population demographics

Understanding and quantifying past and current population dynamics is one of the key goals of many conservation studies, and genetic data can provide an important perspective in this analysis, especially in the absence of other data. Genetic studies often include the estimation of effective population size (N_e) and the detection of population bottlenecks. Populations that have

been through a bottleneck or a significant reduction in size are more susceptible to adverse extrinsic events (Cristescu et al. 2010), likewise, populations with low N_e have intrinsically reduced genetic diversity and adaptive potential (Palstra and Ruzzante 2008). These potential reductions in population viability make the estimation of past demographic parameters important to wildlife managers and conservationists.

Whereas census population size (N_C) describes the total number of individuals in the population, N_e estimates the number of individuals within the census population who are contributing offspring to the next generation (Charlesworth 2009), with N_e often being smaller than N_C in wild populations, e.g. due to variation in breeding success among individuals and sexes (Hoban et al. 2021a). N_e is reflective of the evolutionary potential of a population, along with its susceptibility to stochastic processes and therefore its probability of survival, which is often more relevant to conservation and management than N_C (Frankham 1995b). Estimating N_e from demographic data is complex and requires information on the reproductive success of individuals (Leberg 2005), however N_e may also be directly estimated from genetic data, where it is defined as the size of an idealised population (i.e. one at mutation-drift equilibrium) that would experience the same loss of genetic diversity, through random genetic drift, as the population in question (Wright 1931).

Population bottlenecks reduce both N_C and N_e, and in the absence of immigration are expected to decrease the genetic diversity of the population. Reductions in N_e can take a significantly longer time to recover than N_C, and sequential bottlenecks or fluctuations in population size may lead to situations where - despite having a large contemporary overall size (N_C) - a population remains at risk due to persistent low N_e (Frankham 1995b).

Following major declines in the 19th and 20th centuries, population recoveries, especially of large carnivores, have been occurring in the 21st century as anthropogenic pressures have been eased through legal protections (Chapron *et al*. 2014). While these population expansions are positive for conservation, it is important that census population increases in isolation are not regarded as indicating a successful recovery. Genetic monitoring allows us to estimate contemporary N_e (which is a complementary indicator of population viability to N_C) as well as determine the severity of past population bottlenecks. Both elements provide information that is vital to the development of effective conservation management plans.

4.1.2 Detecting population size changes using genetic data

There are several methods of estimating past population size change using genetic data. One of the most commonly used methods is implemented in the software BOTTLENECK, a moment-based method which uses genetic data summary statistics to detect departures from theoretical distributions (Cornuet and Luikart 1996) and has been used by a wide range of studies (over 2900

citations to date). These departures are transient and therefore it is designed to detect recent changes in population size. Other approaches such as that implemented using MSVAR (Beaumont 1999), which uses likelihood-based methods coupled with Monte Carlo integration, or ABCtoolbox (Wegmann et al. 2010), which uses Approximate Bayesian Computation (ABC), are based on coalescent theory (Kingsman 1982) and therefore better at detecting more historic populations size changes.

For example, a simulation study by Girod et al. (2011) indicated that MSVAR was very efficient at detecting demographic changes that were both ancient (greater than 50 generations ago) and/or severe (a 100-fold change in population size for either contractions or expansions), recent declines (within the last 10 generations) went largely undetected except for very severe population contractions (1000-fold change in population size, e.g. Goossens et al. 2006). This is not surprising given that for methods based on coalescent theory (i.e. MSVAR and ABCtoolbox) very recent demographic events are unlikely to have detectably altered the gene genealogy making them harder to distinguish from those expected in a stable population with the same size as the ancestral one. BOTTLENECK is therefore a more suitable approach for datasets where recent or contemporary population history is being investigated.

During a population bottleneck event allelic diversity is reduced more quickly than heterozygosity as rare alleles are likely to be lost from the population as it rapidly declines, effectively sub-sampling the available genetic diversity (Nei 1987; Maruyama and Fuerst 1985). This results in a transient deficiency in the number of alleles found in a given sample of individuals given the observed heterozygosity at a locus. This deficiency can be used to detect population size change from the genetic data, i.e. the observed number of alleles in a sample is less than the expected number of alleles based on the observed heterozygosity for the same sample but with the assumptions that the population is at mutation drift equilibrium (where the rate of loss of alleles due to genetic drift is equal to the rate of replacement through mutation) and the loci are selectively neutral. Testing for an allele deficiency via the relationship between observed heterozygosity and expected number of alleles is complex and can be biased (Zouros 1979). Therefore, for example, BOTTLENECK (Cournet and Luikart 1996; Piry et al. 1999) instead uses four tests designed to detect differences between the observed heterozygosity of the sample, and the expected observed heterozygosity estimated from simulated populations under mutation drift equilibrium with the same number of alleles as the observed number of alleles in the sample.

Comparisons of observed and expected observed heterozygosity can help detect a recent change in the effective population size of the sample, with a heterozygosity excess (i.e. observed heterozygosity > expected observed heterozygosity) indicating a recent bottleneck (Cornuet and Luikart 1996). Although BOTTLENECK and the tests within were initially designed specifically to

detect reductions in the effective population size of a sample (i.e. population bottlenecks) it was also shown that using the same methods, a heterozygosity deficiency (i.e. observed heterozygosity < expected observed heterozygosity) can indicate recent population expansion, with empirical data from the Sardinian human population, which has been undergoing a long period of population growth without immigration (Di Rienzo et al. 1994) used to confirm this result (Cornuet and Luikart 1996). The method was also used to detect population expansions in two free-living mosquito species (*Anopheles gambiae* and *Anopheles arabiensis*) in Kenya (Donnelly et al. 2001). Heterozygosity excess or deficit at a single locus might be a sign of selective advantage or disadvantage (Watterson 1978) (rather than indicating changes in effective population size), BOTTLENECK averages the deviations found across all loci and therefore is robust as long as a sufficiently large set of loci are used (10 – 15, Luikart *et al.* 1998c).

This approach relies on several assumptions including that each sample is representative of a well-defined population i.e. no immigration, no population structure and selectively neutral loci. The presence of immigration can be particularly problematic, especially if this is from a genetically divergent population, as it is likely to increase the number of rare alleles in the dataset without measurably altering the heterozygosity. This increase in number of alleles will effectively emulate a signal of population expansion and could also mask the presence of a population bottleneck in the dataset. A similar bias can be introduced by including individuals with admixed genotypes in the dataset.

These approaches to detecting population bottlenecks are most likely to detect severe population declines (or large expansions) as the likelihood of a significant result is related to the magnitude of the difference between the observed heterozygosity and the expected observed heterozygosity. Therefore, smaller or less dramatic changes in population size may go undetected. However, given that from a conservation perspective, populations that have gone through severe bottlenecks are likely to be the ones most needing ongoing monitoring or intervention, the results are nonetheless useful in prioritising populations or species of concern. Additionally, a severe demographic bottleneck (reduction in census population size) does not always mean a severe genetic bottleneck (reduction in effective population size), especially if the population has a history of low effective population size (N_e) due to historical fluctuations in size (Frankham 1995b). As such, analysis of genetic markers can help disentangle whether an observed reduction of census population size has had a negative effect on the population at the genetic level or not.

4.1.3 Population size estimation

Effective population size (N_e) is a key determinant of population survival. If N_e is very low then no matter how large N_C may be, the population remains vulnerable to stochastic events and therefore extinction. For some species, N_C is relatively easy to determine via direct observation of

a population, but until relatively recently N_e was much harder to calculate, as detailed data on breeding success are required for its estimation from demographic models (Leberg 2005). The rapid development of genetic approaches in recent decades means that N_e can nowadays be directly estimated from genetic data (Harris and Allendorf 1989; Luikart et al. 2010; Palstra and Fraser 2012; Hoban et al 2021a) with the caveat that different methods can result in different estimates due to varying assumptions and confidence intervals can be large. For species that are elusive or live at low densities, direct observation of N_C is problematic, and estimates of N_e using genetic data are now more accessible. Therefore, a ratio is often applied to translate estimates of N_e from genetic data into estimates of N_C (Frankham 1995b). Across studies, the modal estimate of N_C has been found to be approximately 10-11x larger than N_e (Frankham 1995b; Hoban et al. 2021a) but can range from 10^{-6} in Pacific oysters (*Crassostrea gigas*) to 0.994 in humans, however this range is reduced when only datasets with comprehensive data are used.

Estimating N_e from genetic data has mainly been achieved by using, where possible, two-sample or temporal methods which used data taken at two points in time, preferably multiple generations apart to detect changes in allele frequencies caused by genetic drift and thus produce an estimate for N_e. These two-sample estimators have employed a number of methods including; temporal F-statistics (Ne-estimator, Do et al. 2014 and TempoFs, Jorde and Ryman 2007), pseudomaximum likelihood methods (MLNE, Wang 2001) and coalescent-based Bayesian methods (TM3, Berthier et al. 2002 and CoNe, Anderson 2005). However, this requirement for two sets of genetic data, generations apart, can be problematic for species or populations that are not routinely monitored, or which have long generation times. Consequently, a set of methods that require data from only one time point, known as one-sample estimators, have been developed. These estimators take a variety of different approaches to estimating N_e including; approximate Bayesian computation (ONeSAMP, Tallmon et al. 2008), sibship assignment or parentage (Colony2, Wang 2009 and AgeStruct, Wang et al. 2010) and linkage-disequilibrium (LDNe, Waples and Do 2008). Both one-sample and two-sample estimators of N_e assume discrete generations, however, which can be problematic in many sampling regimes and for species which show temporally overlapping generations. Recent studies into the relative performance of various methods of N_e estimation showed that LDNe is a robust single sample estimator (Gilbert and Whitlock 2015). There have also been recent advances and studies into the use of LDNe with overlapping generations (Waples et al. 2014).

4.1.4 The influence of genetic structure on estimates

A plethora of population genetic analyses can be influenced by unaccounted genetic structure within a dataset leading to erroneous signals or estimates, due to the common assumption of an idealised population such as the Wright-Fisher model (Fisher 1930; Wright 1931). Likewise,

bottleneck detection and N_e estimates can also be biased by unaccounted population structure (Chikhi et al. 2010; Kopatz et al. 2017).

A subdivided population will exhibit an excess of rare alleles due to structuring and non-random mating. If unaccounted for, then bottleneck analysis will detect this excess of rare alleles as a signal of recent population expansion (Cournet and Luikart 1996). Admixed individuals between genetically differentiated populations may also produce a similar bias. Similarly, unaccounted genetic structure has been shown to cause large downward biases in the estimation of N_e using the linkage disequilibrium method (Kopatz et al. 2017). This is due to the effect that non-random mating (i.e. due to population sub-division) exerts on the extent of LD in the population through non-random allele distributions across loci (England et al. 2010; Waples and England 2011). This results in cases where global N_e estimates are considerably lower than the sum of subpopulation N_e estimates, although large-scale empirical studies of this interaction are still relatively rare (Kopatz et al 2017). A previous study on the recovering population of brown bears (*Ursus arctos*) in Finland illustrated that as well as genetic structure, the presence of admixed individuals in the dataset was a second important factor in LDNe estimates. Kopatz et al. (2017) found that including admixed individuals strongly increased N_e estimates and suggested more work was needed on their potential high influence in estimating N_e as well as number of breeders (N_b) through further studies in other species and systems. It follows, therefore, that understanding the genetic structure present in a dataset is important to help ensure estimates are unbiased.

4.1.5 Otter population history in the UK and national survey data as evidence of otter population growth

The persecution of otters in the UK likely began as far back as the middle ages (Lovegrove 2007). Historic records indicate a steady decline in numbers from the 18th century onwards due to anthropogenic predator control, sport hunting and pollution (Jefferies 1989). However, it was not until the 1950's that hunting records showed a sudden and rapid decline in otter numbers, with southern England the most severely affected area. The decline was parallel to that seen in predatory bird populations which suggested that dieldrin, along with other organochlorine chemicals, was the cause (Chanin and Jefferies 1978).

Dieldrin was introduced in the 1950's as a sheep dip and seed coating and was subsequently detected in 81% of otters examined between 1963-1973 (Mason *et al* 1986). Voluntary restrictions were placed on the chemical in the 1960's-1970's followed by a mandatory ban in the 1980's. As a response to the dramatic population decline in otters, systematic national surveys were set up in Wales, England and Scotland, with the first undertaken in the late 1970's (Crawford *et* al. 1979; Green and Green 1980; Lenton *et al*. 1980).

Successive national surveys for otters in both Wales and England have shown a steady increase in detection of positive sign for otters at survey sites (Crawford 2010; Strachan 2015 – see Figure 2.1). However, although more frequent and spatially widespread detection of sign, such as otter spraint, indicates that otters have now returned to previously extirpated areas, it is impossible to estimate the change in population size with any degree of certainty (Sainsbury *et al.* 2018; Matthews *et al*. 2018). Despite the recent recolonisation, significant and persistent genetic structure is observed in otter populations (Hobbs et al. 2011; Stanton et al. 2014 and Chapter 2) suggesting that re-establishment of contact between previously isolated sub-populations has not yet resulted in homogenisation. Estimation of population size from existing monitoring data requires the application of broad assumptions, such as the consistent density of otters per km of river across a broad spatial scale (Matthews et al. 2018). Such assumptions are unlikely to be appropriate given the different degree and rate of population decline and subsequent recovery across Wales and England, along with the heterogeneous nature of the landscape and prey availability. As such, estimation of current or past population sizes from national survey data should be treated with extreme caution.

The unusually well-known history of the otter population in the UK, and the availability of data describing population structure provide this project with the information necessary to evaluate the detection of recent bottlenecks and to estimate effective population size. The rich dataset also provides the opportunity to explicitly test the biases that unaccounted for population structure and admixture are likely to present to both estimates.

Hypotheses to be tested in this chapter fall into two broad areas – on bottleneck detection itself and on N_e estimation:

Bottleneck detection

- When testing for a bottleneck without accounting for the strong regional sub-structuring in the dataset, significant changes in N_e will be detected, with a sign of population expansion due to an excess of rare alleles. Conversely, when regional sub-populations are tested independently, bottleneck tests are unlikely to detect population expansions in all regions.
- When regional sub-populations are tested independently, the populations most likely to exhibit population expansions are those in the South West and Wales and Borders regions as they represent population strong-hold areas and had the fastest recoveries. The region most likely to show significant signal of a bottleneck is in Eastern England, as it exhibited the most severe population decline and has been the slowest to recover.
- The detection of population bottlenecks will be more likely to be significant when admixed individuals are removed from the regional sub-populations due to the removal of

mixed genotype individuals (i.e. alleles of inter-regional origin), whereas the detection of populations expansions will be less likely to be significant when admixed individuals are removed from the regional sub-populations.

Effective population size estimation

- The estimate of N_e for the whole dataset will be less than the sum of the estimates of N_e for the regional sub-populations due to the presence of mixture LD in the first dataset but not the second.
- The inclusion of admixed individuals will upwardly bias Ne estimates, based on the observations of Kopatz *et al.* (2017).
- Temporally replicated estimates of N_e over time will show larger N_e at the later time point for all regions and clusters due to relative increasing presence of otter sign across all areas of Wales and England over the study period.

4.2 Methods

4.2.1 Samples, genotyping and dataset production

We used a georeferenced dataset of 407 muscle tissue samples from predominantly road-killed otters held in the Cardiff University Otter Project archive. Host DNA was extracted and genotyped using 15 microsatellite loci (see methods in Chapter 2 for details). The samples were collected between 1993 and 2014 from across the United Kingdom, although sampling was sparse and temporally restricted in Scotland and Northern Ireland meaning some analysis was only conducted on samples from Wales and England.

Admixed individuals in the dataset were identified using a membership value $q<0.8$ from the STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) outputs obtained in Chapter 2. The cluster assignment at K=3 was chosen for this purpose as it had strong support using a combination of the Evanno method (Evanno et al. 2005), likelihood of K (Pritchard & Wen 2003), was biologically plausible (Janes et al. 2017) and from the results of Chapter 2 best captured the overarching genetic structure in the population (Figure 4.1). This allowed the production of datasets at various spatial and temporal scales, as well as allowing analysis with admixed individuals included and excluded to test the assumptions and biases of each analysis.

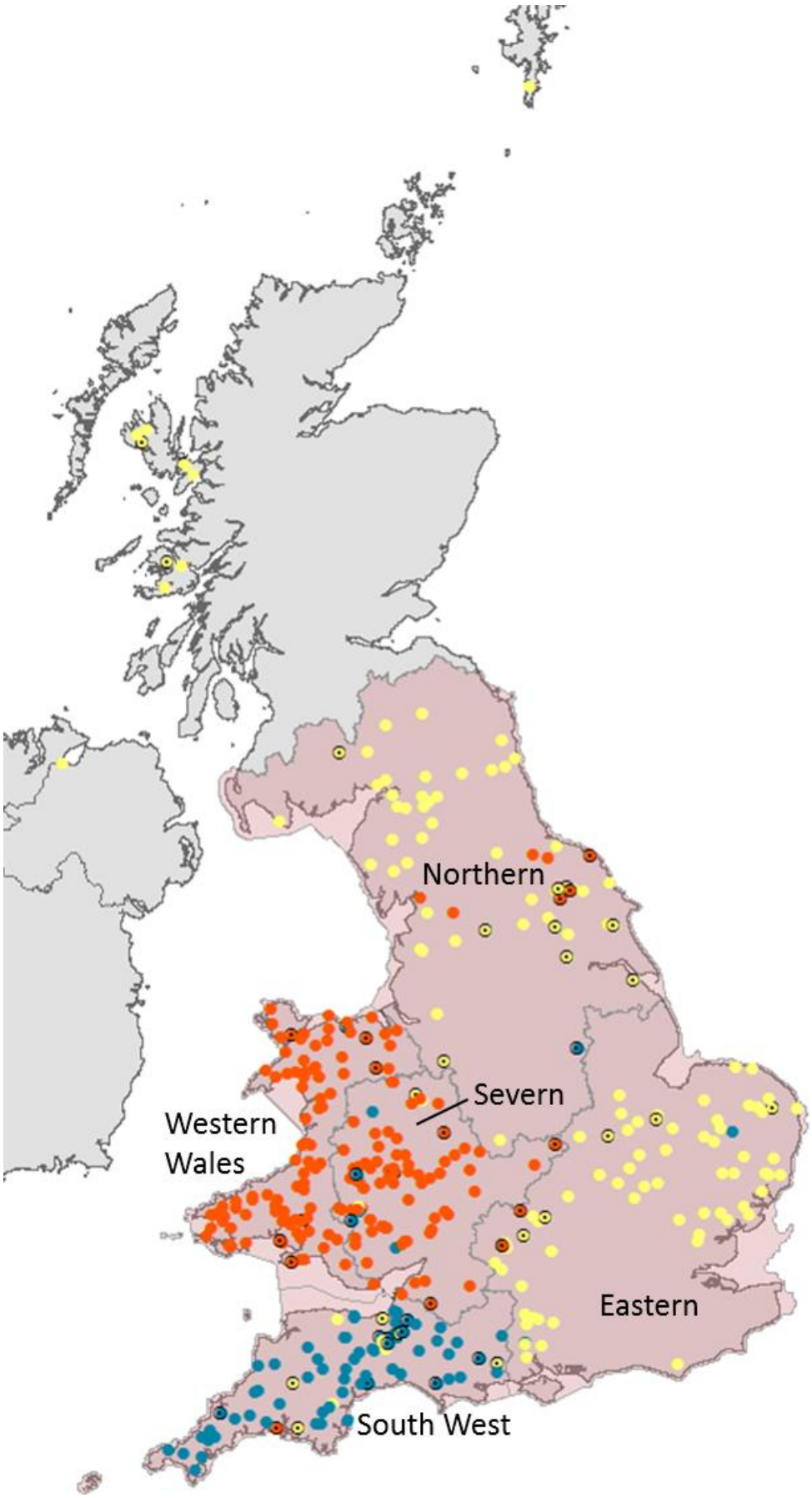


Figure 4.1: Map of the United Kingdom (grey) showing the locations of the 407 individuals genotyped for this study. Assignment to a genetic cluster at K=3 is denoted by the following colours; orange – Wales and Borders, blue – Southwest, yellow – central and Northern England. A black outline with a black dot in the centre indicates that the individual has a membership value q<0.8 to the given cluster and is therefore considered 'admixed'. The five River Basin District regions are outlined, and shaded in red.

4.2.2 Population bottleneck analysis

To test for recent changes in effective population size in the dataset we used BOTTLENECK v1.2.02 (Piry, Luikart and Cornuet 1999) which uses allele frequency data to detect recent bottleneck or expansion events (Cornuet and Luikart 1996; Luikart *et al.* 1998). We used all four tests available in BOTTLENECK; the sign test, standardised differences test, Wilcoxon sign-rank test and the allele frequency distribution or mode shift indicator, but given that the sign test suffers from low statistical power and the standardised differences test requires at least 20 loci, we focused on the results from the Wilcoxon sign-rank test and the more qualitative allele frequency distribution. The Wilcoxon sign-rank test has been shown to have relatively high power in detecting population size changes, and although it can be used with as few as 4 polymorphic loci and any number of individuals, to achieve this high power of detection it is recommended to use 10-15 polymorphic loci and 15-40 individuals (Luikart *et al.* 1998c). All data combinations tested had 14 or 15 polymorphic loci, while all had more than 15 individuals and many had more than 40 individuals (22/28 datasets, with those with N<40 limited to the temporally restricted analysis), indicating that there should have been high power to detect population bottlenecks or expansions using this test.

Table 4.1: Frequency of stepwise mutation of microsatellites with different length repeat units. Frequency of stepwise mutation model (SMM) is the proportion of single-step mutations among single- and multiple-step microsatellite mutations surveyed by Ellegren 2000, Number of loci refers to the number of loci (of the total 15 used in this study) that are either dinucleotide or tetranucleotide, Freq SMM in multiplex – the estimated frequency of SMM in our multiplex based on the number of loci of different repeat units.

Microsatellite repeat unit	Frequency of SMM	Number of loci
Dinucleotide	83.8%	4
Tetranucleotide	89.5%	11
Freq SMM in multiplex	88%	15

We used data available on the mutation processes in human microsatellite DNA sequences (Ellegren 2000) to estimate the frequency of adherence to the stepwise mutation model for both dinucleotide and tetranucleotide microsatellites. The frequencies were estimated to be 83.8% and 89.5% for dinucleotide and tetranucleotide microsatellites respectively. We then applied this estimation to the specific multiplex of microsatellites used to genotype individual otters in this study, based on whether each locus had a dinucleotide or tetranucleotide repeat unit (**Table 4.*1***). This custom frequency was then used as input data for the Two-phased model of mutation (TPM) in BOTTLENECK. This model is considered more appropriate for microsatellite data than the Stepwise Mutation Model (SMM) or the Infinite Allele Model (IAM) as it is likely that mutation consists of mostly one-step mutations with occasional multi-step changes (Di Rienzo *et al.* 1994; Luikart *et al.* 1998b). Additionally, if an analysis is run using the strict SMM mutation model and loci deviate even slightly from this a bottleneck signal will be produced due to the excess

heterozygosity, and therefore this model can be biased towards producing bottleneck signals when using microsatellite data.

Each dataset was run twice for 1,000 iterations using the TPM with proportion of SMM set to 88% and the variance set as either 12 or 30. To assess the impact of both underlying genetic structure and admixed individuals on the detection of effective population size change using these methods, the dataset followed a hierarchical selection process (Table 4.2) where the presence of both genetic structure and admixed individuals in the dataset were systematically and sequentially accounted for. This process was repeated with just the most contemporary samples from 2009 and 2014, and also with early (1993-1995 and 1993-1999) and late (2009-2014 and 2014) datasets from the Wales and Borders (Western Wales and Severn RBD regions) where there was sufficient sampling to allow a temporal comparison (i.e. N>15).

4.2.3 Effective population size estimation

Effective population size (N_e) was estimated using the linkage-disequilibrium (LDNe) method (Waples and Do 2008) in NeEstimator v2.1 (Do *et al.* 2014). LDNe uses the Burrows method to estimate linkage-disequilibrium with a correction factor to account for using unlinked loci such as microsatellites (Waples 2006). We excluded rare alleles which can upwardly bias Ne estimates using the P_{crit} function (Waples and Do 2010) based on the formula $1/(2 \times N) < P_{crit} < 1/N$ which highlighted that different values of P_{crit} were appropriate for different datasets as sample size varied by at least an order of magnitude. Confidence intervals were determined using the jack-knife method which has been shown to perform better than parametric methods (Waples and Do 2008).

We estimated N_e both with and without accounting for both the underlying genetic structure in the dataset and the presence of admixed individuals. Unaccounted genetic structure can lead to a mixture LD (England *et al.* 2010; Waples and England 2011) and downwardly bias N_e by combining more than one gene pool in the analysis, while there is some evidence that admixed individuals may upwardly bias N_e estimates (Kopatz et al. 2017).

Estimates of N_e for RBD regions and genetic clusters were repeated using only data from 2009 and 2014 to calculate the most contemporary figures across Wales and England, both to compare with those calculated from the last national survey data (Sainsbury *et al.* 2018; Matthews *et al.* 2018) and to see if these differed significantly from estimates made using the whole temporal spread of the data. In addition, to these 'late' estimates of N_e, a set of 'early' estimates were computed for the genetic clusters and RBD regions including samples collected up to 2004. This allowed comparison of the estimated N_e at two different time points during the population expansion.

For comparison with population size estimates derived from national survey data a value for the census population size (N_C) was needed, not N_e as calculated by LDNe. Frankham (1995) conducted a wide-ranging review of N_e/N_C ratios using data from 192 species and determined that broadly N_e was likely to be 0.10-0.11 of N, therefore we used this ratio to put the effective N_e estimates in context of the national survey population estimates.

4.3 Results

4.3.1 Population Bottleneck Analysis

The results from BOTTLENECK from the whole dataset (1993 – 2014) when population genetic structure was ignored (i.e. with the dataset being analysed as one panmictic population) suggested that population expansion had occurred. Once the dataset was split geographically into the RBD regions to account for the genetic structure present, many of these signals of expansion disappeared (Table 4.2; for full results see Supplementary 4.1). The Eastern RBD region showed a significant bottleneck signal whether admixed individuals were included or not, while the Severn RBD region showed significant expansion. The South West RBD region had a significant signal of expansion when admixed individuals were included, which became marginal once they were excluded, while both Northern and Western Wales RBD regions showed only signatures of population stability. Analysis by genetic cluster indicated a significant expansion for the Wales and Borders cluster, while both the Central England and South West clusters gave signals consistent with stable populations whether admixed individuals were included or not. None of the datasets showed any mode shift in allele frequencies.

Analysis of the most contemporary data (from 2009 and 2014) showed broadly similar patterns to the full dataset (Table 4.3). The Eastern RBD region showed a weaker bottleneck signal, whereas the South West RBD region showed an increased signal of expansion and the Severn RBD region changed from an expansion to a stable population signal. Both Northern and Western Wales RBD regions continued to show signals of population stability (for full results see Supplementary 4.2).

The results from temporal sampling of the Wales and Borders region genetic cluster showed that both the earliest (1993-1995) and latest (2014) samples, respectively, showed no significant signal of either a bottleneck or population expansion, and thus the population at both timepoints was considered stable (Table 4.3). Given the small sample size at these two time points (N=25 and N=28, respectively) the analysis was repeated using a broader timescale of 1993-1999 and 2009-2014. In this second iteration, the earlier dataset gave no signal of recent population size change, but the later dataset had a significant signal for population expansion (for full results see Supplementary 4.3).

Table 4.2: Bottleneck results with and without accounting for genetic structure and admixture. Dataset, the geographic or genetic extent of the data; N, number of individuals included in the analysis; Mean He, the mean heterozygosity across samples in the dataset; Possible biases – genetic structure, Y indicates genetic structure present in the dataset, N indicates genetic structure accounted for; Possible biases – admixed individuals; Y indicates admixed individuals (based in data from 15 microsatellite loci and a STRUCUTRE q-value threshold of 0.8) were included in the dataset, N indicates admixed individuals excluded from the dataset; TPM (88%, 30) Wilcoxon p-value, the p-value for the relevant one-tailed Wilcoxon test using the Two-Phase Model of mutation with 88% stepwise mutations and variance of 30; TPM (88%, 12) Wilcoxon p-value as previously by with a variance of 12; Mode Shift, the distribution of allele frequencies; Detected Signal, population state according to the Wilcoxon test. N/A indicates p-value was non-significant ($p<0.05$), and underlined p-values show values approaching significance, where second model was significant.

Dataset	N	Mean He	Possible biases		TPM (88%, 30) Wilcoxon p-value	TPM (88%, 12) Wilcoxon p-value	Mode Shift	Detected Signal
			Genetic structure	Admixed individuals				
All Data	407	0.68	Y	Y	0.05	0.02	Normal L-shaped	Expansion
Wales and England	396	0.68	Y	Y	0.05	0.02	Normal L-shaped	Expansion
	347	0.68	Y	N	0.05	0.02	Normal L-shaped	Expansion
Eastern RBD Region	74	0.72	Y	Y	0.01	0.02	Normal L-shaped	Bottleneck
	64	0.71	N	N	0.01	0.01	Normal L-shaped	Bottleneck
Northern RBD Region	59	0.70	Y	Y	N/A	N/A	Normal L-shaped	Stable
	42	0.69	N	N	N/A	N/A	Normal L-shaped	Stable
South West RBD Region	77	0.57	Y	Y	0.05	0.02	Normal L-shaped	Expansion
	58	0.57	N	N	<u>0.08</u>	0.04	Normal L-shaped	Expansion*
Severn RBD Region	84	0.56	Y	Y	0.008	0.002	Normal L-shaped	Expansion
	71	0.55	N	N	0.01	0.004	Normal L-shaped	Expansion
Western Wales RBD Region	102	0.54	Y	Y	N/A	N/A	Normal L-shaped	Stable
	95	0.54	N	N	N/A	N/A	Normal L-shaped	Stable
Central England Cluster	132	0.73	N	Y	N/A	N/A	Normal L-shaped	Stable
	112	0.73	N	N	N/A	N/A	Normal L-shaped	Stable
South West Cluster	78	0.60	N	Y	N/A	N/A	Normal L-shaped	Stable
	65	0.59	N	N	N/A	N/A	Normal L-shaped	Stable
Wales & Borders Cluster	186	0.58	N	Y	0.002	0.001	Normal L-shaped	Expansion
	170	0.56	N	N	0.01	0.005	Normal L-shaped	Expansion

* one p-value significant for expansion, the other p value approaching significance for the same signal

Table 4.3: Bottleneck results for temporally restricted analyses. Dataset, the geographic or genetic extent of the data; Years, the years the dataset covers, N, number of individuals included in the analysis; Mean He, the mean heterozygosity across samples in the dataset; P-values, the p-values for the relevant one-tailed Wilcoxon test using the two-phase model of mutation with 88% stepwise mutation and variance 30 and 12 respectively; Detected Signal, population state according to the Wilcoxon test. NA indicates p-value was non-significant.

Dataset	Years	N	Mean He	P-values	Detected Signal
Eastern RBD Region	2009-2014	43	0.71	0.05 / 0.05	Bottleneck
Northern RBD Region	2009-2014	20	0.69	N/A	Stable
South West RBD Region	2009-2014	37	0.58	0.02 / 0.01	Expansion
Severn RBD Region	2009-2014	22	0.54	N/A	Stable
Western Wales RBD Region	2009-2014	31	0.56	N/A	Stable
Wales and Borders Cluster	1993-1995	25	0.50	N/A	Stable
Wales and Borders Cluster	2014	28	0.53	N/A	Stable
Wales and Borders Cluster	1993-1999	59	0.52	N/A	Stable
Wales and Borders Cluster	2009-2014	53	0.56	0.02 / 0.02	Expansion

4.3.2 Effects of population structure and admixture on effective population size estimates

Estimates of N_e obtained from tests which did not account for population genetic structuring (i.e. all data combined, and all Wales and England data treated as one population), resulted in N_e values that were considerably lower than when population structure was accounted for (approximately 3 and 4 times lower than for summed genetic clusters and summed RBD regions respectively - Figure 4.2). Excluding admixed individuals from the datasets resulted in different outcomes for different regions. For example, with admixed individuals excluded the estimate of N_e for the Northern RBD region was lower than with admixed individuals included, whereas for both the South West and the Severn and Western Wales RBD regions the opposite was true (Figure 4.2A). N_e estimates by genetic cluster varied less when admixed individuals were removed than by RBD region (Figure 4.2B). In all cases the 95% confidence intervals of the estimates both with and without admixed individuals overlapped indicating no significant difference in the estimates. Regardless of the inclusion of admixed individuals, all estimates of N_e were low with the sum of the RBD region results totalling 170.6 (95% CI: 102.1 – 348.3), and the sum of the genetic clusters totalling 121.3 (95% CI: 88.4 – 171.2, see Supplementary 4.4 for full details).

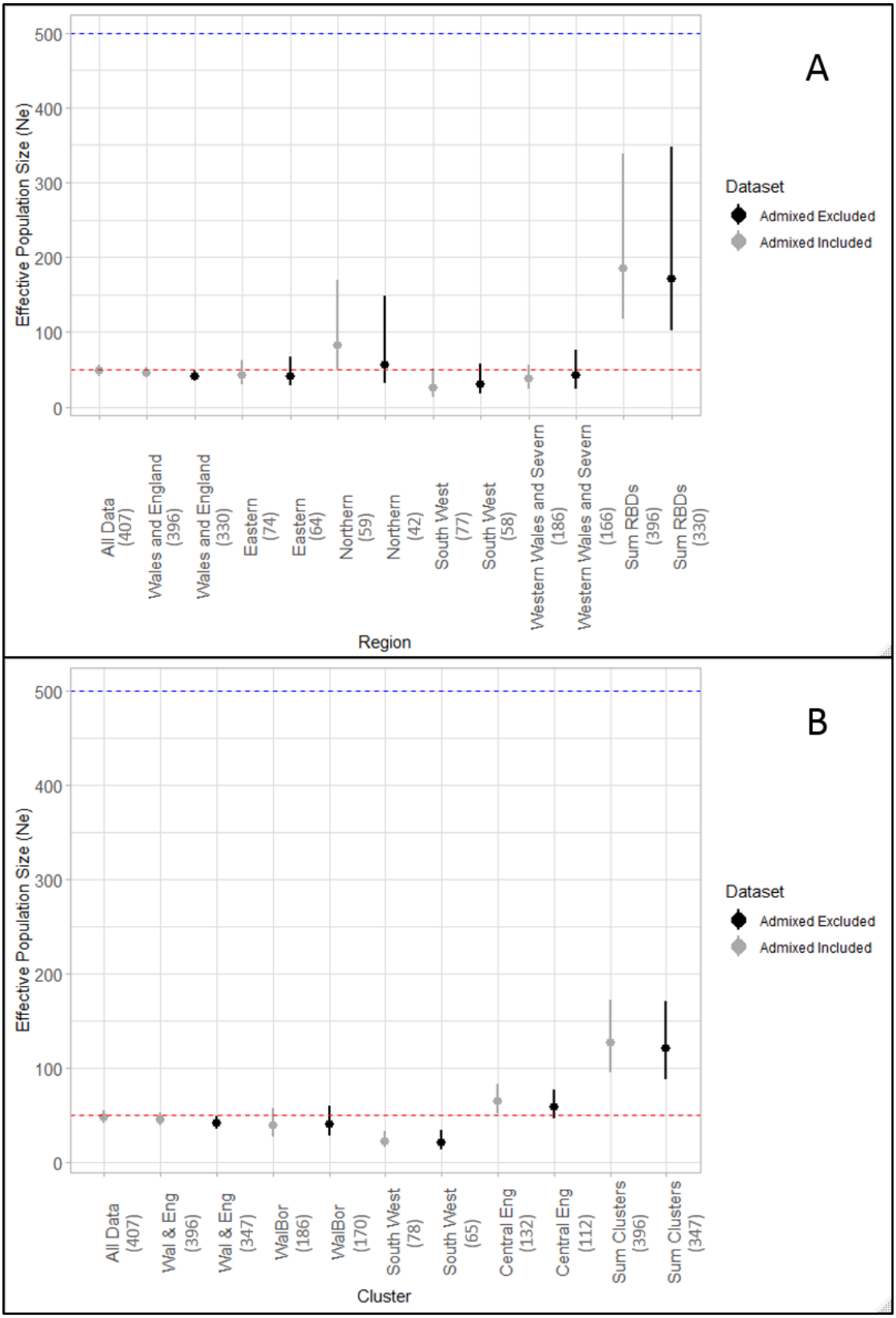


Figure 4.2: Estimates of effective population size (N_e) based on single-sample linkage disequilibrium (LD) methods with and without accounting for population structure and admixed individuals. Confidence intervals (95%) are based on the jack-knife across samples method. Numbers in brackets indicate sample size. Horizontal dashed lines indicate two critical values of N_e (as proposed by Franklin 1980): 50 in red, to reduce the risk of inbreeding depression; and 500 in blue, to maintain adaptive potential. (A) analysis performed on a geographic basis using River Basin District regions to split the data into populations; (B) analysis performed on a genetic cluster basis using average assignment across 10 STRUCTURE runs at K=3. In both plots 'All Data' and 'Wales and England' or 'Wal & Eng' refer to the datasets run without consideration of genetic structure. Grey indicates analyses where admixed individuals were included and black indicates analyses where admixed individuals were excluded on a q<0.8 basis.

4.3.3 Temporal changes to effective population size estimates

For both RBD regions and genetic clusters the majority of N_e estimates increased when using the 'late' dataset compared with the 'early' dataset (Figure 4.3 and Supplementary 4.5). The one exception was the Northern RBD region, where the point estimate for the late dataset was lower than the earlier one, however the upper confidence interval for this estimate tended to infinity indicating low reliability. Cluster-based estimates had narrower confidence intervals than RBD region-based estimates, with N_e values falling below $N_e = 50$ for all three genetic clusters although confidence intervals overlapped this minimum viable population boundary. None of the Ne estimates, including the sum of the cluster estimates, crossed the Ne = 500 threshold to maintain long-term adaptive potential. The more contemporary estimates tended to have a lower precision than their respective early data or all data counterparts as illustrated by the wider confidence intervals. The sum of both RBD region and genetic cluster estimates for N_e were significantly larger than estimates for the whole population without consideration for genetic structure across all temporal groupings. The summed estimates of N_e were approximately four times larger for the RBD regions (4.0 – 4.2) and nearly three times larger for the genetic clusters (2.6 - 2.9), additionally the 95% confidence intervals for all data versus summed data estimates did not overlap in all instances except for the 'early' dataset using genetic clusters.

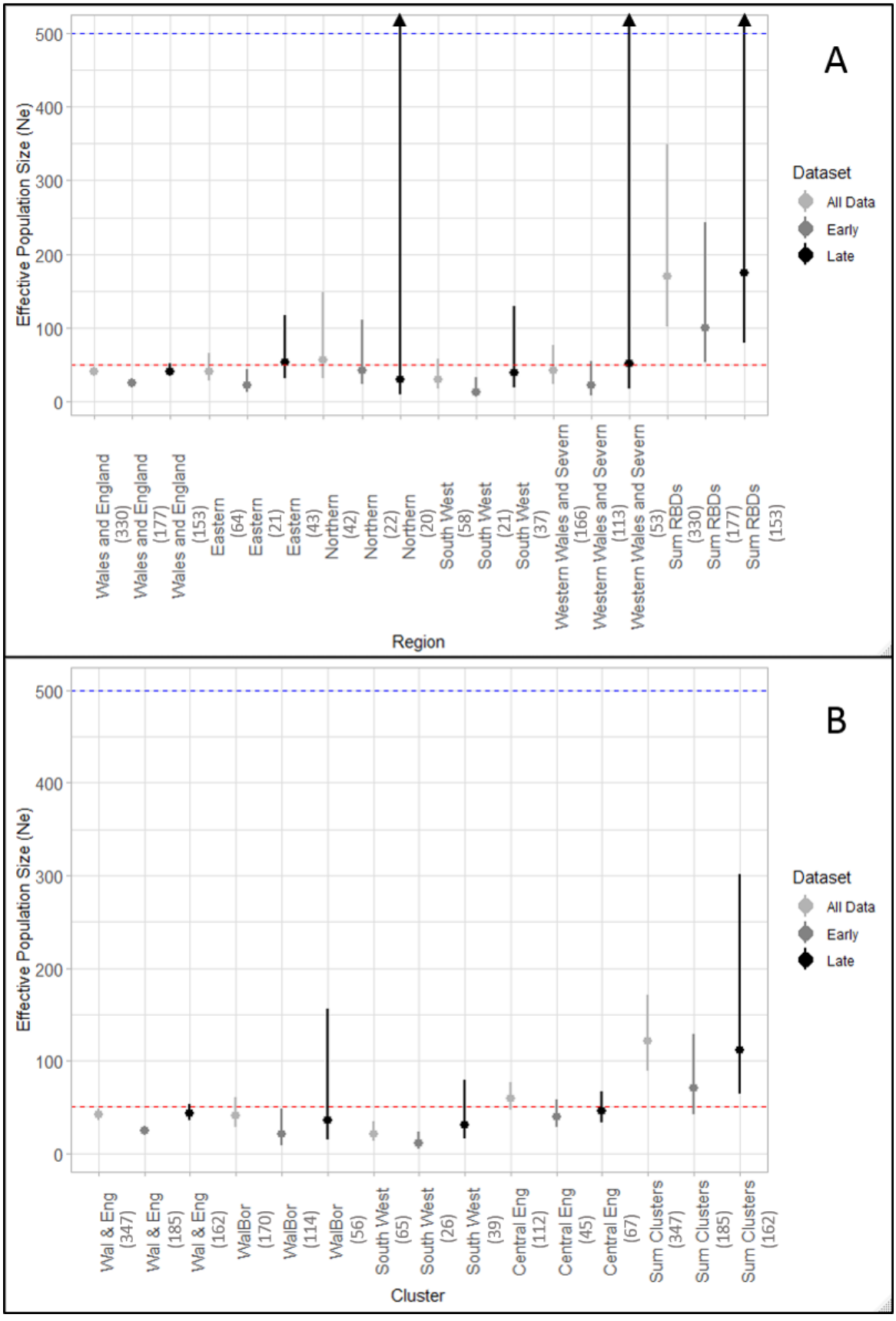


Figure 4.3: Estimates of effective population size (N_e) based on single-sample linkage disequilibrium (LD) methods with temporal comparison. Confidence intervals (95%) are based on the jack-knife across samples method. Numbers in brackets indicate sample size. Horizontal dashed lines indicate two critical values of N_e (as proposed by Franklin 1980): 50 in red, to reduce the risk of inbreeding depression; and 500 in blue, to maintain adaptive potential. (A) analysis performed on a geographic basis using River Basin District regions to split the data into populations; (B) analysis performed on a genetic cluster basis using average assignment across 10 STRUCTURE runs at K=3. Black arrows indicate cases where the upper 95% confidence interval was estimated to be infinity.

4.4 Discussion

Understanding the recent demographic history and current status of populations is vital for their conservation. Laboratory studies have shown that past bottlenecks can affect the extinction risk of a population even after it recovers to its previous size (Bijlsma et al. 2000), therefore detecting bottleneck events and understanding which populations are more likely to be at greater risk of extinction irrespective of their current population size is important. Fluctuations in population size can also lead to a lower N_e/N_C ratio than expected (Wright 1938; Vucetich et al 1997) through persistently lower N_e even after N_C has recovered. The UK otter population has been through a documented severe population decline followed by recovery over the last 40 years. Understanding the genetic effects of these population fluctuations is important in determining whether the current population has recovered as robustly as has been implied from national survey data (Crawford 2010; Strachan 2015).

4.4.1 Population bottleneck detection

Our results suggest that the Eastern RBD region is the only region likely to have gone through a significant recent population bottleneck, and that populations in the Severn and South West RBD regions have recently expanded while the Western Wales and Northern RBD regions have not. These results largely support our hypotheses based on the relative severity of population decline and speed of recovery across the regions recorded in the national surveys (Crawford 2010; Strachan 2015).

As predicted, when population genetic structure was not accounted for in the dataset there was a signal of population expansion. It is likely that this signal was the result of the violation of assumptions, rather than a true signal, and when the regions were analysed separately, and admixed individuals were removed from the regional datasets, there was an increase in the significance of the Wilcoxon's tests in the case of population bottlenecks and a decrease in the significance of population expansions.

Our results differ from those previously found in a pan-European study (Randi et al. 2003) where no population bottlenecks were detected, including for samples located within the UK. However, the sample sizes per country in the study were small (N ranged from three to 29, with only five samples from the UK), largely below the threshold advised by Cornuet and Luikart (1996) for analysis with BOTTLENECK. A previous genetic study of otters in the UK (Hobbs et al. 2011) also found no evidence of population bottlenecks, apart from in one sub-population covering the Northern England and the Scottish Borders region. The very different results presented in the current study are likely to reflect differences in the sub-populations tested (Hobbs et al. (2011) tested 11 sub-populations across the UK designated through progressive partitioning analysis,

whereas the current study focused on three overarching genetic clusters and four RDB regions as designation of sub-populations), or temporal changes in the populations picked up through more contemporary sampling in the current study.

4.4.2 Effective population size

As predicted, the N_e estimates for the datasets without consideration of genetic structure were considerably lower than the estimates based on the sum of either the RBD regions, or the genetic clusters. Such an under-estimation of N_e when genetic substructure is not taken into consideration is of similar magnitude to that found by Kopatz et al. (2017), whose estimates were nearly three times smaller when substructure in the dataset was unaccounted for (the values in this study ranged from 2.6 – 2.9 and 4.0 – 4.2 for genetic clusters and RBD regions, respectively).

The inclusion of individuals with admixed genotypes had less predictable effects on the analyses, with some regions or clusters having increased N_e estimates once admixed individuals were removed (e.g. South West RBD region) and others exhibiting decreased N_e estimates (e.g. Northern RBD region). This is somewhat in contrast with the findings of Kopatz et al. (2017) who found that the inclusion of admixed individuals caused a large upward bias in N_e estimates in brown bear populations in Finland. These differences may be because there was less structure in the brown bear population (which had only two sub-populations, potentially resulting in less complex patterns of admixture (Kopatz et al. 2017). Another factor likely to affect the magnitude of any changes in N_e could be the relative proportion of admixed individuals in each dataset, although Kopatz et al. (2017) detected 12% of admixed individuals in their overall dataset which is similar to the rate observed in our dataset. Recent work has shown that studies utilizing a moderate number of microsatellites (such as the current study) may underestimate actual admixture that becomes apparent from genomic-scale datasets (Gómez-Sánchez et al. 2018). The effect of admixed individuals on estimates of N_e may differ between study systems and populations, therefore understanding such potential biases clearly warrants further investigation.

For the estimates using data from all time points, the South West genetic cluster exhibited the smallest N_e, with both the point estimate and the confidence interval below 50, while the Northern RBD region had the highest N_e, with an upper 95% CI of over 100. This larger population size estimate could be due to the genetic contiguity of Northern English otters with the Scottish population, or the fact that the population in the 1990's was augmented by releases of rehabilitated otters likely from other parts of the UK (Green 1997). The obtained N_e estimate for Northern England could therefore represent an area larger than that which was actively sampled.

Census population size estimates derived from national survey data across Wales and England give a population estimate of 3,900 for the study area (Matthews et al. 2018). The highest

estimates from this study (the summed N_e from across RBD regions with and without admixed individuals) suggest N_e values of 185.6 and 170.6 individuals respectively, which, using the 0.1 ratio generalisation of N_e/N_C (Frankham 1995b; Hoban 2021a) translates to census population sizes of 1,856 and 1,706 for Wales and England. This suggests that the census population size estimate of 3,900 otters based on national survey results (which is considered to have low reliability due to the methods used) overestimates the true population size – although upper 95% confidence intervals from the current study encompass estimates of up to 3,387 and 3,483 (with and without admixed individuals respectively) which are nearer, but still lower than, the estimate calculated by Matthews et al. (2018).

LDNe (as with other single and two-sample estimators of N_e) assumes discrete generations, an assumption which otter population demography violates. Overlapping generations within a dataset have been shown to produce estimates that are more reflective of the number of breeders than of N_e, but if the number of cohorts sampled is enough to represent a generation then the estimate will be approximately equal to N_e (Waples et al. 2014). Therefore despite the overlapping generations in our dataset, due to the number of years covered by the sampling regime we would expect the estimates to be approximately equal to N_e, although estimates using all of the data time points may be more reliable estimates of N_e than those using temporal sub-sampling. Recent population size change may also have effects on N_e estimation by altering the pattern of linkage disequilibrium and may bias the estimates either upwards or downwards for a few generations (Waples 2005). Given the population history of otters in the UK and the results from the BOTTLENECK analyses in this study, the estimates may be biased due to recent changes in size, further reinforcing the need for continued genetic monitoring of the population.

4.4.3 Minimum viable populations

The discussion of what constitutes a minimum viable population (MVP) has been ongoing in the field of conservation genetics since 1980 when Franklin (1980) first proposed the '50/500 rule'. This rule stipulates that to avoid inbreeding depression in the short-term, a minimum $N_e \geq 50$ is required, with a larger minimum $N_e \geq 500$ required to preserve evolutionary potential and adaptive variation in the long-term.

As more studies, especially on wild populations, have accumulated, questions have been raised over whether the two minimum N_e sizes of 50 and 500 are large enough to avoid detrimental loss of genetic diversity over their respective time frames. Frankham et al. (2014) proposed that the rule be changed to 100/1000, based on new evidence over the last 30 years, although others maintain that the 50/500 rule is sufficient (Jamieson and Allendorf 2012).

The estimates of N_e for the otter population in Wales and England from this study fall clearly below either of the proposed minimum values (500 or 1000) for long-term viability, indicating that the evolutionary potential of the population and its ability to adapt to future environmental changes and stressors is currently at risk. In the short term, many of the geographic regions and genetic clusters are estimated to have an N_e of around 50 indicating that they could also be at risk of inbreeding depression.

The emphasis on genetic diversity in monitoring and management of wild populations has been increasing over the last decade (Frankham 2010; Hoban et al. 2013), with recent renewed calls for genetic monitoring to be included in international policy (Laikre et al. 2020). Specifically, N_e and the proportion of populations with $N_e \geq 500$ has been proposed as one of three genetic indicators for the 2030 Convention on Biological Diversity (CBD) targets (Hoban et al. 2020). While discussions in the scientific community around how best to word the target to policy makers and conservation managers are ongoing (Frankham 2021; Laikre 2021), there is broad agreement that an indicator that uses N_e to track the maintenance of genetic diversity in wild populations and species as well as domesticated ones is vital.

Here, we have shown that despite being hailed as a conservation success story (Crawford 2010), otter populations across Wales and England have not yet reached the N_e necessary for long-term viability, with many estimates additionally falling within the bounds of questionable viability over the short-term. While other work in this thesis (Chapter 2) indicates that gene flow between populations is increasing, the high and maintained F_{ST} values between RBD regions relating to the three main genetic clusters across the study area indicate that substantial genetic structure remains.

4.4.4 Further work

Continued genetic monitoring of the otter population in Great Britain is advised, to track N_e and other genetic diversity metrics, as the population continues to recover. The national otter surveys, which provided a means of monitoring otter presence across the landscape over the last 40 years have not been updated since 2010 (Crawford 2010; Strachan 2015 - although an updated national survey in Wales is due to be released imminently) leaving a gap in our knowledge of both the current distribution and continued expansion of the otter population, particularly in England. The low N_e estimates reported in our study indicate the importance of including genetic monitoring of species into national monitoring plans: importantly, the conclusions drawn from successive national surveys using otter sign (e.g. spraint and footprints), namely of a robust population close to panmixia, are not supported by the genetic evidence.

A previous study by Stanton et al. (2014) indicated that the most pronounced genetic divide in the otter population of Great Britain was a North-South split, with otters in the area approximately equivalent to the Northern RBD region grouped with those in Scotland. Limitations in the available sample size and temporal coverage of Scotland meant that it was not possible to appropriately investigate this area in the current study. Extending the genetic monitoring to include the Scottish population of otters would not only provide a fuller view of the situation across Great Britain, but also resolve whether the Northern RBD region is genetically contiguous with this population, and allow N_e estimates for this region to be put in more detailed context.

4.5 Conclusions

Despite a well-documented range expansion and accompanying population expansion over the last 40 years, otters in Wales and England still exhibit small effective population sizes that are well below those required for long-term viability. Several of the genetic sub-populations and regions also fall below the effective population size required to avoid inbreeding depression and maintain viability in the short-term. The South-East of England where the population decline was most severe and the population has taken longest to recover still shows the genetic signature of a population bottleneck, whereas other areas, such as the Severn have signatures of population expansion. These results paint a more precarious picture than that of the last national surveys in Wales and England which showed otter presence at 90% and 59% of surveyed sites respectively (Crawford 2010; Strachan 2015) and highlight the need for continued monitoring of the otter population. Small effective population sizes may reduce the ability of the otter sub-populations across Wales and England to respond to future environmental changes and threats as their adaptive potential is reduced.

Supplementary Material

S4.1 Full Bottleneck results with and without accounting for genetic structure and admixture. Dataset, the geographic or genetic extent of the data; N, number of individuals included in the analysis; Mean He, the mean heterozygosity across samples in the dataset; Possible biases – genetic structure, Y indicates genetic structure present in the dataset, N indicates genetic structure accounted for; Possible biases – admixed individuals; Y indicates admixed individuals (based in data from 15 microsatellite loci and a STRUCUTRE q-value threshold of 0.8) were included in the dataset, N indicates admixed individuals excluded from the dataset; Test, refers to the specific test within Bottleneck; Test Stat, is the test statistic; p-value, is the p-value of the given test; SMM, single mutation model; TPM (88%, 30), Two-Phase Model of mutation with 88% stepwise mutations and variance of 30; TPM (88%, 12), Two-Phase Model of mutation with 88% stepwise mutations and variance of 12; Mode Shift, the distribution of allele frequencies. Significant p-values (p<0.05) are highlighted in bold.

		Possible biases				SMM		TPM (88%, 30)		TPM (88%, 12)		Mode-Shift
Dataset	**N**	**Mean He**	**Genetic structure**	**Admixed individuals**	**Test**	**Test Stat**	**p- value**	**Test Stat**	**p- value**	**Test Stat**	**p- value**	
All Data	407	0.68	Yes	Yes	Sign Test (No. loci with heterozygosity excess)	2	**<0.001**	5	**0.040**	5	**0.039**	Normal L-shaped
					Standardized differences*	-7.781	**<0.001**	-2.073	**0.020**	-2.778	**0.003**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.047**		**0.021**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.958		0.982	
					Wilcoxon test (two-tail)		**<0.001**		0.095		**0.041**	
Wales and England	396	0.68	Yes	Yes	Sign Test (No. loci with heterozygosity excess)	2	**<0.001**	5	**0.041**	5	**0.044**	Normal L-shaped
					Standardized differences*	-8.089	**<0.001**	-2.083	**0.019**	-2.996	**0.001**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		0.0535		**0.018**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.953		0.985	
					Wilcoxon test (two-tail)		**<0.001**		0.107		**0.040**	
Wales and England	347	0.68	Yes	No	Sign Test (No. loci with heterozygosity excess)	3	**0.002**	5	0.043	5	**0.039**	Normal L-shaped
					Standardized differences*	-7.763	**<0.001**	-2.234	0.013	-2.904	**0.002**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.047**		**0.021**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.958		0.982	
					Wilcoxon test (two-tail)		**0.001**		0.095		**0.041**	
Eastern RBD Region	74	0.72	No	Yes	Sign Test (No. loci with heterozygosity excess)	10	0.385	13	0.023	13	**0.023**	Normal L-shaped
					Standardized differences*	-0.852	0.197	1.412	0.079	1.183	0.118	

					Wilcoxon test (one-tail for deficiency i.e. expansion)		0.849		0.989		0.982	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.165		**0.013**		**0.021**	
					Wilcoxon test (two-tail)		0.330		**0.026**		**0.041**	
Eastern RBD Region	64	0.71	No	No	Sign Test (No. loci with heterozygosity excess)	10	0.386	12	0.077	13	**0.024**	Normal L-shaped
					Standardized differences*	-0.666	0.253	1.392	0.0820	1.113	0.133	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		0.661		0.991		0.989	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.360		**0.011**		**0.013**	
					Wilcoxon test (two-tail)		0.720		**0.022**		**0.026**	
Northern RBD Region	59	0.7	No	Yes	Sign Test (No. loci with heterozygosity excess)	5	0.0383	10	0.388	10	0.385	Normal L-shaped
					Standardized differences*	-2.867	**0.002**	-0.067	0.473	-0.451	0.326	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**0.047**		0.533		0.489	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.958		0.489		0.533	
					Wilcoxon test (two-tail)		0.095		0.978		0.978	
Northern RBD Region	42	0.69	No	No	Sign Test (No. loci with heterozygosity excess)	7	0.230	10	0.378	10	0.376	Normal L-shaped
					Standardized differences*	-1.727	0.042	0.415	0.339	0.055	0.478	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		0.227		0.700		0.598	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.789		0.319		0.423	
					Wilcoxon test (two-tail)		0.454		0.639		0.847	
South West RBD Region	77	0.57	No	Yes	Sign Test (No. loci with heterozygosity excess)	2	**<0.001**	6	0.108	6	0.101	Normal L-shaped
					Standardized differences*	-4.798	**<0.001**	-1.909	**0.028**	-2.227	**0.013**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.047**		**0.024**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.958		0.979	
					Wilcoxon test (two-tail)		**<0.001**		0.095		**0.048**	
South West RBD Region	58	0.57	No	No	Sign Test (No. loci with heterozygosity excess)	3	**0.003**	6	0.125	6	0.125	Normal L-shaped
					Standardized differences*	-4.16	**<0.001**	-1.495	0.067	-1.769	**0.038**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		0.076		**0.042**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.932		0.964	

					Wilcoxon test (two-tail)		**<0.001**		0.151		0.083	
Severn RBD Region	84	0.56	No	Yes	Sign Test (No. loci with heterozygosity excess)	1	**<0.001**	5	**0.039**	4	**0.011**	Normal L-shaped
					Standardized differences*	-5.96	**<0.001**	-2.8	**0.003**	-3.236	**<0.001**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.008**		**0.002**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.994		0.999	
					Wilcoxon test (two-tail)		**<0.001**		**0.015**		**0.003**	
Severn RBD Region	71	0.55	No	No	Sign Test (No. loci with heterozygosity excess)	1	**<0.001**	4	**0.011**	4	**0.011**	Normal L-shaped
					Standardized differences*	-5.405	**<0.001**	-2.415	**0.008**	-2.838	**0.002**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.013**		**0.005**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.989		0.996	
					Wilcoxon test (two-tail)		**<0.001**		**0.026**		**0.010**	
Western Wales RBD Region	102	0.54	No	Yes	Sign Test (No. loci with heterozygosity excess)	2	**<0.001**	7	0.248	7	0.247	Normal L-shaped
					Standardized differences*	-3.151	**<0.001**	-0.76	0.224	-1.084	0.139	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**0.004**		0.244		0.180	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.998		0.773		0.835	
					Wilcoxon test (two-tail)		**0.008**		0.489		0.359	
Western Wales RBD Region	95	0.54	No	No	Sign Test (No. loci with heterozygosity excess)	2	**<0.001**	8	0.425	5	0.040	Normal L-shaped
					Standardized differences*	-2.852	**0.002**	-0.661	0.254	-0.97	0.167	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**0.001**		0.262		0.126	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.999		0.756		0.885	
					Wilcoxon test (two-tail)		**0.003**		0.524		0.252	
Central England Cluster	132	0.73	No	Yes	Sign Test (No. loci with heterozygosity excess)	7	0.227	9	0.578	9	0.586	Normal L-shaped
					Standardized differences*	-2.395	**0.008**	0.843	0.200	0.337	0.368	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		0.084		0.849		0.738	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.924		0.165		0.281	
					Wilcoxon test (two-tail)		0.169		0.330		0.561	
Central England Cluster	112	0.73	No	No	Sign Test (No. loci with heterozygosity excess)	7	0.235	10	0.370	8	0.413	Normal L-shaped

					Standardized differences*	-1.916	**0.028**	0.906	0.182	0.482	0.315	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		0.076		0.916		0.820	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		0.932		0.094		0.195	
					Wilcoxon test (two-tail)		0.151		0.188		0.389	
South West Cluster	78	0.6	No	Yes	Sign Test (No. loci with heterozygosity excess)	3	**0.002**	7	0.062	7	0.232	Normal L-shaped
					Standardized differences*	-3.804	**<0.001**	-1.096	0.136	-1.52	0.064	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		0.262		0.151	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.756		0.8616	
					Wilcoxon test (two-tail)		**0.001**		0.524		0.303	
South West Cluster	65	0.59	No	No	Sign Test (No. loci with heterozygosity excess)	1	**<0.001**	8	0.423	7	0.238	Normal L-shaped
					Standardized differences*	-4.011	**<0.001**	-1.369	0.086	-1.684	**0.046**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		0.244		0.151	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.773		0.862	
					Wilcoxon test (two-tail)		**<0.001**		0.489		0.303	
Wales and Borders Cluster	186	0.57	No	Yes	Sign Test (No. loci with heterozygosity excess)	1	**<0.001**	3	**0.002**	3	**0.002**	Normal L-shaped
					Standardized differences*	-9.155	**<0.001**	-4.064	**<0.001**	-4.769	**<0.001**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.002**		**<0.001**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.999		0.999	
					Wilcoxon test (two-tail)		**<0.001**		**0.003**		**0.002**	
Wales and Borders Cluster	170	0.56	No	No	Sign Test (No. loci with heterozygosity excess)	1	**<0.001**	3	**0.002**	3	**0.002**	Normal L-shaped
					Standardized differences*	-7.325	**<0.001**	-2.964	**0.002**	-3.529	**<0.001**	
					Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.011**		**0.005**	
					Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.991		0.996	
					Wilcoxon test (two-tail)		**<0.001**		**0.022**		**0.010**	

*Fewer than the required 20 loci

S4.2 Temporally restricted Bottleneck results using data from 2009 and 2014 only. Dataset, the geographic (RBD region) extent of the data; N, number of individuals included in the analysis; Mean He, the mean heterozygosity across samples in the dataset; Test, refers to the specific test within Bottleneck; Test Stat, is the test statistic; p-value, is the p-value of the given test; SMM, single mutation model; TPM (88%, 30), Two-Phase Model of mutation with 88% stepwise mutations and variance of 30; TPM (88%, 12), Two-Phase Model of mutation with 88% stepwise mutations and variance of 12; Mode Shift, the distribution of allele frequencies. Significant p-values (p<0.05) are highlighted in bold.

Dataset	N	Mean He	Test	SMM		TPM (88%, 30)		TPM (88%, 12)		Mode Shift
				Test Stat	p- value	Test Stat	p- value	Test Stat	p- value	
Eastern	43	0.71	Sign Test (No. loci with heterozygosity excess)	8	0.402	12	0.089	12	0.080	Normal L-shaped
			Standardized differences*	-0.876	0.191	0.999	0.159	0.798	0.212	
			Wilcoxon test (one-tail for deficiency i.e. expansion)		0.719		0.958		0.953	
			Wilcoxon test (one-tail for excess i.e. bottleneck)		0.300		**0.047**		0.054	
			Wilcoxon test (two-tail)		0.600		0.095		0.107	
Northern	20	0.69	Sign Test (No. loci with heterozygosity excess)	8	0.411	9	0.596	9	0.603	Normal L-shaped
			Standardized differences*	-1.879	0.030	-0.201	0.420	-0.36	0.361	
			Wilcoxon test (one-tail for deficiency i.e. expansion)		0.195		0.533		0.511	
			Wilcoxon test (one-tail for excess i.e. bottleneck)		0.820		0.489		0.511	
			Wilcoxon test (two-tail)		0.389		0.978		1.000	
South West**	37	0.58	Sign Test (No. loci with heterozygosity excess)	2	**<0.001**	4	**0.013**	4	**0.013**	Normal L-shaped
			Standardized differences*	-3.448	**<0.001**	-1.415	0.079	-1.77	**0.040**	
			Wilcoxon test (one-tail for deficiency i.e. expansion)		**<0.001**		**0.024**		**0.013**	
			Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.979		0.989	
			Wilcoxon test (two-tail)		**<0.001**		**0.048**		**0.026**	
Severn	22	0.54	Sign Test (No. loci with heterozygosity excess)	5	**0.041**	7	0.234	6	0.121	Normal L-shaped
			Standardized differences*	-1.867	**0.031**	-0.565	0.286	-0.69	0.246	
			Wilcoxon test (one-tail for deficiency i.e. expansion)		0.054		0.340		0.262	
			Wilcoxon test (one-tail for excess i.e. bottleneck)		0.953		0.681		0.756	
			Wilcoxon test (two-tail)		0.107		0.679		0.524	
Western Wales	31	0.56	Sign Test (No. loci with heterozygosity excess)	5	0.036	8	0.420	8	0.417	Normal L-shaped

Standardized differences*	-1.13	0.129	0.211	0.416	0.042	0.483
Wilcoxon test (one-tail for deficiency i.e. expansion)		0.068		0.756		0.555
Wilcoxon test (one-tail for excess i.e. bottleneck)		0.940		0.262		0.467
Wilcoxon test (two-tail)		0.135		0.524		0.934

*Fewer than the required 20 loci

S4.3 Temporally restricted Bottleneck results for the Wales and Borders Region. Dataset, the geographic (RBD region) extent of the data; Years, the years included in the analysis; N, number of individuals included in the analysis; Mean He, the mean heterozygosity across samples in the dataset; Test, refers to the specific test within Bottleneck; Test Stat, is the test statistic; p-value, is the p-value of the given test; SMM, single mutation model; TPM (88%, 30), Two-Phase Model of mutation with 88% stepwise mutations and variance of 30; TPM (88%, 12), Two-Phase Model of mutation with 88% stepwise mutations and variance of 12; Mode Shift, the distribution of allele frequencies. Significant p-values (p<0.05) are highlighted in bold.

Dataset	Years	N	Mean He	Test	SMM		TPM (88%, 30)		TPM (88%, 12)		Mode Shift
					Test Stat	p-value	Test Stat	p-value	Test Stat	p-value	
Wales and Borders	1993-1995	25	0.50	Sign Test (No. loci with heterozygosity excess)	7	0.226	9	0.560	8	0.445	Normal L shaped
				Standardized differences*	-1.039	0.149	0.042	0.483	-0.176	0.430	
				Wilcoxon test (one-tail for deficiency i.e. expansion)		0.262		0.661		0.577	
				Wilcoxon test (one-tail for excess i.e. bottleneck)		0.756		0.360		0.445	
				Wilcoxon test (two-tail)		0.524		0.720		0.890	
Wales and Borders	2014	28	0.53	Sign Test (No. loci with heterozygosity excess)	3	0.002	6	0.112	5	0.039	Normal L shaped
				Standardized differences*	-3.005	0.001	-1.202	0.115	-1.462	0.072	
				Wilcoxon test (one-tail for deficiency i.e. expansion)		0.006		0.104		0.060	
				Wilcoxon test (one-tail for excess i.e. bottleneck)		0.995		0.906		0.947	
				Wilcoxon test (two-tail)		0.012		0.208		0.121	
Wales and Borders	1993-1999	59	0.52	Sign Test (No. loci with heterozygosity excess)	6	0.100	8	0.424	8	0.431	Normal L shaped
				Standardized differences*	-1.819	0.034	-0.299	0.382	-0.387	0.350	
				Wilcoxon test (one-tail for deficiency i.e. expansion)		0.068		0.381		0.360	
				Wilcoxon test (one-tail for excess i.e. bottleneck)		0.940		0.640		0.661	
				Wilcoxon test (two-tail)		0.135		0.762		0.720	
Wales and Borders	2009-2014	53	0.56	Sign Test (No. loci with heterozygosity excess)	2	<0.001	4	0.013	4	0.011	Normal L shaped
				Standardized differences*	-3.597	<0.001	-1.415	0.079	-1.648	0.050	
				Wilcoxon test (one-tail for deficiency i.e. expansion)		<0.001		**0.024**		**0.015**	
				Wilcoxon test (one-tail for excess i.e. bottleneck)		1.000		0.980		0.987	
				Wilcoxon test (two-tail)		<0.001		**0.048**		**0.030**	

*Fewer than the required 20 loci

S4.4 Estimates of effective population size (Ne) based on single-sample linkage disequilibrium (LD) methods with and without accounting for population structure and admixed individuals. Dataset; the geographic or genetic extent of the data; Admixed Individuals, included (Yes) or excluded (No) based on a q<0.8 assignment to one of the three genetic clusters identified through STRUCTURE analysis; N, number of individuals included in the analysis; N_e, the effective population size estimate; L95%; the lower 95% confidence interval as determined through jack-knifing; U95%, the lower 95% confidence interval as determined through jack-knifing; Pcrit, the critical value for exclusion of rare alleles (based on sample size using the formula: 1/(2 x N) < Pcrit < 1/N).

Dataset	Admixed Individuals	N	N_e	L95%	U95%	Pcrit
All Data	Yes	407	47.8	41.1	55.5	None
Wales and England	Yes	396	45.4	38.9	52.8	None
Eastern RBD	Yes	74	41.7	29.7	61.7	0.01
Northern RBD	Yes	59	81.8	50.6	168.9	0.01
South West RBD	Yes	77	25	13	51.8	0.01
Western Wales & Severn RBDs	Yes	186	37.1	24.5	56.3	None
Sum RBD estimates	Yes	396	185.6	117.8	338.7	
Central England Cluster	Yes	132	64.6	51.6	82.5	None
South West Cluster	Yes	78	22.7	16.1	32.5	0.01
Wales and Borders Cluster	Yes	186	39.4	27.4	57.1	None
Sum Cluster Estimates	Yes	396	126.7	95.1	172.1	
Wales and England	No	330	40.7	34.6	47.7	None
Eastern RBD	No	64	41.5	28.2	66.2	0.01
Northern RBD	No	42	56.1	31.4	147.6	0.02
South West RBD	No	58	30.7	18.2	58.2	0.01
Western Wales & Severn RBDs	No	166	42.3	24.3	76.3	None
Sum RBD estimates	No	330	170.6	102.1	348.3	
Wales and England	No	347	41.3	35.2	48.4	None
Central England Cluster	No	112	59.1	46.7	76.7	None
South West Cluster	No	65	21.3	13.6	34.2	0.01
Wales and Borders Cluster	No	170	40.9	28.1	60.3	None
Sum Cluster Estimates	No	347	121.3	88.4	171.2	

S4.5 Estimates of effective population size (Ne) based on single-sample linkage disequilibrium (LD) methods with and without accounting for population structure and admixed individuals. Dataset; the geographic or genetic extent of the data; Time group; indicates the temporal spread of the samples included in the analysis (Early ≤ 2004, late ≥ 2009); N, number of individuals included in the analysis; N_e, the effective population size estimate; L95%; the lower 95% confidence interval as determined through jack-knifing; U95%, the lower 95% confidence interval as determined through jack-knifing; Pcrit, the critical value for exclusion of rare alleles (based on sample size using the formula: 1/(2 x N) < Pcrit < 1/N).

Dataset	Time group	N	N_e	L95%	U95%	Pcrit
Wales and England	All	330	40.7	34.6	47.7	None
Eastern RBD	All	64	41.5	28.2	66.2	0.01
Northern RBD	All	42	56.1	31.4	147.6	0.02
South West RBD	All	58	30.7	18.2	58.2	0.01
Western Wales & Severn RBDs	All	166	42.3	24.3	76.3	None
Sum RBD estimates	All	330	170.6	102.1	348.3	
Wales and England	All	347	41.3	35.2	48.4	None
Central England Cluster	All	112	59.1	46.7	76.7	None
South West Cluster	All	65	21.3	13.6	34.2	0.01
Wales and Borders Cluster	All	170	40.9	28.1	60.3	None
Sum Cluster Estimates	All	347	121.3	88.4	171.2	
Wales and England	Early	177	24.9	21.2	29.2	None
Eastern RBD	Early	21	21.7	13.2	43.6	0.02
Northern RBD	Early	22	42.3	24.4	111.5	0.02
South West RBD	Early	21	13.5	6.7	33.3	0.02
Western Wales & Severn RBDs	Early	113	22.3	8.6	55.1	None
Sum RBD estimates	Early	177	99.8	52.9	243.5	
Wales and England	Early	185	24.8	21.1	29	None
Central England Cluster	Early	45	39.2	28.2	58.4	0.02
South West Cluster	Early	26	10.2	4.6	22.6	0.02
Wales and Borders Cluster	Early	114	21.2	8.8	48.2	None
Sum Cluster Estimates	Early	185	70.6	41.6	129.2	
Wales and England	Late	153	41.8	34.1	51.6	None
Eastern RBD	Late	43	53.9	32.2	117.7	0.02
Northern RBD	Late	20	29.6	10.6	Inf	0.05
South West RBD	Late	37	39.1	19.6	129.2	0.02
Western Wales & Severn RBDs	Late	53	52.1	18.2	Inf	0.01
Sum RBD estimates	Late	153	174.7	80.6	Inf	
Wales and England	Late	162	43	35.2	52.7	None
Central England Cluster	Late	67	45.7	33	67.1	0.01
South West Cluster	Late	39	30.2	15.7	78.5	0.02
Wales and Borders Cluster	Late	56	35.9	15	155.7	0.01
Sum Cluster Estimates	Late	162	111.8	63.7	301.3	

5

General discussion

The past 50 years have seen an average 68% global decline in animal population sizes (WWF Living Planet Report, 2020). These alarming numbers are reflected in a concurrent, but smaller, global reduction in genetic diversity measured across a subset of these taxa (Leigh et al. 2018). Genetic diversity can be lost in as little as a generation, but may take hundreds of generations or more to recover (Nei et al. 1975). Against this backdrop of decline, some previously persecuted taxa have made recoveries in the 21st century (Chapron et al. 2014). If we are to reverse declines in species abundance more widely, and by extension genetic variation (Frankham 1996) to prevent harmful biodiversity loss (Pereira et al. 2013), then monitoring programs must embrace genetic tools alongside more traditional methods (Schwartz et al. 2007; Laikre et al 2020). Importantly, estimating the loss of genetic variation indirectly through population size change or extinction rates is likely inaccurate (Steffen et al. 2015), and therefore direct estimates using genetic data should be encouraged where possible (Laikre et al. 2021). Genetic tools can not only help characterise the loss of genetic diversity, but also track its recovery amongst those populations that are expanding (Hagen et al. 2015). This thesis provides novel insights into the latter using the well-documented otter population recovery in the UK. The conducted research has focussed on documenting spatial and temporal changes in population structure and connectivity (Chapter 2), quantifying environmental barriers and facilitators to gene flow (Chapter 3) and assessing recent changes in effective population size (Chapter 4). Taken together, these analyses build a more complete picture of the on-going genetic recovery of the species in the UK, and indicate that the conclusions drawn of a robust population close to panmixia from successive national surveys using otter sign alone (e.g. spraint and footprints) are not supported by the genetic evidence.

5.1 Main conclusions

In **Chapter 2** I found that despite the well-documented population recovery of otters in the UK, as measured through increased detection of otter sign over successive national surveys, significant population structure persists. Increases in gene flow between the different regions have yet to reduce differentiation between these sub-populations leaving them vulnerable to stochastic events. In **Chapter 3** I found that environmental factors influencing gene flow in otters across the Wales and Borders region are consistent across finer (100 m^2) and coarser (1 km^2) grain scales, with habitat type best explaining the functional connectivity of otters. Multi-surface models showed that elevation and water when combined (respectively) with habitat type provided

alternative explanations for functional connectivity that were equally effective when compared with habitat type alone. In general, higher elevations and habitat types associated with higher elevations (coniferous woodland and semi-natural grassland) as well as urban areas impeded otter geneflow and lower elevations, broadleaf woodland, arable land, rivers and coastline facilitated otter gene-flow. In **Chapter 4** I ascertained that effective population sizes of UK otter sub-populations fall far below those required for long-term population viability, and in some cases also raise concerns regarding short-term population viability. These data suggest that census population size is lower than previously predicted based on otter occupancy data from national surveys.

5.2 Importance of genetic variation

The importance of genetic variation as a cornerstone of biodiversity is globally recognised and emphasised by its inclusion as an Essential Biodiversity Variable (Pereira et al. 2013) and as a component of the planetary variables crucial for human survival on Earth (Mace et al. 2014). Yet, despite the maintenance of genetic variation having been included as one of the Aichi Targets for 2020 (Target 13 - CBD 2011), loss of genetic variation on a global scale is ongoing (Leigh et al. 2019).

Many studies have highlighted the lack of appropriate and comparable monitoring of genetic variation over the last decade (Feld et al. 2009; Pereira et al. 2013; Leigh et al. 2019). While some countries undertake routine genetic monitoring of specific species and populations (e.g. brown bears - Kendall et al. 2019; Kopatz et al. 2021; wolverines - Brøseth et al. 2010), this is far from the norm and often limited to taxa with economic and human-conflict importance (Salmonids – Hansen 2002; Van Doornik et al. 2011; wolves – Dufresnes et al. 2019). The call for a global framework and standardised indicators for genetic diversity is once again intensifying across the conservation community, with calls for specific genetic diversity indicators for wild species (in addition to domestic species) to be included in the next Convention on Biological Diversity (CBD) targets (Laikre et al. 2020; Hoban et al. 2020; Diaz et al. 2020; Laikre et al. 2021).

This thesis highlights that genetic monitoring of a population can add depth to our knowledge of population status and change, so complementing traditional occupancy monitoring. In some cases, insights from genetic monitoring might even contradict conclusions reached through traditional monitoring. It is rare to have available the spatial and temporal extent of occupancy data that were available to this study (from the five successive national surveys for otter presence, conducted from the late 1970's until 2010 Crawford 2010; Findlay 2015; Strachan 2015, though see Wei et al. 2018 for an example in giant pandas). Such data have been invaluable in monitoring the expansion of otter populations in the UK but have potentially led to a degree of complacency around their recovery (as reflected in the lack of updated national surveys since

2010, although a contemporary Welsh national survey is now awaiting publication and an English survey pending). Results from this thesis have challenged the assumption that spatial connectivity of previously isolated sub-populations constitutes recovery of the population as a whole, thereby emphasising the importance of genetic data for population monitoring. Chapter 2 showed that despite some increases in gene flow between regions in Wales and England, significant population structure and differentiation remains, along with regional differences in genetic diversity. This not only suggests that the otter population might be less robust than previously assumed but also highlights the role that genetic data can play in delineating non-panmictic populations through illuminating otherwise cryptic structure.

Another important factor to consider is the time lag between a demographic or environmental change, and the corresponding effect on genetic variation. Simulation studies by Landguth et al. (2010) and Epps and Keyghobadi (2015) have highlighted how, due to this time lag effect, both past demographic events and landscape composition may continue to be reflected in the contemporary genetic structuring of a population more strongly than current processes. Length of time lag is highly dependent on a variety of species-specific factors that affect the speed at which mutation-drift equilibrium is reached, with dispersal ability being particularly important (Epps and Keyghobadi 2015). In the case of reconnected populations dispersal ability is also linked to migration, and therefore changes in levels of functional connectivity, between previously isolated populations. Simulated loss and gain of barriers to gene flow indicated that for species with high dispersal abilities, such as otters, genetic time lags should be less than 15 generations with legacy effects lost by this point (Landguth et al. 2010). In wild populations other demographic factors can also play a role in time lags, for example - Keller et al. (2001) demonstrated how immigration, even at relatively low levels, was directly linked to a rapid post-bottleneck recovery of genetic variation in an insular population of song sparrows (*Melospiza melodia*). The study was able to clearly demonstrate that immigrant individuals were drivers of the restoration in genetic variation, with number of alleles and expected heterozygosity returning to pre-bottleneck levels in just one and three years respectively. In the current study, we observed little change in regional genetic variation over the study period, despite increases in inter-regional gene flow and in the proportion of admixed individuals between detected genetic clusters.

5.3 Demographic and genetic connectivity

Commonly used population monitoring tools in field ecology include direct counts, sign detection and methods using newer technologies such as unmanned aerial vehicles (Hodgson et al. 2016). However, count and occupancy data provides limited information on the connectivity of populations. Even individual tracking data (i.e telemetry) does not provide data on effective (reproductive) migration and therefore functional connectivity. For example, Riley et al. (2006)

demonstrated through a combination of genetic and telemetry data, that even though a large highway in California acted as a significant barrier to both bobcats (*Lynx rufus*) and coyotes (*Canis latrans*), 5-32% of sampled individuals dispersed across the road over the 7-year study period. Despite this quite modest level of migration, however, genetic analysis revealed that the rate of effective migration was considerably lower at less than 0.5%. This disparity between demographic and genetic connectivity was also highlighted by Schregel et al. (2017) in their large-scale study of Scandinavian brown bears (*Ursus arctos*), where substantial population recovery over recent decades had failed to erode population structure in either sex despite high mobility and male-biased dispersal. Their conclusion that both demographic and genetic connectivity must be estimated to avoid false assumptions about wildlife populations is echoed by the results in this thesis.

Relatively few empirical studies to date have investigated temporal changes in population structure and connectivity in expanding populations and conclusions have often been contrasting, even within the same species. For example, unlike in the study by Schregel et al. (2017), brown bears in Finland showed rapid declines in population differentiation over a short period of time (1.5 generations) during population expansion across a geographically similar area (Hagen et al. 2015). This highlights the need for more genetic studies on the dynamics of expanding populations and genetic monitoring of populations if we are to be able to make informed generalisations about genetic patterns in changing wildlife populations.

5.4 Genetic monitoring

While genetic assessments of populations are now relatively common, true genetic monitoring, i.e. temporally replicated sampling to detect changes and trends in genetic variation (Schwartz et al. 2007) is still rare and usually reserved for populations of economic interest (Van Doornik et al. 2011) or iconic species (Kendall et al. 2019). Yet, these data are vital for development and evaluation of suitable conservation and management actions (Frankham 2010). A simulation study by Hoban et al. (2014) indicated that samples from 50 individuals genotyped with 20 microsatellite markers at just two time points performed well in detecting signatures of genetic erosion. Despite this relatively cost-efficient protocol, genetic monitoring still lacks wide-scale uptake among conservation managers and practitioners, perhaps due to the difficulty in interpreting genetic signals at the policy/management level, especially in natural wildlife (i.e. non-model) populations (Pierson et al. 2015).

Cryptic population structure has been discovered in a wide range of highly mobile species with continuous distributions (Sacks et al. 2004; Pilot et al. 2006). Observations of 'continuous distributions' in these species may result in the incorrect assumption that populations are largely panmictic, therefore genetic data are vital in determining meta-population structure and the

connectivity between populations. Additionally, spatial processes can bias many genetic parameter estimates through violation of the Wright-Fischer idealised population model (Fisher 1930; Wright 1931), and therefore unaccounted for genetic structure can lead to erroneous results. Several studies, including this thesis have shown this to be true when estimating effective population size (Neel et al. 2013; Kopatz et al. 2017) and therefore an understanding of the underlying population structure is vital.

5.5 Pre-empting potential fragmentation events

The use of genetic data in assessing fragmentation of populations due to anthropogenic barriers is relatively well established (Reviewed in Keyghobadi 2007) with negative impacts recorded even for large, highly mobile species (Bighorn sheep - Epps et al. 2005; bobcats and coyotes - Riley et al. 2006; grizzly bears - Kendall et al. 2009). The pre-emptive use of genetic data to determine areas where connectivity is likely be affected by future anthropogenic actions is much rarer (e.g. Ma et al. 2018; Fedorca et al. 2019), however, as an additional piece of work during the production of this thesis, a report on the relatedness of otters in South-East Wales across a proposed major road development (the M4 relief road) was undertaken to provide evidence as to the current genetic connectivity of individuals across the area and the likely impacts that the new infrastructure would have (Appendix 1). This is an example of the timely use of genetic data in providing evidence to the legislative process, and one that ultimately helped in the decision to abandon the project due to negative environmental and biodiversity concerns. Instances such as this and the work of Fedorca et al. (2019) are necessary if we are to move away from strategies that attempt to reverse and mitigate fragmentation, and towards those that avoid further fragmentation of populations.

5.6 Caveats and implications for future monitoring

The research in this thesis has highlighted the importance of genetic monitoring, and the conclusions have indicated that incorporating an element of genetic monitoring into the current traditional occupancy surveys for otters in the UK would provide extra information on the population and help fill important gaps in the current knowledge. The current research was conducted using tissue samples from the Cardiff University Otter Project's archive obtained from deceased individuals (predominantly road-traffic accidents) reported by the public and collected by the statutory bodies.

The *ad-hoc* nature of these collections mean that sampling strategies may only be applied to the bank of geolocated samples already archived, and not to the landscape itself. This type of sampling may lead to biases or patchiness in the data relative to true otter abundance, including lack of samples where otter populations are small, mortality varies due to road density or where

there are geographic differences in reporting rates relating to varying engagement with the project across the UK. In the current study this resulted in small sample sizes at some time points especially for the Eastern and Northern river basin district regions. Ideally a minimum number of samples would be collected across all regions to allow for sufficient sample sizes for all analyses. Additionally, sampling locations would represent a continuum across the landscape to avoid any potential incidences of false detection of genetic clusters where isolation-by-distance better explained genetic differences due to sampling design (Schwartz and McKelvey 2009).

Sampling from deceased individuals combines the advantages of invasive sampling (i.e. higher quality DNA sources such as tissue) with those of non-invasive sampling (i.e. no need to capture or handle live individuals). However, it also means that only one location per individual can be established. For many demographic parameters, such as home range size, census population size and population density, estimation requires multiple 'captures' of the same individual which can be achieved via non-invasive genetic sampling (Waits et al. 2016). Studies using DNA collected from otter spraint samples would help improve estimation of these parameters in the UK, but the use of non-invasive sampling methods in otters has been hampered by low genotyping success rate, especially in temperate climates such as that of the UK (Coxon et al. 1999; White et al. 2013). Next generation sequencing (NGS) technologies have been shown to be more sensitive for low quality DNA samples than traditional capillary electrophoresis (Fordyce et al. 2015) and may provide a means of increasing genotyping success rates (De Barba et al. 2017). As such, the utility of these new technologies should be explored in relation to increasing genotyping success rates and therefore reducing costs.

Microsatellites are widely used in population genetic studies due to their relative abundance in the literature and low cost of genotyping (Selkoe and Toonen 2006; Kim and Sappington 2013). Due to the methods of analysing and genotyping individuals (i.e. traditionally via capillary electrophoresis) it can be difficult to amalgamate data from different studies and a calibration process is required to standardise datasets (Kraus et al. 2015). More recently genotyping by sequencing (GBS) using NGS technologies has provided an alternative analysis option that could facilitate data sharing through the generation of sequence data (Pimental et al. 2018) and such methodologies should be explored in relation to genetic monitoring. Advantages could include more replicable methodologies through the use of standardised bioinformatic pipelines in processing sequence data, increased automation of data production and the ability to scale up laboratory protocols to increase the number of individuals sampled.

The dataset used in this thesis is made up of predominantly road-killed animals. Those individuals killed on the roads of the UK, and therefore submitted to Cardiff University, may not be completely reflective of the wider population. There is a paucity of reliable assessment of the

demographics of wild otters across the UK due to the difficulty of directly observing live individuals as well as determining age and sex from non-invasive samples such as spraint sampling making significant deviations within the dataset harder to detect.

One study of otters in the South-West of England estimated sex ratios and genetic composition to be similar between carcass data and spraint sampling (Dallas et al. 2003). This suggests that otter carcasses can provide a sample that is unbiased in terms of sex ratio. The authors estimated an M:F ratio of 31:21 from the spraint samples and 41:29 from the carcasses, the current study has a sex ratio of 29:21 which is inline with these previous findings.

The age structure of samples in the dataset may not reflect the natural age structure of the population, with the potential for sub-adult individuals to be over-represented in the dataset due to juvenile or sub-adult dispersal behaviour, and therefore increased chance of road traffic collision (RTC), related to dispersal. However, this possibility is to some extent refuted by the proportions observed where measurable (5.0% juveniles; 35.5% sub-adults; 59.5% adults). In these instances, long-distance dispersers would be the individuals most likely to skew the dataset. If these long-distance dispersers are a particular morph or genetically pre-disposed to the behaviour i.e. naked mole rats (O'Riain et al 1996), then their effects on the conclusions drawn may be significant. However, if long-distance dispersers are simply a random sample of the population, then their effects on the results should be less problematic, as there is an equal chance that any individual could disperse a long distance. Otters, as far as we know, are not genetically pre-disposed to long distance dispersal and variation in dispersal distance is more likely due to random chance.

Volatile organic compounds from scent glands have been used to roughly age otters using the presence or absence of specific compounds (juvenile or adult with the interim assumed to indicate sub-adult; Kean et al. 2011), however initial investigations into using this method on spraint were unsuccessful. Cementum rings on the teeth of otters have also been used to estimate age, however under UK climatic conditions these rings are less clear than in other species i.e. badger (Meles meles; Harris et al. 1992) and stoats (Mustela erminea; King 1991) and otters from other locations (Denmark and Norway; Heggberget 1984) leading to ambiguity (Sherrard-Smith et al. 2010). It is possible that telomeres could be explored as a method of determining age in otters from genetic samples. Telomere length has been shown to be linked to age and lifespan in wild species (Fossel 2012; Dantzer et al. 2015) and has been used as a proxy for age in a study of age-related immune response in badgers (Beirne et al. 2016). However, each species requires individual calibration to estimate chronological age from telomere length and the result is age which requires samples from known age individuals either in the wild or from captivity (Haussmann et al. 2003).

Another option for future studies would be to target samples in the archive from non-collision (RTC) individuals. However, as over 85% of the otters received by the project are killed on the roads this would severely limit the pool of samples that could be genotyped. A broad scale genetic assessment of non-RTC individuals may be possible which could then be compared to the results obtained from RTC individuals to determine any significant differences between the two datasets. Similarly supplementary samples from live individuals, either through trapping or spraint collection, could be undertaken in specific key areas to address any sample bias from using RTC individuals. This would have to be geographically limited due to both the difficulty of live trapping (both legislatively and practically) and both the sampling effort required to obtain fresh spraint samples and the low genotyping success rates of these samples.

Partitioning of samples using the available metadata, i.e. sex and age class, can also be helpful in determining if genetic patterns are driven by a specific age group or sex. In the spatial-autocorrelation analysis the patterns of autocorrelation were the same over the distance classes for all individuals, adult females only and adult males only, suggesting that sex and age class are not specific drivers in the genetic patterns of relatedness that are observed. Further partitioning of samples for other genetic analyses may be useful in the future to confirm these findings.

Often genetic assessments and monitoring of populations are conducted in an *ad-hoc* way or after the negative impacts of anthropogenic effects are already being observed. For some populations historic or museum samples may be used as a proxy for past genetic variation (Pertoldi et al. 2005; Poulsen et al. 2006; Lonsinger et al. 2018), however the temporal spacing of samples, which can be large across historic samples, as well as the lower DNA quality and thus increased chance of genotyping errors can influence the genetic metrics calculated (Schwartz et al. 2007; Gauthier et al. 2020). Ideally genetic monitoring should take a Before-After-Control-Impact (BACI) type approach to ensure genetic metrics can be compared before and after an intervention (Schwartz et al. 2007) and genetic variation across a wide range of species should be monitored periodically to provide baseline levels of genetic variation in populations. Ensuring replicable methodologies across studies and analysis protocols is vital to ensure that key metrics are calculated in equivalent ways and are comparable (Hoban et al. 2021b).

In many countries, the organisations undertaking genetic assessments of populations (i.e. universities and research institutions) are uncoupled from those organisations responsible for the management of natural resources (i.e. government bodies). Where this is the case, better communication across the sectors to design and implement genetic monitoring programs in conjunction that are linked to broader conservation policy is vital (Hoban et al. 2014, Hoban et al. 2021b) if we are to avoid 'describing the world's fate ever more precisely while doing nothing to avoid it' (Fischer et al. 2012). As such, genetic monitoring programs should have clearly linked

actions that are implemented if agreed thresholds are crossed (Martin et al. 2012; Hoban et al. 2021b) to avoid perpetuation of the current situation where despite laws and policies existing, failure of proper implementation results in continued failure to achieve conservation goals (Schwartz 2020).

5.7 Future directions and conclusions

The findings in this thesis have highlighted many areas for future work and throughout Chapters 2-4 I have specified avenues for further research.

- Further investigation of the genetic changes that occur during population expansions across a variety of species and locations, to help uncover general patterns and indicators that could be used to track future population recoveries.
- Continued genetic monitoring of the UK otter population including tracking changes in effective population size, genetic differentiation and gene flow.
- Replicated landscape genetic studies using the same methods in different areas of the UK (or in the wider Eurasian otter range) to allow direct comparison of results and to determine how universal the environmental factors found to facilitate and impede gene flow in this study are across different landscapes.
- Further work on the estimation of effective population sizes in non-model wildlife populations, where complex demographic and genetic dynamics can confound results.
- Temporal replication of effective population size estimates across populations in various states (i.e. stable, expanding, decreasing) to determine any generalisations in trends.

In conclusion, this thesis has highlighted the value of genetic data, even for populations where a significant (non-genetic) monitoring program has existed for several decades. The conclusions reached about population status using genetic data were very different from those based on occupancy data, and highlight inherent dangers in assuming that spatial connectivity is an indicator of genetic connectivity.

References

Allendorf, F.W. and Luikart, G., 2009. *Conservation and the genetics of populations*. John Wiley & Sons.

Almond, R.E.A., Grooten, M. and Peterson, T., 2020. *Living Planet Report 2020-Bending the curve of biodiversity loss*. World Wildlife Fund.

Amos, W. and Harwood, J., 1998. Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society B*. 353: 177–186.

Andrews, E., Crawford, A., 1986. *Otter Survey of Wales 1984–85*. Vincent Wildlife Trust, London, UK.

Anderson, E.C., 2005. An efficient Monte Carlo method for estimating N_e from temporally spaced samples using a coalescent-based likelihood. *Genetics*, 170(2), pp.955-967.

Anderson, C.D., Epperson, B.K., Fortin, M.J., Holderegger, R., James, P.M., Rosenberg, M.S., Scribner, K.T. and Spear, S., 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular ecology*, 19(17), pp.3565-3575.

Andrews, E., Howell, P., Johnson, K., 1993. *Otter Survey of Wales 1991*. Vincent Wildlife Trust, London, UK.

Applied Biosystems. 2006. *Genemapper v4.0*. Applied Biosystems, Paisley, United Kingdom.

Arenas, M., Ray, N., Currat, M. and Excoffier, L., 2012. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution.* 29, pp.207–218.

Arrendal, J., Walker, C.W., Sundqvist, A.K., Hellborg, L. and Vilà, C., 2004. Genetic evaluation of an otter translocation program. *Conservation Genetics*, 5(1), pp.79-88.

Banks, S.C., Lindenmayer, D.B., Ward, S.J. and Taylor, A.C., 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.

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 Peninsula. *Biological conservation*, 114(3), pp.377-387.

Barton, N. H. & Slatkin, M., 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56: 409-415.

Bates, D., Mächler, M., Bolker, B. and Walker, S., 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv*:1406.5823.

Beaumont, M.A., 1999. Detecting population expansion and decline using microsatellites. *Genetics*, 153(4), pp.2013-2029.

Beirne, C., Waring, L., McDonald, R.A., Delahay, R. and Young, A., 2016. Age-related declines in immune response in a wild mammal are unrelated to immune cell telomere length. *Proceedings of the Royal Society B: Biological Sciences*, 283(1825), p.20152949.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), pp.289-300.

Berthier, P., Beaumont, M.A., Cornuet, J.M. and Luikart, G., 2002. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, 160(2), pp.741-751.

Bijlsma, R., Bundgaard, J. and Boerema, A.C., 2000. Does inbreeding affect the extinction risk of small populations?: predictions from *Drosophila*. *Journal of Evolutionary Biology*, 13(3), pp.502-514.

Bolliger, J., Lander, T. and Balkenhol, N., 2014. Landscape genetics since 2003: status, challenges and future directions. *Landscape Ecology*, 29(3), pp.361-366.

Bonesi, L., Hale, M. and Macdonald, D.W., 2013. Lessons from the use of non-invasive genetic sampling as a way to estimate Eurasian otter population size and sex ratio. *Acta theriologica*, 58(2), pp.157-168.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R. and Cavalli-Sforza, L.L., 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 368(6470), pp.455-457.

Broadmeadow, S. and Nisbit, T.R., 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.

Brøseth, H., Flagstad, Ø., Wärdig, C., Johansson, M. and Ellegren, H., 2010. Large-scale noninvasive genetic monitoring of wolverines using scats reveals density dependent adult survival. *Biological Conservation*, 143(1), pp.113-120.

Burnham, K.P. and Anderson, D.R., 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd Ed. New York: Springer-Verlag.

Cardillo, M., Mace, G.M., Jones, K.E., Bielby, J., Bininda-Emonds, O.R.P., Sechrest, W., Orme, C.D.L. and Purvis, A. 2005. Multiple causes of high extinction risk in large mammal species. *Science,* 309(5738), pp.1239–1241.

Chadwick, E.A., 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.R.F. and Jefferies, D.J., 1978. The decline of the otter *Lutra lutra L.* in Britain: an analysis of hunting records and discussion of causes. *Biological Journal of the Linnean Society*, 10(3), pp.305-328.

Chanin, P., 2003. *Monitoring the otter*. Conserving Natura 2000 Rivers Monitoring Series 10. English Nature, Peterborough

Chapron, G., Kaczensky, P., Linnell, J.D., von Arx, M., Huber, D., Andrén, H., López-Bao, J.V., Adamec, M., Álvares, F., Anders, O. and Balčiauskas, L., 2014. Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science*, 346(6216), pp.1517-1519.

Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), pp.195-205.

Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B. and Beaumont, M.A., 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, 186(3), pp.983-995.

Clark, R.W., Brown, W.S., Stechert, R. and Zamudio, K.R., 2008. Integrating individual behaviour and landscape genetics: the population structure of timber rattlesnake hibernacula. *Molecular ecology*, 17(3), pp.719-730.

Clarke, R.T., Rothery, P. and Raybould, A.F., 2002. Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(3), pp.361-372.

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

Cornuet, J.M. and Luikart, G., 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144(4), pp.2001-2014.

Coxon, K., Chanin, P., Dallas, J. and Sykes, T., 1999. *The use of DNA fingerprinting to study population dynamics of otters (Lutra lutra) in Southern Britain: a feasibility study*. R&D Technical Report W202, Environment Agency, Bristol, United Kingdom

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M. and Wayne, R.K., 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*. 15: 290-295.

Crawford, A., Evans, D., Jones, A., McNulty, J., 1979. *Otter Survey of Wales 1977–1978*. Nature Conservancy Council, London, UK.

Crawford, A., 2003. *Fourth otter survey of England 2000-2002*. Environmental Agency, Bristol, UK.

Crawford, A., 2010. *Fifth Otter Survey of England 2009-2010*. Environment Agency Technical Report 126, Bristol, UK.

Cristescu, R., Sherwin, W.B., Handasyde, K., Cahill, V. and Cooper, D.W., 2010. Detecting bottlenecks using BOTTLENECK 1.2. 02 in wild populations: the importance of the microsatellite structure. *Conservation Genetics*, 11(3), pp.1043-1049.

Cullingham, C.I., Miller, J.M., Peery, R.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C. and Janes, J.K., 2020. Confidently identifying the correct K value using the ΔK method: When does K= 2?. *Molecular Ecology*, 29(5), pp.862-869.

Currat, M. and Excoffier, L., 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *Plos biol*, 2(12), p.e421.

Currat, M., Excoffier, L., Maddison, W., Otto, S.P., Ray, N., Whitlock, M.C. and Yeaman, S., 2006. Comment on" Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*" and" Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". *Science*, 313(5784), pp.172-172.

Cushman, S.A., McKelvey, K.S., Hayden, J. and Schwartz, M.K., 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, 168(4), pp.486-499.

Dallas, J.F. and Piertney, S.B., 1998. Microsatellite primers for the Eurasian otter. *Molecular Ecology*, 7(9), pp.1248-1251.

Dallas, J.F., Bacon, P.J., Carss, D.N., Conroy, J.W., 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(1-2), pp.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(1), pp.15-28.

Dallas, J.F., Coxon, K.E., Sykes, T., Chanin, P.R., 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), pp.275-282.

Dantzer, B. and Fletcher, Q.E., 2015. Telomeres shorten more slowly in slow-aging wild animals than in fast-aging ones. *Experimental Gerontology*, 71, pp.38-47.

De Barba, M., Miquel, C., Lobréaux, S., Quenette, P.Y., Swenson, J.E. and Taberlet, P., 2017. High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, 17(3), pp.492-507.

De Vos, J.M., Joppa, L.N., Gittleman, J.L., Stephens, P.R. and Pimm, S.L., 2015. Estimating the normal background rate of species extinction. *Conservation biology*, 29(2), pp.452-462.

Devictor, V., Whittaker, R.J. and Beltrame, C., 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*. 16: 354–362.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. and Freimer, N.B., 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences*, 91(8), pp.3166-3170.

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H., Chan, K.M. and Garibaldi, L.A., 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, 366(6471).

Díaz, S., Zafra-Calvo, N., Purvis, A., Verburg, P.H., Obura, D., Leadley, P., Chaplin-Kramer, R., De Meester, L., Dulloo, E., Martín-López, B., Shaw, M.R., Visconti, P., Broadgate, W., Bruford, M.W., Burgess, N.D., Cavender-Bares, J., DeClerck, F., Fernádez-Palacios, J.M., Garibaldi, L.A., Hill, S.L.L., Isbell, L.F., Khoury, C.K., Krug, C.B., Liu, J., Maron, M., McGowan, P.J.K., Pereira, H.M., Reyes-García, V., Rocha, J., Rondinini, C., Shannon, L., Shin, Y.J., Snelgrove, P.V.R., Spehn, E.M., Strassburg, B., Subramanian, S.M., Tewksbury, J.J., Watson, J.E.M. and Zanne, A.E., 2020. Set ambitious goals for biodiversity and sustainability. *Science*, 370(6515), pp.411-413.

Dickson, B.G., Albano, C.M., Anantharaman, R., Beier, P., Fargione, J., Graves, T.A., Gray, M.E., Hall, K.R., Lawler, J.J., Leonard, P.B. and Littlefield, C.E., 2019. Circuit-theory applications to connectivity science and conservation. *Conservation biology*, 33(2), pp.239-249.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. and Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular ecology resources*, 14(1), pp.209-214.

Donnelly, M.J., Licht, M.C. and Lehmann, T., 2001. Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Molecular biology and evolution*, 18(7), pp.1353-1364.

Dufresnes, C., Remollino, N., Stoffel, C., Manz, R., Weber, J.M. and Fumagalli, L., 2019. Two decades of non-invasive genetic monitoring of the grey wolves recolonizing the Alps support very limited dog introgression. *Scientific reports*, 9(1), pp.1-9.

D'Urban Jackson, J., Dos Remedios, N., Maher, K.H., Zefania, S., Haig, S., Oyler-McCance, S., Blomqvist, D., Burke, T., Bruford, M.W., Székely, T. and Küpper, C., 2017. Polygamy slows down population divergence in shorebirds. *Evolution*, 71(5), pp.1313-1326.

Durbin, L.S., 1998. Habitat selection by five otters *Lutra lutra* in rivers of northern Scotland. *Journal of Zoology*, 245, pp.85–92.

Earl D, vonHoldt B., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, pp.359–361.

Ellegren, H., 2000. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature genetics*, 24(4), pp.400-402.

England, P.R., Luikart, G. and Waples, R.S., 2010. Early detection of population fragmentation using linkage disequilibrium estimation of effective population size. *Conservation genetics*, 11(6), pp.2425-2430.

Environment Agency, 2015. *WFD River Basin Districts Cycle 2,* Environment Agency, viewed 12th December 2018, <https://data.gov.uk/dataset/368ae5fb-65a1-4f19-98ff-a06a1b86b3fe/wfd-river-basin-districts-cycle-2>

Epps, C.W., Palsbøll, P.J., Wehausen, J.D., Roderick, G.K., Ramey, R.R. and McCullough, D.R., 2005. Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology letters*, 8(10), pp.1029-1038.

Epps, C.W. and Keyghobadi, N., 2015. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular ecology*, 24(24), pp.6021-6040.

Erlinge, S., 1967. Home Range of the Otter *Lutra lutra L*. in Southern Sweden. *Oikos*. 18(2), 186-209.

Erlinge, S., 1968. Territoriality of the otter *Lutra lutra L*. *Oikos*, pp.81-98.

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

Excoffier, L., 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular ecology*, 13(4), pp.853-864.

Excoffier, L., and Ray, N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23(7), pp.347–351.

Excoffier, L., Foll, M. and Petit, R.J., 2009. Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics,* 40, pp.481–501.

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

Faubet, P., Waples, R.S. and Gaggiotti, O.E., 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular ecology*, 16(6), pp.1149-1166.

Fedorca, A., Russo, I.R.M., Ionescu, O., Ionescu, G., Popa, M., Fedorca, M., Curtu, A.L., Sofletea, N., Tabor, G.M. and Bruford, M.W., 2019. Inferring fine-scale spatial structure of the brown bear (*Ursus arctos*) population in the Carpathians prior to infrastructure development. *Scientific reports*, 9(1), pp.1-12.

Feld, C.K., Martins da Silva, P., Paulo Sousa, J., De Bello, F., Bugter, R., Grandin, U., Hering, D., Lavorel, S., Mountford, O., Pardo, I. and Pärtel, M., 2009. Indicators of biodiversity and ecosystem services: a synthesis across ecosystems and spatial scales. *Oikos*, 118(12), pp.1862-1871.

Findlay, M., Alexander, L. and Macleod, C., 2015. *Site condition monitoring for otters (Lutra lutra) in 2011-12*. Scottish Natural Heritage.

Fischer, J., Dyball, R., Fazey, I., Gross, C., Dovers, S., Ehrlich, P.R., Brulle, R.J., Christensen, C. and Borden, R.J., 2012. Human behavior and sustainability. *Frontiers in Ecology and the Environment*, 10(3), pp.153-160.

Fisher, R.A., 1930. *The genetical theory of natural selection*. Oxford: Clarendon.

Fordyce, S.L., Mogensen, H.S., Børsting, C., Lagacé, R.E., Chang, C.W., Rajagopalan, N. and Morling, N., 2015. Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. *Forensic Science International: Genetics*, 14, pp.132-140.

Fossel, M., 2012. Use of telomere length as a biomarker for aging and age-related disease. *Current Translational Geriatrics and Experimental Gerontology Reports*, 1(2), pp.121-127.

Frankham, R., 1995a. Conservation genetics. *Annual Review of Genetics*. 29, pp.305–327.

Frankham, R., 1995b. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research*, 66(2), pp.95-107.

Frankham, R., 1996. Relationship of genetic variation to population size in wildlife. *Conservation biology*, 10(6), pp.1500-1508.

Frankham, R., 2010. Challenges and opportunities of genetic approaches to biological conservation. *Biological conservation*, 143(9), pp.1919-1927.

Franklin, I.R., 1980. Evolutionary change in small populations. In M. E. Soulé, & B. A. Wilcox (Eds.), *Conservation biology: An evolutionary-ecological perspective* (pp. 135–149). Sunderland, MA: Sinauer Associates.

Garnier-Géré, P. and Chikhi, L., 2001. Population subdivision, Hardy–Weinberg equilibrium and the Wahlund effect. *eLS*.

Gauthier, J., Pajkovic, M., Neuenschwander, S., Kaila, L., Schmid, S., Orlando, L. and Alvarez, N., 2020. Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Molecular ecology resources*, 20(5), pp.1191-1205.

Gilbert, K.J. and Whitlock, M.C., 2015. Evaluating methods for estimating local effective population size with and without migration. *Evolution*, 69(8), pp.2154-2166.

Girod, C., Vitalis, R., Leblois, R. and Fréville, H., 2011. Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. *Genetics*, 188(1), pp.165-179.

Gómez-Sánchez, D., Olalde, I., Sastre, N., Enseñat, C., Carrasco, R., Marques-Bonet, T., Lalueza-Fox, C., Leonard, J.A., Vilà, C. and Ramírez, O., 2018. On the path to extinction: inbreeding and admixture in a declining grey wolf population. *Molecular ecology*, 27(18), pp.3599-3612.

Graciá, E., Botella, F., Anadón, J.D., Edelaar, P., Harris, D.J. and Giménez, A., 2013. Surfing in tortoises? Empirical signs of genetic structuring owing to range expansion. *Biology letters*, 9(3), p.20121091.

Green, J. and Green, R., 1980. *Otter Survey of Scotland 1977-79*. Vincent Wildlife Trust, London.

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

Green, R., 1997. Reintroduction of Otters a Successful Conservation Strategy. *IUCN Otter Specialist Group Bulletin*, 14(2), pp.62-74.

Goossens, B., Chikhi, L., Ancrenaz, M., Lackman-Ancrenaz, I., Andau, P. and Bruford, M.W., 2006. Genetic signature of anthropogenic population collapse in orang-utans. *PLoS Biol*, 4(2), p.e25.

Goudet, J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of heredity*, 86(6), pp.485-486.

Hagen, S.B., Kopatz, A., Aspi, J., Kojola, I. and Eiken, H.G., 2015. Evidence of rapid change in genetic structure and diversity during range expansion in a recovering large terrestrial carnivore. *Proceedings of the Royal Society B: Biological Sciences*, 282(1807), p.20150092.

Hájková, P., Pertoldi, C., Zemanová, B., Roche, K., Hajek, B., Bryja, J. and Zima, J., 2007. Genetic structure and evidence for recent population decline in Eurasian otter populations in the Czech and Slovak Republics: implications for conservation. *Journal of Zoology*, 272(1), pp.1-9.

Hallatschek, O., Hersen, P., Ramanathan, S. and Nelson, D.R., 2007. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50), pp.19926-19930.

Hansen, M.M., 2002. Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: An approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology*, 11(6), pp.1003-1015.

Hardy, G.H., 1908. Mendelian proportions in a mixed population. *Science,* 18, pp. 49-50.

Harris, R.B. and Allendorf, F.W., 1989. Genetically effective population size of large mammals: an assessment of estimators. *Conservation Biology*, 3(2), pp.181-191.

Harris, S., Cresswell, W.J. and Cheeseman, C.L., 1992. Age determination of badgers (*Meles meles*) from tooth wear: the need for a pragmatic approach. *Journal of Zoology*, 228(4), pp.679-684.

Hauer, S., Ansorge, H. and Zinke, O., 2002. Mortality patterns of otters (*Lutra lutra*) from eastern Germany. *Journal of Zoology*, 256(3), pp.361-368.

Haussmann, M.F., Vleck, C.M. and Nisbet, I.C., 2003. Calibrating the telomere clock in common terns, *Sterna hirundo*. *Experimental Gerontology*, 38(7), pp.787-789.

Heggberget, T.M., 1984. Age determination in the European otter *Lutra lutra lutra*. *Zeitschrift für Säugetierkunde*, 49(5), pp.299-305.

Heppenheimer, E., Cosio, D.S., Brzeski, K.E., Caudill, D., Van Why, K., Chamberlain, M.J. and Hinton, J.W., 2018. Demographic history influences spatial patterns of genetic diversity in recently expanded coyote (*Canis latrans*) populations. *Heredity*, 120(3), p.183.

Herfindal, I., Linnell, J.D., Odden, J., Nilsen, E.B. and Andersen, R., 2005. Prey density, environmental productivity and home-range size in the Eurasian lynx (*Lynx lynx*). *Journal of Zoology*, 265(1), pp.63-71.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*. 405, pp.907–913.

Hoban, S.M., Hauffe, H.C., Pérez-Espona, S., Arntzen, J.W., Bertorelle, G., Bryja, J., Frith, K., Gaggiotti, O.E., Galbusera, P., Godoy, J.A. and Hoelzel, A.R., 2013. Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources*, 5(2), pp.593-598.

Hoban, S., Arntzen, J.A., Bruford, M.W., Godoy, J.A., Rus Hoelzel, A., Segelbacher, G., Vilà, C. and Bertorelle, G., 2014. Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evolutionary applications*, 7(9), pp.984-998.

Hoban, S., Bruford, M., Jackson, J.D.U., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P.A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C. and Aitken, S., 2020. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, 248, p.108654.

Hoban, S., Paz-Vinas, I., Aitken, S., Bertola, L., Breed, M.F., Bruford, M., Funk, C., Grueber, C., Heuertz, M., Hohenlohe, P. and Hunter, M., 2021a. Effective population size remains a suitable, pragmatic indicator of genetic diversity for all species, including forest trees. *Biological Conservation*, 253, p.108906.

Hoban, S., Bruford, M.W., Funk, W.C., Galbusera, P., Griffith, M.P., Grueber, C.E., Heuertz, M., Hunter, M.E., Hvilsom, C., Kalamujic Stroil, B. and Kershaw, F., Khoury, C.K., Laikre, L., Lopesfernandes, M., Macdonald, A.J., Mergeay, J., Meek, M., Mittan, C., Mukassabi, T.A., O'Brien, D., Ogden, R., Palma-Silva, C., Ramakrishnan, U., Segelbacher, G., Shaw, R.E, Sjögren-Gulve, P., Veličković, N. and Vernesi, C., 2021b. Global commitments to conserving and monitoring genetic diversity are now necessary and feasible. *Bioscience*.

Hobbs, G., Chadwick, E., Slater, F. and Bruford, M., 2006. Landscape genetics applied to a recovering otter (*Lutra lutra*) population in the UK: preliminary results and potential methodologies. *Hystrix, the Italian Journal of Mammalogy*, 17(1).

Hobbs, G.I., 2010. *Population genetic structure of a recovering otter (Lutra lutra) population in the UK.* PhD Thesis. Cardiff University.

Hobbs, G.I., Chadwick, E.A., Bruford, M.W. and Slater, F.M., 2011. Bayesian clustering techniques and progressive partitioning to identify population structuring within a recovering otter population in the UK. *Journal of Applied Ecology*, 48(5), pp.1206-1217.

Hodgson, J.C., Baylis, S.M., Mott, R., Herrod, A. and Clarke, R.H., 2016. Precision wildlife monitoring using unmanned aerial vehicles. *Scientific reports*, 6(1), pp.1-7.

Hohnen, R., Tuft, K.D., Legge, S., Hillyer, M., Spencer, P.B., Radford, I.J., Johnson, C.N. and Burridge, C.P., 2016. Rainfall and topography predict gene flow among populations of the declining northern quoll (*Dasyurus hallucatus*). *Conservation Genetics*, 17(5), pp.1213-1228.

Holderegger, R. and Wagner, H.H., 2008. Landscape genetics. *BioScience*. 58, pp.199–207.

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), pp.314-316.

Ibrahim, K.M., Nichols, R.A. and Hewitt, G.M., 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77(3), pp.282-291.

Jackson, D.L., 2000. Guidance on the Interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater): Definitions and the Relationship with other Habitat Classifications. *Peterborough, UK: JNCC*.

Jamieson, I.G. and Allendorf, F.W., 2012. How does the 50/500 rule apply to MVPs?. *Trends in ecology & evolution*, 27(10), pp.578-584.

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

Janssens, X., Fontaine, M.C., Michaux, J.R., Libois, R., de Kermabon, J., Defourny, P. and Baret, F.V., 2008. Genetic pattern of the recent recovery of European otters in southern France. *Ecography*. 31: 176-186.

Jefferies, D.J., Wayre, P. & Shuter, R., 2000. A brief history of the Otter Trust's successful programme of repopulating lowland England with otters bred in captivity with a special emphasis on East Anglia. *Otters, Journal of the Otter Trust*, 2000 3(4): 105-117.

Jenkins, D., 1980. Ecology of otters in northern Scotland: I otter (*Lutra lutra*) breeding and dispersion in mid Deeside, Aberdeenshire, in 1974-79. *Journal of Animal Ecology*. 49, 713–735.

Jessop, R.M. and Cheyne, D.L., 1992. The reintroduction of European otter into lowland England carried out by the Otter Trust 1983–1992: a progress report. *Otters: the Journal of the Otter Trust*, pp.11-16.

Jin, L. and Chakraborty, R., 1995. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity*, 74(3), p.274.

Jombart, T., Pontier, D. and Dufour, A.B., 2009. Genetic markers in the playground of multivariate analysis. *Heredity*, 102(4), p.330.

Jombart, T., Devillard, S. and Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11(1), pp.1-15. doi:10.1186/1471-2156-11-94

Jones, T. and Jones, D., 2004. *Otter survey of Wales 2002*. Environment Agency, Bristol, UK.

Jorde, P.E. and Ryman, N., 2007. Unbiased estimator for genetic drift and effective population size. *Genetics*, 177(2), pp.927-935.

Kalinowski, S.T., 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5(1), pp.187-189.

Kean, E.F., Müller, C.T. and Chadwick, E.A., 2011. Otter scent signals age, sex, and reproductive status. *Chemical Senses*, 36(6), pp.555-564.

Kean, E.F., Bruford, M.W., Russo, I-R.M., Müller, C.T. and Chadwick, E.A., 2017. Odour dialects among wild mammals. *Scientific Reports*, 7(1), pp.1-6.

Keller, L.F., Jeffery, K.J., Arcese, P., Beaumont, M.A., Hochachka, W.M., Smith, J.N. and Bruford, M.W., 2001. Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1474), pp.1387-1394.

Kendall, K.C., Stetz, J.B., Boulanger, J.B., Mcleod, A.C., Paetkau, D. and White, G.C., 2009. Demography and genetic structure of a recovering grizzly bear population. *Journal of Wildlife Management*. 73, pp.3–17.

Kendall, K.C., Graves, T.A., Royle, J.A., Macleod, A.C., McKelvey, K.S., Boulanger, J. and Waller, J.S., 2019. Using bear rub data and spatial capture-recapture models to estimate trend in a brown bear population. *Scientific reports*, 9(1), pp.1-11.

Keyghobadi, N., 2007. The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology*, 85(10), pp.1049-1064.

Kim, K.S. and Sappington, T.W., 2013. *Microsatellite data analysis for population genetics. In Microsatellites* (pp. 271-295). Humana Press, Totowa, NJ.

Kimura, M. and Crow, J.F., 1963. The measurement of effective population number. *Evolution*, pp.279-288.

King, C.M., 1991. A review of age determination methods for the stoat *Mustela erminea*. *Mammal Review*, 21(1), pp.31-49.

Kingman, J.F.C., 1982. The coalescent. *Stochastic processes and their applications*, 13(3), pp.235-248.

Kivimäki, I., Shimbo, M. and Saerens, M., 2014. Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A: Statistical Mechanics and its Applications*, 393, pp.600-616.

Kopatz, A., Eiken, H.G., Schregel, J., Aspi, J., Kojola, I. and Hagen, S.B., 2017. Genetic substructure and admixture as important factors in linkage disequilibrium-based estimation of effective number of breeders in recovering wildlife populations. *Ecology and evolution*, 7(24), pp.10721-10732.

Kopatz, A., Kleven, O., Kojola, I., Aspi, J., Norman, A.J., Spong, G., Gyllenstrand, N., Dalén, L., Fløystad, I., Hagen, S.B. and Kindberg, J., 2021. Restoration of transborder connectivity for Fennoscandian brown bears (*Ursus arctos*). *Biological Conservation*, 253, pp.108936.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. and Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*, 15(5), pp.1179-1191.

Kozakiewicz, C.P., Ricci, L., Patton, A.H., Stahlke, A.R., Hendricks, S.A., Margres, M.J., Ruiz-Aravena, M., Hamilton, D.G., Hamede, R., McCallum, H. and Jones, M.E., 2020. Comparative landscape genetics reveals differential effects of environment on host and pathogen genetic structure in Tasmanian devils (*Sarcophilus harrisii*) and their transmissible tumour. *Molecular Ecology*, 29(17), pp.3217-3233.

Kraus, R.H., Vonholdt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., Förster, D.W., Fickel, J., Roos, C. and Nowak, C., 2015. A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in E urope based on noninvasively collected samples. *Molecular ecology resources*, 15(2), pp.295-305.

Kruuk, H. and Conroy, J.W.H., 1987. Surveying otter *Lutra lutra* populations: a discussion of problems with spraints. *Biological Conservation*, 41, pp.179–183.

Kruuk, H., 2006. *Otters: Ecology, Behaviour and Conservation*. Oxford University Press, Oxford.

Laikre, L., Hoban, S., Bruford, M.W., Segelbacher, G., Allendorf, F.W., Gajardo, G., Rodríguez, A.G., Hedrick, P.W., Heuertz, M., Hohenlohe, P.A., Jaffé, R., Johannesson, K., Liggins, L., MacDonald, A.J., Orozco-terWengel, P., Reusch, T.B.H., Rodríguez-Correa, H., Russo, I.M., Ryman, N. and Vernesi, C., 2020. Post-2020 goals overlook genetic diversity. *Science*, 367(6482), pp.1083-1085.

Laikre, L., Hohenlohe, P.A., Allendorf, F.W., Bertola, L.D., Breed, M.F., Bruford, M.W., Funk, W.C., Gajardo, G., González-Rodríguez, A., Grueber, C.E., Hedrick, P.W., Heuertz, M., Hunter, M.E., Johannesson, K., Liggins, L., MacDonald, A.J., Mergeay, J., Moharrek, F., O'Brien, D., Ogden, R., Orozco-terWengel, P., Palma-Silva, C., Pierson, J., Paz-Vinas, I., Russo, I.M., Ryman, N., Segelbacher, G., Sjögren-Gulve, P., Waits, L.P., Vernesi, C. and Hoban S., 2021. Authors' Reply to Letter to the Editor: Continued improvement to genetic diversity indicator for CBD. *Conservation Genetics*, pp.1-4.

Lande, R., 1988. Genetics and demography in biological conservation. *Science*, 241(4872), pp.1455-1460.

Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M. and Luikart, G., 2010. Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, 19(19), pp.4179-4191.

Lanszki, J., Hidas, A., Szentes, K., Révay, T., Lehoczky, I. and Weiss, S., 2008. Relative spraint density and genetic structure of otter (*Lutra lutra*) along the Drava River in Hungary. *Mammalian Biology*, 73(1), pp.40-47.

Leberg, P., 2005. Genetic approaches for estimating the effective size of populations. *The Journal of Wildlife Management*, 69(4), pp.1385-1399.

Leigh, D.M., Hendry, A.P., Vázquez-Domínguez, E. and Friesen, V.L., 2019. Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, 12(8), pp.1505-1512.

Lenton, E.J., Chanin, P.R. and Jefferies, B.J., 1980. *Otter survey of England 1977-79*. Nature Conservancy Council, London, UK.

Lessa, E.P., Cook, J.A. and Patton, J.L., 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences of the United States of America*, 100, pp.10331–10334.

Levin, S.A., 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, 73(6), pp.1943-1967.

Lonsinger, R.C., Adams, J.R. and Waits, L.P., 2018. Evaluating effective population size and genetic diversity of a declining kit fox population using contemporary and historical specimens. *Ecology and Evolution*, 8(23), pp.12011-12021.

Luikart, G., Allendorf, F.W., Cornuet, J.M. and Sherwin, W.B., 1998a. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of heredity*, 89(3), pp.238-247.

Luikart, G., Allendorf, F.W., Piry, S. and Cornuet, J.M., 1998b. Molecular genetic test identifies endangered populations. *Conservation Biology*, 12(1), pp.228-237.

Luikart, G. and Cornuet, J.M., 1998c. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation biology*, 12(1), pp.228-237.

Luikart, G., Ryman, N., Tallmon, D.A., Schwartz, M.K. and Allendorf, F.W., 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics*, 11(2), pp.355-373.

Ma, T., Hu, Y., Russo, I.R.M., Nie, Y., Yang, T., Xiong, L., Ma, S., Meng, T., Han, H., Zhang, X. and Bruford, M.W., 2018. Walking in a heterogeneous landscape: Dispersal, gene flow and conservation implications for the giant panda in the Qinling Mountains. *Evolutionary applications*, 11(10), pp.1859-1872.

Macdonald, S.M. and Mason, C.F., 1983. Some factors influencing the distribution of otters (*Lutra lutra*). *Mammal review*, 13(1), pp.1-10.

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), pp.189–197.

Marrotte, R.R. and Bowman, J., 2017. The relationship between least-cost and resistance distance. *PloS one*, 12(3), p.e0174212.

Marsh, D.M. and Trenham, P.C., 2008. Current trends in plant and animal population monitoring. *Conservation Biology*, 22(3), pp.647-655.

Martin, T.G., Nally, S., Burbidge, A.A., Arnall, S., Garnett, S.T., Hayward, M.W., Lumsden, L.F., Menkhorst, P., McDonald-Madden, E. and Possingham, H.P., 2012. Acting fast helps avoid extinction. *Conservation Letters*, 5(4), pp.274-280.

Maruyama, T. and Fuerst, P.A., 1984. Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. *Genetics*, 108(3), pp.745-763.

Maruyama, T. and Fuerst, P.A., 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics*, 111(3), pp.675-689.

Mason, C.F. and Macdonald, S.M., 1986. *Otters: ecology and conservation*. Cambridge University Press.

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), pp.148–152.

Mathews, F., Kubasiewicz, L.M., Gurnell, J., Harrower, C.A., McDonald, R.A. and Shore, R.F., 2018. *A review of the population and conservation status of British mammals*. Natural England.

Matthysen, E., 2005. Density-dependent dispersal in birds and mammals. *Ecography*, 28(3), pp.403-416.

McRae, B.H., Dickson, B.G., Keitt, T.H. and Shah, V.B., 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, 89(10), pp.2712-2724.

McRae, B.H., Shah, V.B. and Mohapatra, T.K., 2013. *Circuitscape 4 user guide*. The nature conservancy. https://www.circuitscape.org

Miller, J.M., Cullingham, C.I. and Peery, R.M., 2020. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity*, 125(5), pp.269-280.

Mills, L.S. and Allendorf, F.W., 1996. The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology*, 10, pp.1509–1518.

Monteith, D.T., Hildrew, A.G., Flower, R.J., Raven, P.J., Beaumont, W.R.B., Collen, P., Kreiser, A.M., Shilland, E.M. and Winterbottom, J.H., 2005. Biological responses to the chemical recovery of acidified fresh waters in the UK. *Environmental Pollution*, 137(1), pp.83-101.

Moreau, C., Bhérer, C., Vézina, H., Jomphe, M., Labuda, D. and Excoffier, L., 2011. Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science*, 334(6059), pp.1148-1150.

Morton, R.D., Rowland, C.S., Wood, C.M., Meek, L., Marston, C.G. and Smith, G.M., 2014. Land Cover Map 2007 (25m raster, GB) v1.2. NERC Environmental Information Data Centre. https://doi.org/10.5285/a1f88807-4826-44bc-994d-a902da5119c2

Mucci, N., Arrendal, J., Ansorge, H., Bailey, M., Bodner, M., Delibes, M., Ferrando, A., Fournier, P., Fournier, C., Godoy, J.A. and Hajkova, P., 2010. Genetic diversity and landscape genetic structure of otter (*Lutra lutra*) populations in Europe. *Conservation Genetics*, 11(2), pp.583-599.

Natural Resources Wales 2015, *Water Framework Directive (WFD) River Basin Districts Cycle 2*, Natural Resources Wales, viewed 12th December 2018,

<https://lle.gov.wales/catalogue/item/WaterFrameworkDirectiveRiverBasinDistrictsCycle2/?lang=en>

Natural Resources Wales, 2020. *State of Natural Resources Report (SoNaRR) for Wales 2020*. Natural Resources Wales.

Neel, M.C., McKelvey, K., Ryman, N., Lloyd, M.W., Bull, R.S., Allendorf, F.W., Schwartz, M.K. and Waples, R.S., 2013. Estimation of effective population size in continuously distributed populations: there goes the neighborhood. *Heredity*, 111(3), pp.189-199.

Nei, M., Maruyama, T. and Chakraborty, R., 1975. The bottleneck effect and genetic variability in populations. *Evolution*, pp.1-10.

Nei, M., 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.

Néill, L.Ó., Veldhuizen, T., de Jongh, A. and Rochford, J., 2009. Ranging behaviour and socio-biology of Eurasian otters (*Lutra lutra*) on lowland mesotrophic river systems. *European Journal of Wildlife Research*, 55(4), pp.363-370.

Nilsson, T., 2013. *Population viability analyses of the Scandinavian populations of bear (Ursus arctos), lynx (Lynx lynx) and wolverine (Gulo gulo).* Swedish Environmental Protection Agency [Naturvårdsverket].

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N. and Clune, J., 2021. A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, 12(1), pp.150-161.

Nussey, D.H., Coltman, D.W., Coulson, T., Kruuk, L.E., Donald, A., Morris, S.J., Clutton-Brock, T.H. and Pemberton, J., 2005. Rapidly declining fine-scale spatial genetic structure in female red deer. *Molecular Ecology*. 14: 3395–3405.

O'Riain, M.J., Jarvis, J.U. and Faulkes, C.G., 1996. A dispersive morph in the naked mole-rat. *Nature*, 380(6575), pp.619-621.

Olah, G., Theuerkauf, J., Legault, A., Gula, R., Stein, J., Butchart, S., O'Brien, M. and Heinsohn, R., 2018. Parrots of Oceania–a comparative study of extinction risk. *Emu-Austral Ornithology*, 118(1), pp.94-112.

Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Marasini, G.G., Visconti, P. and Rondinini, C., 2013. Generation length for mammals. *Nature Conservation*, 5, p.89.

Pagacz, S., 2016. The effect of a major drainage divide on the gene flow of a semiaquatic carnivore, the Eurasian otter. *Journal of Mammalogy,* 97(4), pp.1164-1176.

Palstra, F.P. and Ruzzante, D.E., 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular ecology*, 17(15), pp.3428-3447.

Palstra, F.P. and Fraser, D.J., 2012. Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and evolution*, 2(9), pp.2357-2365.

Parham, J., Crall, J., Stewart, C., Berger-Wolf, T. and Rubenstein, D.I., 2017, January. Animal population censusing at scale with citizen science and photographic identification. In *AAAI Spring Symposium-Technical Report*.

Park, S.D.E., 2001. *Trypanotolerance in West African cattle and the population genetic effects of selection*. Ph.D. thesis, University of Dublin

Peakall, R.O.D. and Smouse, P.E., 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), pp.288-295.

Peakall, R.O.D. and Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28(19), p.2537.

Pérez-Espona, S., Pérez-Barbería, F.J., McLeod, J.E., Jiggins, C.D., Gordon, I.J. and Pemberton, J.M., 2008. Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology*, 17(4), pp.981-996.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C. and Coops, N.C., 2013. Essential biodiversity variables. *Science*, 339(6117), pp.277-278.

Pertoldi, C., Loeschcke, V., Randi, E., Bo Madsen, A., Hansen, M.M., Bijlsma, R. and Van De Zande, L., 2005. Present and past microsatellite variation and assessment of genetic structure in Eurasian badger (*Meles meles*) in Denmark. *Journal of Zoology*, 265(4), pp.387-394.

Peterman, W.E., Connette, G.M., Semlitsch, R.D. and Eggert, L.S., 2014. Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, 23(10), pp.2402-2413.

Peterman, B., 2017 *A Vignette/Tutorial to use ResistanceGA.* http://petermanresearch.weebly.com/uploads/2/5/9/2/25926970/resistancega.pdf

Peterman, W.E., 2018. RESISTANCEGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9(6), 1638–1647.

Peterson, D.L. and Parker, V.T., 1998. *Dimensions of scale in ecology, resource management, and society*. In Peterson, D.L. and Parker, V.T., editors, Ecological scale: theory and applications, New York: Columbia University Press, pp.499–522.

Petkova, D., Novembre, J. and Stephens, M., 2016. Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48(1), pp.94-100.

Pfaff, C.L., Parra, E.J., Bonilla, C.A.R.O.L.I.N.A., Hiester, K., McKeigue, P.M., Kamboh, M.I., Hutchinson, R.G., Ferrell, R.E., Boerwinkle, E. and Shriver, M.D., 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *The American Journal of Human Genetics*, 68(1), pp.198-207.

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(5), pp.748-761.

Pierson, J., Luikart, G. and Schwartz, M., 2015. The application of genetic indicators in wild populations: potential and pitfalls for genetic monitoring. *Indicators and surrogates of biodiversity and environmental change (DB Lindenmayer, P. Barton, and JC Pierson, eds.)*. CSIRO Publishing, Melbourne, Australia, pp.149-159.

Pilot, M., Jedrzejewski, W., Branicki, W., Sidorovich, V.E., Jedrzejewska, B., Stachura, K. and Funk, S.M., 2006. Ecological factors influence population genetic structure of European grey wolves. *Molecular ecology*, 15(14), pp.4533-4553.

Pimentel, J.S., Carmo, A.O., Rosse, I.C., Martins, A.P., Ludwig, S., Facchin, S., Pereira, A.H., Brandão-Dias, P.F., Abreu, N.L. and Kalapothakis, E., 2018. High-throughput sequencing strategy for microsatellite genotyping using neotropical fish as a model. *Frontiers in Genetics*, 9, p.73.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M. and Sexton, J.O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *science*, 344(6187).

Piry, S., Luikart, G. and Cornuet, J.M., 1999. Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of heredity*, 90(4), pp.502-503.

Poulsen, N.A., Nielsen, E.E., Schierup, M.H., Loeschcke, V. and Grønkjaer, P., 2006. Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). *Molecular Ecology*, 15(2), pp.321-331.

Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945–959.

Pritchard, J.K. and Przeworski, M., 2001. Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics*, 69(1), pp.1-14.

Pritchard, J. K. and Wen, W., 2003. *Documentation for STRUCTURE Software: Version 2*. Chicago: University of Chicago Press. Retrived from: http://web.stanford.edu/group/pritchardlab/software/structure2_1.html.

Pruett, C.L. and Winker, K., 2008. The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. *Journal of Avian Biology*, 39(2), pp.252-256.

Quaglietta, L., Fonseca, V.C., Hájková, P., Mira, A. and Boitani, L., 2013. Fine-scale population genetic structure and short-range sex-biased dispersal in a solitary carnivore, *Lutra lutra*. *Journal of Mammalogy*, 94(3), pp.561-571.

Quaglietta, L., Hájková, P., Mira, A. and Boitani, L., 2015. Eurasian otter (*Lutra lutra*) density estimate based on radio tracking and other data sources. *Mammal Resources*, 60, pp.127–137.

R Core Team, 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A. and Drummond, A., 2003. *Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes.* University of Edinburgh, UK.

Randi, E., Davoli, F., Pierpaoli, M., Pertoldi, C., Madsen, A.B. and Loeschcke, V., 2003. Genetic structure in otter (*Lutra lutra*) populations in Europe: implications for conservation. *Animal Conservation*, 6(2), pp.93-100.

Rannala, B., 2007. *BayesAss edition 3.0 user's manual*. University of California, Davis.

Ray, N., Currat, M. and Excoffier, L., 2003. Intra-deme molecular diversity in spatially expanding populations. *Molecular biology and evolution*, 20(1), pp.76-86.

Raymond, M. and Rousset, F., 1995. An exact test for population differentiation. *Evolution,* 49, pp.1280-1283.

Reed, D.H., 2004. Extinction risk in fragmented habitats. *Animal Conservation*, 7, pp.181–191.

Riley, S.P., Pollinger, J.P., Sauvajot, R.M., York, E.C., Bromley, C., Fuller, T.K. and Wayne, R.K., 2006. FAST-TRACK: A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular ecology*, 15(7), pp.1733-1741.

Robitaille, J.F. and Laurence, S., 2002. Otter, *Lutra lutra*, occurrence in Europe and in France in relation to landscape characteristics. *Animal Conservation*, 5(4), pp.337-344.

Roos, A., Loy, A., de Silva, P., Hajkova, P. and Zemanová, B., 2015. *Lutra lutra. The IUCN Red List of Threatened Species 2015*: e.T12419A21935287.

Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources*, 8(1), pp.103-106.

Rueness, E.K., Jorde, P.E., Hellborg, L., Stenseth, N.C., Ellegren, H. and Jakobsen, K.S., 2003. Cryptic population structure in a large, mobile mammalian predator: the Scandinavian lynx. *Molecular Ecology*, 12(10), pp.2623-2633.

Ruiz-Olmo, J., 1998. Influence of altitude on the distribution, abundance and ecology of the otter (*Lutra lutra*). In *Behaviour and ecology of riparian mammals*: 114-116. Dunstone, N. and Gorman, M., (Eds). Cambridge: Cambridge University Press.

Ruiz-Olmo, J., López-Martín, J.M. and Palazón, S., 2001. The influence of fish abundance on the otter (*Lutra lutra*) populations in Iberian Mediterranean habitats. *Journal of Zoology*, 254(3), pp.325-336.

Russo, I.R.M., Sole, C.L., Barbato, M., Von Bramann, U. and Bruford, M.W., 2016. Landscape determinants of fine-scale genetic structure of a small rodent in a heterogeneous landscape (Hluhluwe-iMfolozi Park, South Africa). *Scientific reports*, 6(1), pp.1-14.

Rutledge, L.Y., Garroway, C.J., Loveless, K.M. and Patterson, B.R., 2010. Genetic differentiation of eastern wolves in Algonquin Park despite bridging gene flow between coyotes and grey wolves. *Heredity,* 105(6), pp.520–531.

Rybicki, B.A., Iyengar, S.K., Harris, T., Liptak, R., Elston, R.C., Maliarik, M.J. and Iannuzzi, M.C., 2002. Prospects of admixture linkage disequilibrium mapping in the African-American genome. *Cytometry: The Journal of the International Society for Analytical Cytology*, 47(1), pp.63-65.

Sacks, B.N., Brown, S.K. and Ernest, H.B., 2004. Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, 13(5), pp.1265-1275.

Schneider, D.C., 1994. *Quantitative Ecology: Spatial and Temporal Scaling*. Academic Press, San Diego, California.

Schregel, J., Kopatz, A., Eiken, H.G., Swenson, J.E. and Hagen, S.B., 2017. Sex-specific genetic analysis indicates low correlation between demographic and genetic connectivity in the Scandinavian brown bear (*Ursus arctos*). *PLoS ONE*, 12(7), p.e0180701.

Schwartz, M.K., Luikart, G. and Waples, R.S., 2007. Genetic monitoring as a promising tool for conservation and management. *Trends in ecology & evolution*, 22(1), pp.25-33.

Schwartz, M.K. and McKelvey, K.S., 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, 10(2), p.441.

Schwartz, M.W., 2020. Conservation lessons from taboos and trolley problems. *Conservation Biology*.

Scrucca, L., 2013. GA: a package for genetic algorithms in R. *Journal of Statistical Software*, 53(4), pp.1-37.

Selkoe, K.A. and Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, 9(5), pp.615-629.

Sherrard-Smith, E. and Chadwick, E.A., 2010. Age structure of the otter (*Lutra lutra*) population in England and Wales, and problems with cementum ageing. *IUCN Otter Specialist Group Bulletin*, 27(1), pp.42-49.

Shirk, A.J., Landguth, E.L. and Cushman, S.A., 2018. A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, 18(1), pp.55-67.

Sidorovich, V.E., 1991. Structure, reproductive status and dynamics of the otter population in Byelorussia. *Acta Theriologica*, 36(1-2), pp.153-161.

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

Slatkin, M., 1994. Linkage disequilibrium in growing and stable populations. *Genetics*, 137(1), pp.331-336.

Stanton, D.W.G., Hobbs, G.I., Chadwick, E.A., Slater, F.M. and Bruford, M.W., 2009. Mitochondrial genetic diversity and structure of the European otter (*Lutra lutra*) in Britain. *Conservation Genetics*, 10(3), pp.733-737.

Stanton, D.W., Hobbs, G.I., McCafferty, D.J., Chadwick, E.A., Philbey, A.W., Saccheri, I.J., Slater, F.M. and Bruford, M.W., 2014. Contrasting genetic structure of the Eurasian otter (*Lutra lutra*) across a latitudinal divide. *Journal of Mammalogy*, 95(4), pp.814-823.

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A. and Folke, C., 2015. Planetary boundaries: Guiding human development on a changing planet. *Science*, 347(6223).

Strachan, R., Birks, J.D.S., Chanin, P.R.F., Jefferies, D.J., 1990. *Otter Survey of England 1984–1986*. JNCC, Peterborough, UK.

Strachan, R. and Jefferies, D.J., 1996. *Otter Survey of England 1991-1994: A Report on the Decline and Recovery of the Otter in England and on its Distribution, Status and Conservation in 1991-1994*. Vincent Wildlife Trust.

Strachan, R., 2015. *Otter survey of Wales 2009-10*. Natural Resources Wales. https://naturalresources.wales/media/4590/osw-5-english-24-06-2015.pdf

Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L. and Waits, L.P., 2007. Putting the 'landscape' in landscape genetics. *Heredity*, 98, pp.128–142.

Tajima, F., 1989. The effect of change in population size on DNA polymorphism. *Genetics*, 123(3), pp.597-601.

Tallmon, D.A., Koyuk, A., Luikart, G. and Beaumont, M.A., 2008. COMPUTER PROGRAMS: onesamp: a program to estimate effective population size using approximate Bayesian computation. *Molecular ecology resources*, 8(2), pp.299-301.

Tallmon, D.A., Waples, R.S., Gregovich, D. and Schwartz, M.K., 2012. Detecting population recovery using gametic disequilibrium-based effective population size estimates. *Conservation Genetics Resources*, 4(4), pp.987-989.

Trowbridge, B.J., 1983. *Olfactory communication in the European otter (Lutra l. lutra)*. Doctoral dissertation, University of Aberdeen.

Van Doornik, D.M., Waples, R.S., Baird, M.C., Moran, P. and Berntson, E.A., 2011. Genetic monitoring reveals genetic stability within and among threatened Chinook salmon populations in the Salmon River, Idaho. *North American Journal of Fisheries Management*, 31(1), pp.96-105.

Van Etten, J., 2017. R package gdistance: distances and routes on geographical grids. *Journal of Statistical Software*, 76(1), pp.1-21.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P. and Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3), pp.535-538.

Van Strien, M.J., Keller, D. and Holderegger, R., 2012. A new analytical approach to landscape genetic modelling: Least-cost transect analysis and linear mixed models. *Molecular ecology*, 21(16), pp.4010-4023.

Vercayie, D. and Herremans, M., 2015. Citizen science and smartphones take roadkill monitoring to the next level. In: Seiler A, Helldin J-O (Eds) Proceedings of IENE 2014 International Conference on Ecology and Transportation, Malmö, Sweden. IENE 2014. *Nature Conservation*. 11: 29–40.

Verner, J., 1985. Assessment of counting techniques. In *Current ornithology* (pp. 247-302). Springer, Boston, MA.

Vucetich, J.A., Waite, T.A. and Nunney, L., 1997. Fluctuating population size and the ratio of effective to census population size. *Evolution*, pp.2017-2021.

Wahlund, S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11: 65-106. English translation. In: Weiss KM, Ballonoff PA, editors. 1975. Demographic Genetics. Dowden, Hutchinson and Ross, Stroudsburg. p. 224–263.

Waits, L.P. and Paetkau, D.W., 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, 69, pp.1419–1433.

Waits, L.P., Lonsinger, R.C. and Woodruff, S.P., 2016. *Monitoring Species of Concern Using Noninvasive Genetic Sampling and Capture-Recapture Methods*. University of Idaho Moscow United States.

Wang, J., 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research*, 78(3), pp.243-257.

Wang, J., 2009. A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular ecology*, 18(10), pp.2148-2164.

Wang, J., Brekke, P., Huchard, E., Knapp, L.A. and Cowlishaw, G., 2010. Estimation of parameters of inbreeding and genetic drift in populations with overlapping generations. *Evolution: International Journal of Organic Evolution*, 64(6), pp.1704-1718.

Wang, J., Santiago, E. and Caballero, A., 2016. Prediction and estimation of effective population size. *Heredity*, 117(4), pp.193-206.

Waples, R.S., 2005. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply?. *Molecular Ecology*, 14(11), pp.3335-3352.

Waples, R.S., 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, 7(2), p.167.

Waples, R.S. and Do, C.H.I., 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular ecology resources*, 8(4), pp.753-756

Waples, R.S. and Do, C.H.I., 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary applications*, 3(3), pp.244-262.

Waples, R.S. and England, P.R., 2011. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics*, 189(2), pp.633-644.

Waples, R.S., Antao, T. and Luikart, G., 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*, 197(2), pp.769-780.

Waples, R.S., 2015. Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity*, 106(1), pp.1-19.

Watterson, G.A., 1986. The homozygosity test after a change in population size. *Genetics*, 112(4), pp.899-907.

Wayre, P., 1985. A successful reintroduction of European otters. *Oryx*, 19(3), pp.137-139.

Wegmann, D., Leuenberger, C., Neuenschwander, S. and Excoffier, L., 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC bioinformatics*, 11(1), pp.1-7.

Wei, W., Swaisgood, R.R., Dai, Q., Yang, Z., Yuan, S., Owen, M.A., Pilfold, N.W., Yang, X., Gu, X., Zhou, H. and Han, H., Zhang, J., Hong, M. and Zhang, Z., 2018. Giant panda distributional and habitat-use shifts in a changing landscape. *Conservation Letters*, 11(6), p.e12575.

Weir, B.S. and Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), pp.1358-1370.

Weinberg, W., 1908. On the demonstration of heredity in man. SH Boyer (Ed.), (1963) *Papers on human genetics*, Prentice-Hall, Englewood Cliffs, NJ.

White, P.C., McClean, C.J. and Woodroffe, G.L., 2003. Factors affecting the success of an otter (*Lutra lutra*) reinforcement programme, as identified by post-translocation monitoring. *Biological Conservation*, 112(3), pp.363-371.

White, S., O'Neill, D., O'Meara, D.B., Shores, C., Harrington, A.P., O'Reilly, C., Weyman, G. and Sleeman, D.P., 2013. A non-invasive genetic survey of otters (*Lutra lutra*) in an urban environment: A pilot study with citizen scientists. *IUCN Otter Specialist Group Bulletin*, 30(2), pp.103-111.

Wilson, G. A., and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163, pp.1177–1191.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics*, 16, pp.97–259.

Wright, S., 1938. Size of population and breeding structure in relation to evolution. *Science*, 87, pp.430-431.

Yamamichi, M. and Innan, H., 2012. Estimating the migration rate from genetic variation data. *Heredity*, 108(4), pp.362.

Zhan, X., Li, M., Zhang, Z., Goossens, B., Chen, Y., Wang, H., Bruford, M.W. and Wei, F., 2006. Molecular censusing doubles giant panda population estimate in a key nature reserve. *Current biology*, 16(12), pp.R451-R452.

Zouros, E., 1979. Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics*, 92(2), pp.623-646.

Evidence of Eurasian Otter (*Lutra lutra*) population connectivity across the M4 Corridor around Newport Proposed Motorway.

Nia E Thomas, Michael W Bruford and Elizabeth A Chadwick

School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3AX

Summary

The segments of our genome that we use to provide DNA profiles are found in all animal species and have been very useful in understanding wildlife movements, including otters. Most DNA profiling methods have a very strong statistical likelihood of identifying relatives (especially parent and offspring, siblings, etc). We DNA profiled tissue samples collected by the Cardiff University Otter Project in the Conservation Genetics laboratory, following laboratory best practise procedures. These genetic data constitute new evidence on otter movements around the proposed development area and are used to derive genetic relatedness values and estimate kinship between individual otters, to illustrate the connectivity and continuity of the population in the vicinity of the proposed M4 Corridor around Newport (CaN) motorway.

Key findings:

1. Individual otters disperse freely throughout the area, both East-West across the River Usk, and North-South between coastal / inland habitats, as evidenced by the level of genetic mixing observed.

2. Otters tested form a single demographic unit.

3. Construction of a motorway, such as that proposed for the M4 Corridor around Newport (CaN), across this area is expected to reduce habitat connectivity and consequently fragment this population.

Acknowledgements

The samples used were collected for post-mortem examination as part of a long term national research programme (www.otterproject.cf.ac.uk). Core funding has in the past been provided by the Environment Agency and Natural Resources Wales, while current funders are the Esme Fairbairn Foundation and an anonymous charitable donor. Nia Thomas's PhD project is funded by the National Environment Research Council's GW4+ Doctoral Training Partnership and by Somerset Otter Group (Grant Code: NE/L002434/1). Otter collection has been assisted by members of the public, Natural Resources Wales (NRW, previously Countryside Council for Wales), Trunk Roads Agencies, County Ecologists, Wildlife Trusts, and other environmental bodies.

Suggested Citation: *Thomas NE, Bruford MW & Chadwick EA (2017) Evidence of Eurasian Otter (*Lutra lutra*) population connectivity across the M4 corridor around Newport proposed motorway. Cardiff University Otter Project Technical Report for the Gwent Wildlife Trust.*

Table of Contents

List of Tables

List of Figures

Introduction

This report has been prepared to examine otter connectivity in the vicinity of the proposed M4 Corridor around Newport (CaN) motorway, which crosses the River Usk and bisects the Gwent Levels, to assess the risk of fragmenting the Eurasian otter (*Lutra lutra*) population which resides in this area.

Observations of otters in the wild are difficult due to their solitary and elusive nature as well as their nocturnal habits (Kruuk, 1995). As such, many studies across Europe make use of genetic methods to DNA profile tissue samples from dead individuals and spraint (faeces) left behind on rocks and other prominent locations (Coxon *et al.* 1999; O'Neill *et al.* 2009; Lerone *et al.* 2014). The information gained from these genetic data can provide a wealth of knowledge on otter movement, relatedness of individuals and population structure (Hobbs *et al.* 2011; Pagacz 2016).

Many studies have shown the negative effects that roads have on wildlife (e.g. Mader 1984 – mice and carabid beetles in Germany; Baur A & Baur B 1990 - land snails in Sweden; Vos & Chardon 1998 – Moor frogs in the Netherlands; Lodé 2000 – 97 species of road kill in France; Koenig *et al*. 2001 – lizards in Australia; Dyer *et al.* 2002 – caribou in Canada; Rondinini & Doncaster 2002 – hedgehogs in the UK; Bhattacharya et al. 2003 – bumblebees in the USA; McGregor et al. 2008 – mice and chipmunks in Canada; Shepard *et al.* 2008 – turtles in USA). These effects are consistent across a wide range of species as well as geographic locations, and are due to a barrier effect caused both by increased mortality rates via road traffic accidents (RTAs), and by behavioural avoidance of roads. Altered behaviour/ mortality rate leads to more highly fragmented populations, and increased population sub-structuring as gene flow across these barriers is decreased. This fragmentation is evident even in large, mobile species such as grizzly bears (*Ursus arctos*) in Montana, where genetic methods have revealed population sub-structure and low gene flow across Highway 2 (Kendall *et al.* 2009). Riley *et al.* (2006) use similar methods to demonstrate that freeways are barriers to gene flow in Californian carnivores.

DNA profiling can be applied to estimate relatedness between individuals, and this information when coupled with geographic data can be used to help understand population connectivity. Importantly, such information can be used not only to detect fragmentation signals and their likely causes after dispersal barriers have been constructed, but also to pre-emptively determine population connectivity levels during the planning process, ahead of proposed developments, in order to reduce the likelihood of human induced population fragmentation.

The aims of the current study are **to estimate population structure and relatedness among otters in the vicinity of the proposed development, to predict the likely impact of the proposed M4 CaN motorway on population connectivity for this EU protected species.**

Methodology

The presence of otters on UK rivers such as the Usk and surrounding area can be recorded in a variety of ways, for example by surveying for spraint (faeces) deposited close to river banks or, as here, through the identification and collection of individuals killed in road traffic accidents (RTAs) via the Cardiff University Otter Project. Otters found dead in Wales have been collected for post-mortem examination by Cardiff University Otter Project for over 20 years, and have previously been used for genetic studies of this expanding population (Hobbs et al 2011). Full details of the molecular techniques used in this study (as developed at Cardiff University and elsewhere) are provided in Appendix I, and summarised below; best practice laboratory procedures were used throughout.

Tissue samples from the hind leg of geo-referenced otters were collected during post-mortem and stored in ethanol at -20°C. DNA was extracted from the tissue using a QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, #65906) following the 'Isolation of total DNA from animal tissues' protocol.

Otters were DNA profiled using genetic markers designed for the Eurasian otter; and refined for the UK population in previous work (Hobbs *et al.* 2006). We analysed 10 tissue samples from RTA otters at 15 DNA profiling markers (known as microsatellite loci).

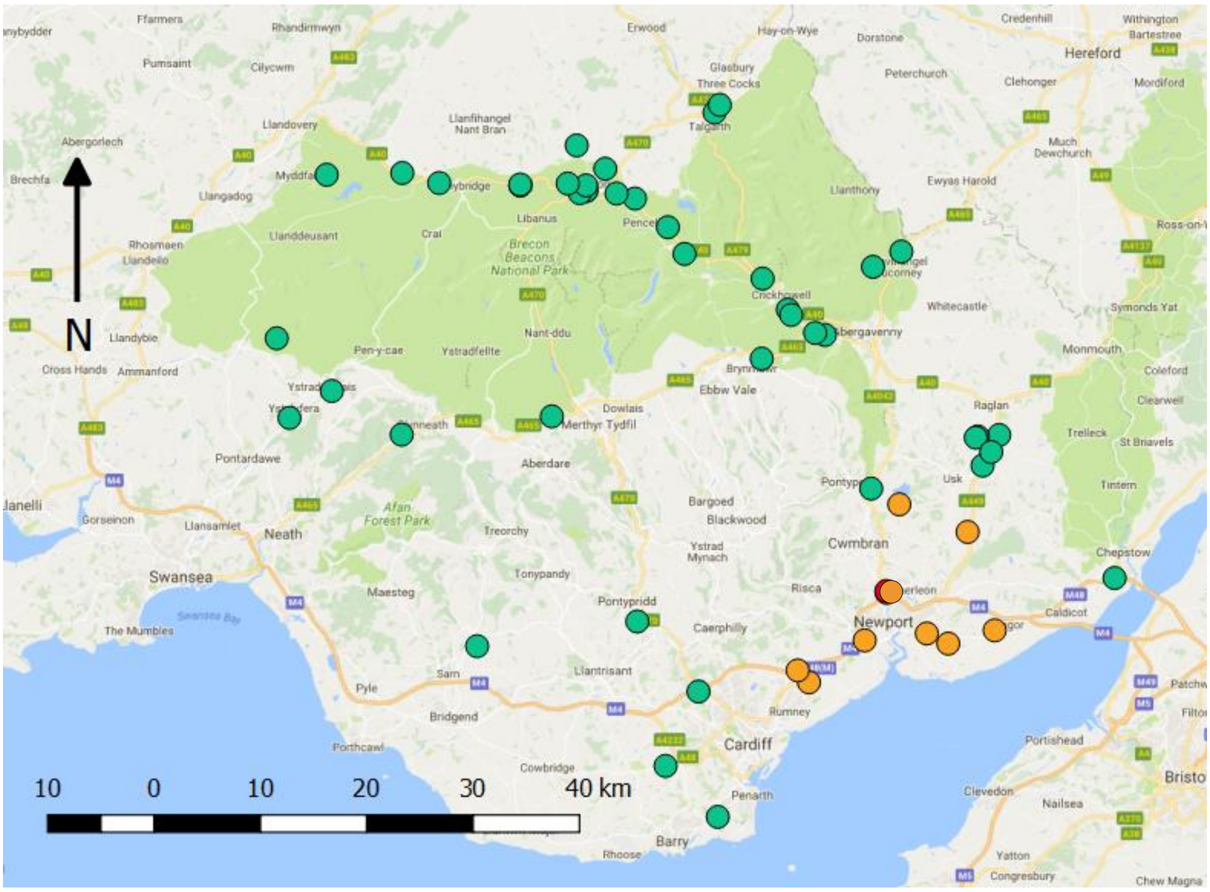


Figure 1: Source locations of the otters used in this study. *Orange circles show the 9 otters genotyped for this work, the red circle is otter 807 which failed analysis and the green circles indicate otters from the SE Wales area used to provide background allele frequencies for analysis.*

Relatedness among samples was quantified via two methods.

1. POSA: The proportion of shared alleles (POSA) between individuals was calculated by pairwise comparison, where the number of shared alleles was divided by the total number of alleles (30, two alleles for each of the 15 loci) in the software package MSA (Dieringer & Schlötterer, 2003).

2. Coefficient of relatedness, R: A more precise measure of relatedness can be calculated by taking into account the background allele frequencies within the wider population (i.e. to reflect the fact that shared *rare* alleles provide more compelling evidence that two individuals are related than shared common alleles). In the current study, we selected 42 previously genotyped samples from the South-East Wales area (See Figure 1) to provide these background allele frequencies. This gives a more conservative measure of relatedness than, for example, using background allele frequencies from across Wales or across the UK. We calculated the coefficient of relatedness (R) value using the ML-Relate package (Kalinowski *et al.* 2006) for each pairwise relationship, using likelihood calculations in Wagner *et al.* (2006). We then performed simulations using 1000 randomisations on the dataset to determine all possible relationships for each pairwise comparison at the 95% confidence level.

Within a theoretical population, relatedness can be quantified using R values which will range between 0 (unrelated individuals) and 1 (identical twins/clones) with many degrees of relation in between (Table 1). In a real population, there is expected to be variation around these theoretical values. ML-Relate applies a series of simulations using the data to statistically determine the likelihood that the relationship inferred is correct; here the relationships inferred from the estimated R values are assigned using a 95% confidence interval.

Table 1: R-values and the relationships they infer within a theoretical population. NB: Individuals from different populations, where background allele frequencies are distinct, can theoretically possess a negative relatedness value.

R-value	Order	Example Relationships
1	Identical	Identical twins, clones
0.5	1st Order	Parent-Offspring pair, Full siblings (may be same, or different litter)
0.25	2nd Order	Grandparent-Grandchild, half siblings, Aunt/Uncle-Niece/Nephew
0.125	3rd Order	Great grandparent-Great grandchild, first cousins, quadruple second cousins and many more complex relationships
0	Unrelated	Unrelated individuals from same population

Visualisation and geographical interpretation of relationships was undertaken using the Geographical Information Systems program QGIS (Quantum GIS Development Team, 2017) using Google Maps as the base map through the OpenLayers plugin. All distance measurements were straight line (i.e. as the crow flies) distances.

Results and Discussion

Nine of the 10 tissue samples (90%) selected for analysis were fully genotyped at the 15 microsatellite loci, with one sample, 807 failing (see Table 2 for sample metadata). The allele size scores making up each genetic profile are given in Appendix II.

Table 2: Otter metadata. Relevant recorded information including sex, age class (determined via post-mortem), date and XY co-ordinates (when and where each otter was found) for the otters analysed.

UWCref	Sex	Age Class	Date Found	X Co-ordinate	Y Co-ordinate
362	Female	Adult	Oct 2000	332600	198500
531	Male	Adult	Sept 2002	331300	190300
809	Male	Sub-Adult	May 2004	341350	186680
956	Male	Sub-Adult	Feb 2006	323940	182070
960	Female	Adult	June 2006	337000	185500
1069	Male	Sub-Adult	Feb 2008	322900	183200
1077	Male	Adult	Sept 2007	329200	185900
2426	Male	Adult	March 2014	335010	186470
2858	Male	Adult	Dec 2015	338935	195859

Proportion of alleles shared (POSA)

Pairwise allele sharing proportions showed that the 9 otters genotyped in the study area share many alleles, with the proportion of shared alleles ranging from 0.4 – 0.77 (i.e. 40-77% shared) (Table 3). Twenty nine out of thirty six pairwise comparisons indicated >50% shared alleles (shaded squares).

Table 3: Proportion of shared alleles. Values shown are pairwise comparisons between individuals to 2 decimal places. *Shaded squares show proportions >0.5*

	362	531	809	956	960	1069	1077	2426	2858
362	-								
531	0.60	-							
809	0.63	0.63	-						
956	0.53	0.67	0.67	-					
960	0.63	0.63	0.60	0.60	-				
1069	0.70	0.57	0.73	0.70	0.60	-			
1077	0.67	0.73	0.67	0.67	0.63	0.77	-		
2426	0.43	0.43	0.47	0.47	0.57	0.40	0.40	-	
2858	0.70	0.70	0.63	0.70	0.73	0.70	0.77	0.40	-

These values indicate a high degree of genetic similarity between the individuals, potentially suggesting the existence of a contiguous otter population with gene flow across the Gwent levels and River Usk SAC.

Coefficient of relatedness, R

Relatedness values (Table 4) confirm that many of the individuals genotyped are likely to be related. Maximum likelihood estimates suggest that there are four pairs of 1st order relatives among our dataset of nine otters, along with 12 pairs of 2nd order relatives while 20 pairings showed no (or a lower order) relationship.

Simulations using 1000 randomisations to establish 95% confidence provides a stringent view of these relationships and shows that for four pairs of individuals (marked R in Table 4) the 95% confidence intervals around relatedness estimates exclude the possibility of zero relatedness, i.e. we can be highly confident that these individuals are related. For six pairs (marked U in Table 4) 95% confidence intervals around relatedness estimates exclude the possibility of being either first or second order relatives, i.e. we can be highly confident that these individuals are not closely related. For 26 pairs (marked ? in Table 4) the confidence intervals span both related and unrelated states and thus the relationship is technically unresolved using this stringent method. For example, although maximum likelihood might infer a pair are 2nd order relatives we cannot with 95% confidence exclude the possibility that they are unrelated, and conversely where ML infers that they are unrelated, we cannot with 95% confidence exclude a relationship (see Appendix III for LnL scores from ML-Relate).

Table 4: Relatedness Estimates (R) based on ML-Relate (Kalinowski *et al.* 2006). In the lower half of the table pairwise R values for the samples analysed are displayed with maximum likelihood inferred relationships denoted as follows: unrelated individuals with no asterisk, 2nd order relatives annotated with one asterisk (*) and 1st order relatives annotated with two asterisks (**). First and second order relationships are shown in bold. The top half of the table shows the possible relationships at 95% confidence using simulations on the dataset with 1000 randomisations; U = unrelated (not 1st or 2nd order relatives), R = definitely related (1st or 2nd order), ? = relationship unresolved.

	362	**531**	**809**	**956**	**960**	**1069**	**1077**	**2426**	**2858**
362	-	?	?	U	?	?	?	U	?
531	0	-	?	?	?	?	?	U	?
809	0	0.08	-	R	?	?	?	?	?
956	0	0.28	**0.47****	-	R	?	?	?	R
960	0.14	**0.14***	0.13	**0.5****	-	?	?	R	?
1069	**0.17***	0	**0.5****	**0.26***	0	-	?	U	?
1077	**0.16***	**0.32***	0.11	0.12	0.01	**0.31***	-	U	?
2426	0	0	0	0.1	**0.23***	0	0	-	U
2858	0.1	**0.2***	0	**0.5****	0.34	**0.37***	**0.3***	0.02	-

Using the metadata collected during post-mortem we analysed the demographic features of all pairs identified as 1st order relatives and/or those assigned as closely related with 95% confidence. Each relationship is also given geographic context in Figures 2 and 3.

Otters 956 and 960 were a sub-adult male and adult female respectively, both killed in 2006, that could be related as mother/offspring or aunt/nephew. Simulations give us 95% confidence that the pair are either 1st or 2nd order relatives. **They were found ~21.7km apart (as the crow flies), both to the South of the proposed route on either side of the Usk River.**

Otter 809 and 956 were both sub-adult males, killed in 2004 and 2006 respectively and thus most likely to be full or half siblings but from litters two years apart. As otter home ranges can remain constant over a number of years, with male and female ranges often overlapping, the same female and male otters may breed together over consecutive years making full siblings in different litters theoretically possible (Kruuk 1995). Again, simulations give us 95% confidence that the pair are either 1st or 2nd order relatives. **They were found ~29km apart (as the crow flies) either side of the Usk River, with 956 South of the proposed CaN motorway and 809 on the route of the proposed CaN motorway.**

Otters 809 and 1069, both sub-adult males, killed in 2004 and 2008 respectively are also likely to share at least one parent but come from different litters. However, it should be noted that although maximum likelihood estimates suggest that this pair are first order relatives, simulations suggest very broad confidence intervals and we are unable to exclude the possibility of unrelatedness. **They were found ~30.2km apart (as the crow flies), either side of the Usk River, with 1069 on the M4 to the West and 809 on the proposed CaN motorway route in the East.**

Otters 956 and 2858 were a sub-adult male killed in 2006 and an adult male killed in 2015 respectively. It is not possible to accurately age otters without undertaking cementum analysis of the teeth (e.g. Sherrard Smith and Chadwick 2010), but otters have been found in the wild up to 16 years of age (Ansorge et al, 1997). The post-mortem report for otter 2858 indicated that he was an old male (inferred from baculum length, testis size and tooth wear), so birth is likely to have been no later than 2012, and more likely around 2010. Otter 956, a sub-adult, must have been born in 2005, implying a gap of 5-7 years between the litters producing these animals. Recorded post mortem information for otter 956 suggest an immature male, therefore ruling out a parent offspring relationship. Thus these individuals are most likely to be second order relatives (Uncle/Nephew or Grandfather/grandson). Simulations provide 95% confidence that this pair are closely related. **They were found ~32.9km apart (as the crow flies) on either side of both the River Usk and the proposed route.**

Otters 960 and 2426 were an adult female and an adult male killed in 2006 and 2014 respectively and are most likely to be 2nd order relatives, probably grandmother-grandson given the time between deaths. Simulations provide 95% confidence that this pair are closely related. **They were found ~3.5km apart (as the crow flies) both to the East of the River Usk but on either side of the proposed route.**

Many of these close relationships span both the river and the proposed route (figure 2), indicating first-order relative (i.e single or bi-generational) dispersal. This is also true using the stringent 95% confidence relationships (figure 3). There is no apparent genetic sub-structuring of the population within this area, allowing us to infer that otters are dispersing, reproducing and contributing to gene-flow between the west and eastern part of the region (across the Usk) and north and south of the proposed CaN motorway.

Road traffic accidents (RTAs) are believed to be the highest cause of otter mortality in the UK, with over 50% of mortalities being observed to have occurred on trunk or A roads despite these making up only 13% of the network (Philcox *et al.* 1999). The current study provides evidence of a homogenous population, mixing freely throughout the Gwent levels area. Construction of the M4 CaN motorway is therefore highly likely to fragment the resident otter population and its habitat, both directly by acting as a physical barrier and indirectly via road avoidance behaviour. Loss of population connectivity and related gene flow might subsequently cause a decrease in genetic diversity of the fragmented populations.

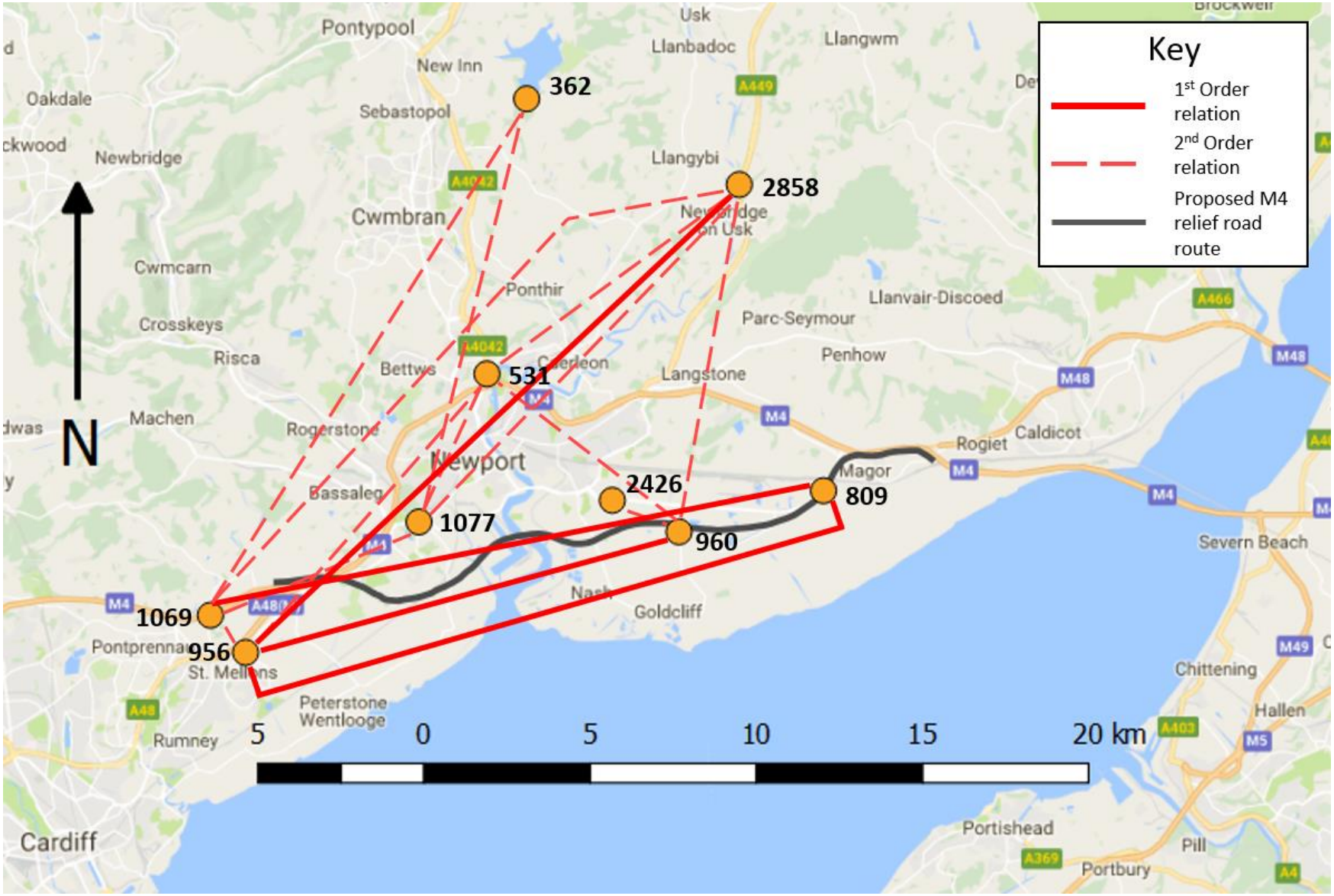


Figure 2: Location of each otter profiled (identifiable by orange points and UWCref codes) and their pairwise relatedness as determined by ML-Relate (Kalinowski *et al.* 2006) using maximum likelihood calculations. *Solid red lines show 1st order relatives and dashed pink lines show 2nd order relatives. The approximate proposed route of the M4 relief road is given by the grey line.*

Appendix I

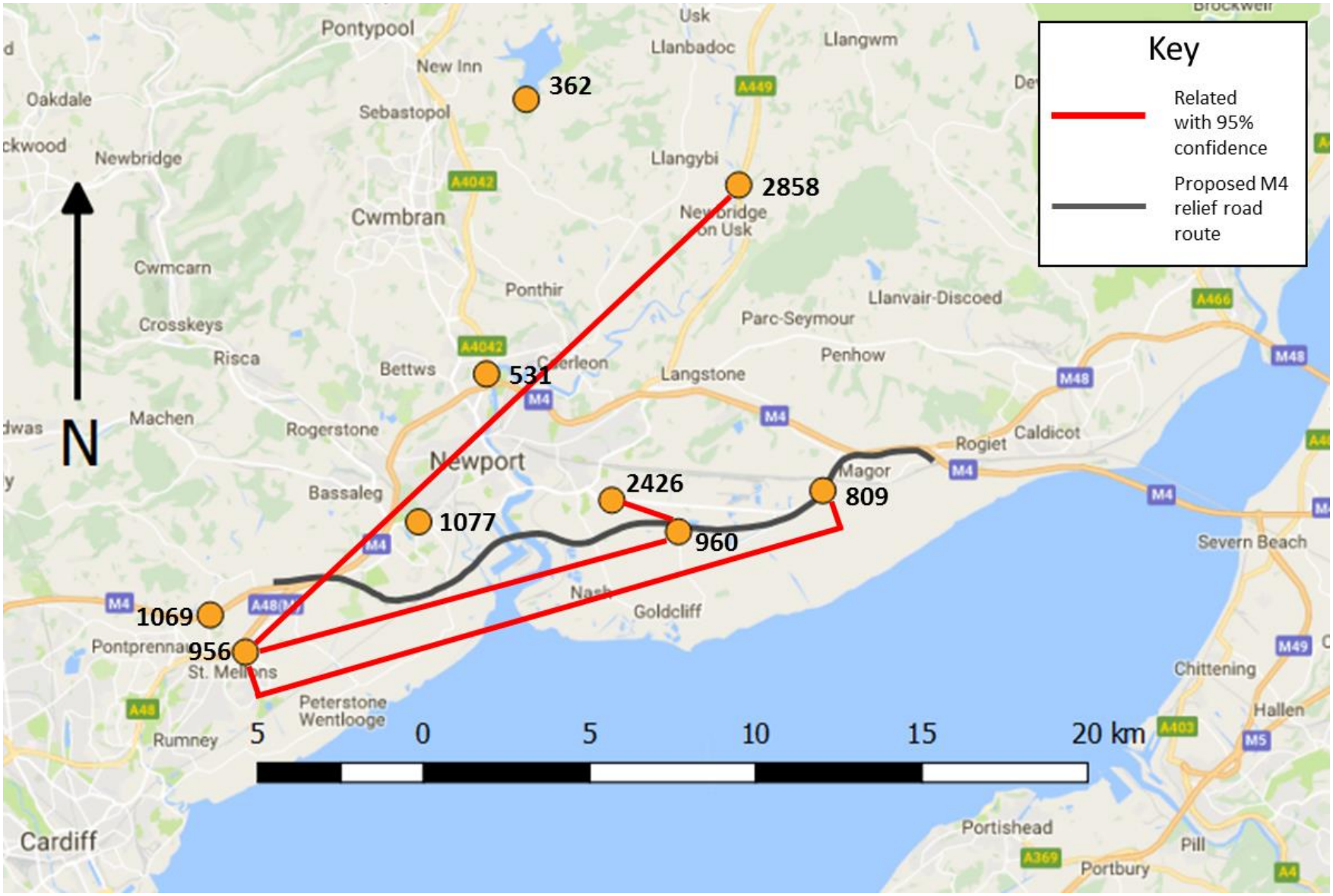


Figure 3: Location of each otter profiled (identifiable by orange points and UWCref codes) with pairwise relations at 95% confidence as determined by ML-Relate (Kalinowski et al. 2006). *Solid red lines show pairwise relations at the 0.25 or 0.5 level (1st or 2nd order relatives) with 95% confidence. The approximate proposed route of the M4 relief road is given by the grey line.*

Conclusions

1. Individual otters regularly disperse and mate across the Gwent Levels and River Usk (i.e. across the proposed M4 CaN motorway route).

2. The populations of otters on either side of the river (and in the river itself) and above and below the proposed route should therefore be treated as a single demographic unit.

3. Construction of a road, such as proposed for the M4 CaN motorway, across this area will impede dispersal, fragment both otter habitat and this population, reducing connectivity and thus gene flow.

References

Ansorge, H, Schipke, R & Zinke O (1997) Population structure of the otter, *Lutra lutra*: parameters and model for a central European region. *Zeitschrift Fur Saugetierkunde*. 62, 143-151.

Baur A & Baur B (1990) Are roads barriers to dispersal in the land snail *Arianta arbustorum*? *Canadian Journal of Zoology*. 68, 613–617.

Bhattacharya M, Primack RB & Gerwein J (2003) Are roads and railroads barriers to bumblebee movement in a temperate suburban conservation area? *Biological Conservation*. 109, 37–45.

Coxon K, Chanin P, Dallas J & Sykes T (1999) *The use of DNA fingerprinting to study population dynamics of otters (Lutra lutra) in Southern Britain: a feasibility study*. R&D Technical Report W202, Environment Agency, Bristol, United Kingdom

Dieringer D & Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes.* 3 (1), 167-169

Dyer SJ, O'Neill JP, Wasel SM & Boutin S (2002) Quantifying barrier effects of roads and seismic lines on movements of female woodland caribou in northeastern Alberta. *Canadian Journal Zoology*. 80, 839–845.

Hauer S, Ansorge H & Zinke O (2002) Reproductive performance of otters *Lutra lutra* (Linnaeus, 1758) in Eastern Germany: low reproduction in a long-term strategy. *Biological Journal of the Linnean Society*. 77, 329-340.

Kalinowski ST, AP Wagner & ML Taper (2006) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*. 6, 576-579.

Kendall KC, Stetz JB, Boulanger J, Macleod AC, Paetkau D & White GC (2009) Demography and genetic structure of a recovering brown bear population. *Journal of Wildlife Management*. 73 (1), 3-17.

Koenig J, Shine R & Shea G (2001) The ecology of an Australian reptile icon: how do blue-tongued lizards (*Tiliqua scincoides*) survive in suburbia? *Wildlife Research*. 28, 215–227.

Kruuk H. (1995) *Wild Otters – Predation and Populations.* Oxford University Press, Oxford UK.

Lerone L, Mengoni C, Carpaneto GM, Randi E & Loy A (2014) Procedures to genotype problematic non-invasive otter (*Lutra lutra*) samples. *Acta Theriologica.* 59 (4), 511-520.

Lodé T (2000). Effect of a motorway on mortality and isolation of wildlife populations. *Ambio*. 29, 163–166.

Mader HJ (1984). Animal habitat isolation by roads and agricultural fields. *Biological Conservation*. 29, 81–96.

McGregor RL, Bender DJ & Fahrig L (2008) Do small mammals avoid roads because of the traffic? *Journal of Applied Ecology*. 45, 117–123.

O'Néill L, Veldhuizen T, de Jongh A, & Rochford J (2009) Ranging behaviour and socio-biology of Eurasian otters (*Lutra lutra*) on lowland mesotrophic river systems. *European Journal of Wildlife Research*. 55, 363–370.

Pagacz S (2016) The effect of a major drainage divide on the gene flow of a semiaquatic carnivore, the Eurasian otter. *Journal of Mammalogy*. 97 (4), 1164-1176.

Philcox CK, Agrogan AL & Macdonald DW (1999) Patterns of otter *Lutra lutra* road mortality in Britain. *Journal of Applied Ecology*. 36, 748–761.

Quantum GIS Development Team (2017*). Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Riley SPD, Pollinger JP, Sauvajot RM, York EC, Bromley C, Fuller TK & Wayne RK (2006) A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology*. 15, 1733–1741.

Rondinini, C & Doncaster CP (2002) Roads as barriers to movement for hedgehogs. *Functional Ecology*. 16, 504–509.

Shepard DB, Kuhns AR, Dreslik MJ & Phillips CA (2008) Roads as barriers to animal movement in fragmented landscapes. *Animal Conservation*. 11, 288–296.

Vos CC & Chardon JP (1998) Effects of habitat fragmentation and road density on the distribution pattern of the moor frog Rana arvalis. *Journal of Applied Ecology*. 35, 44–56.

Wagner AP, Creel S & Kalinowski ST (2006) Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity*. 97, 336–345.

Appendices

Appendix I - Full Molecular Methods and Analysis

Using primers designed for the Eurasian otter; and optimally rarefied for the UK population in previous work (Hobbs *et al.* 2006); we genotyped 9 tissue samples from RTA otters (10 samples were initially selected, but one failed analysis) at 15 microsattelite loci; lut435, lut453, lut717, lut604, lut733, lut615, lut902, lut782, lut701, lut833, lut818, lut715, lut832 (Dallas and Piertney, 1998), 04OT05, 04OT22 (Huang et al., 2005). The 15 loci were split into three multiplex mixes (see Table 1) which all contained 5 primers. The forward primer for each locus was fluorescently labelled (Fam, Hex or Ned dye) to allow fragment size analysis by capillary electrophoresis downstream. polymerase chain reaction's (PCR) were carried out using a QIAGEN Multiplex PCR kit following the 'amplification of microsatellite loci using multiplex PCR' protocol (QIAGEN, #206143), but with a smaller final volume of 10µl. Amplification of PCR products was performed using Veriti™ 96-Well Thermal Cycler (Applied Biosystems) in 10 µl reactions, using 2 µl of template DNA, 5 µl of 1x QIAGEN Multiplex PCR Master Mix (containing HotStarTaq® DNA polymerase, Multiplex PCR buffer (contains 3 mM MgCl2) and dNTP Mix), 1 µl of 10x Primer Mix (0.2 µM of each primer) and 2 µl of sterile water. All multiplexes underwent identical PCR reactions with the following thermal profile: initial denaturation step of 95°C for 15mins, followed by 29 cycles of 94°C for 30s, 58°C for 90s and 72°C for 1 min and a final extension of 60°C for 30 mins. PCR products were visualised on a 1% agarose-gel stained with Invitrogen™ SYBR™ Safe DNA Gel Stain and visualised using a UVP 3UV™ Transilluminator. PCR products were sent for capillary electrophoresis at DNAseq (in Dundee) using an Applied Biosystems 3730XL DNA analyser (Applied Biosystems). Results were analysed and allele sizes scored using GeneMapper™ V4.0 (Applied Biosystems).

Relatedness among samples was quantified via two methods.

1. POSA: The proportion of shared alleles (POSA) between individuals was calculated by pairwise comparison, where the number of shared alleles was divided by the total number of alleles (30, two alleles for each of the 15 loci) in the software package MSA (Dieringer & Schlötterer, 2003).

2. Coefficient of relatedness, R: A more precise measure of relatedness can be calculated by taking into account the background allele frequencies within the wider population (i.e. to reflect the fact that shared rare alleles provide more compelling evidence that two individuals are related than shared common alleles). In the current study, we selected 42 previously genotyped samples from the South-East Wales area (See Figure 1) to provide these background allele frequencies. This gives a more conservative measure of relatedness than, for example, using background allele frequencies from across Wales or across the UK. We calculated the co-efficient of relatedness (R) value using the ML-Relate package (Kalinowski et al. 2006) for each pairwise relationship, using likelihood calculations in Wagner et al. (2006) and all relationships were inferred at the 95% confidence level.

In the undertaking of this work our lab followed laboratory best practise procedures. Although it should be made clear that as a research lab we are not accredited to ISO/IEC 17025:2005 standards (required for human forensic evidence in court), this type of molecular analysis on wildlife is routinely carried out in our lab and published in peer reviewed journals.

Appendix II – Full genotype profiles for samples

Column headings refer to the 15 named microsatellite loci, data show the two alleles (fragment sizes) for each individual at each locus.

UWCref	Lut435		Lut453		04OT05		Lut717		04OT22		Lut604		Lut733		Lut615		Lut902		Lut782		Lut818		Lut701		Lut833		Lut715		Lut832	
362	123	135	127	127	171	179	191	191	156	156	137	137	174	174	219	225	149	170	188	188	158	184	201	201	158	158	204	204	189	189
531	123	135	123	127	179	187	191	191	156	156	137	137	166	174	219	219	149	166	161	188	180	184	201	206	166	166	199	212	189	189
809	123	123	127	127	179	187	191	191	156	160	137	137	170	174	219	225	166	170	188	161	158	184	201	210	154	170	204	212	189	189
956	123	135	123	127	187	187	191	191	156	160	126	137	174	174	219	219	166	166	161	188	176	184	201	210	154	166	204	208	189	189
960	123	135	123	123	179	187	191	191	156	156	126	130	170	174	219	223	166	170	188	188	158	184	201	201	166	166	204	204	189	189
1069	123	135	127	127	179	179	191	191	156	160	137	137	174	174	219	225	166	170	188	188	176	184	201	210	154	166	204	204	189	204
1077	123	135	127	127	179	179	191	191	156	164	137	137	174	174	219	219	166	170	161	188	184	184	201	206	166	166	204	204	189	189
2426	123	135	123	127	175	183	191	191	156	156	126	126	170	170	221	223	166	170	161	165	158	158	201	201	154	158	204	200	189	193
2858	123	135	127	127	179	187	191	191	156	160	126	137	174	178	219	219	149	166	188	188	184	184	201	201	166	166	204	204	189	189

Appendix III – Inferred Relationships and LnL Scores from ML-Relate

Table showing the full results from ML-Relate: ML Relationship shows the relationship with the highest likelihood; R gives the co-efficient of relatedness; LnL(R) indicates the natural logarithm of relatedness (R); the four columns labelled unrelated, half sibs, full sibs and parent-offspring show the differences on a log scale from the LnL(R), if a relationship is excluded by the genetic data then 9999 is reported; the final column indicates all possible relationships for a particular pair at 95% confidence.

UWCref1	UWCref2	ML Relationship	R	LnL(R)	Unrelated	Half Sibs	Full Sibs	Parent Offspring	Relationships at 95% CI
531	362	Unrelated	0	-43.85	-	1.71	4.81	9999	U, HS
809	362	Unrelated	0	-37.52	-	0.93	2.9	9999	U, HS
809	531	Unrelated	0.08	-46.89	-	0.76	2.97	9999	U, HS
956	362	Unrelated	0	-43.88	-	2.62	6.7	9999	U
956	531	Half Sibs	0.28	-51.76	1.48	-	1.17	9999	U, HS, FS
956	809	Parent-Offspring	0.47	-44.59	2.33	0.41	1.64	-	HS, FS, PO
960	362	Unrelated	0.14	-43.16	-	1.66	3.45	9999	U, HS
960	531	Half Sibs	0.14	-52.27	0.26	-	1.9	9999	U, HS, FS
960	809	Unrelated	0.13	-46.2	-	1.72	3.71	9999	U, HS
960	956	Parent-Offspring	0.5	-48.97	3.59	0.8	3.47	-	HS, PO
1069	362	Half Sibs	0.17	-37.79	0.01	-	0.8	9999	U, HS, FS
1069	531	Unrelated	0	-47.17	-	1.86	5.64	9999	U, HS
1069	809	Parent-Offspring	0.5	-39.1	1.75	0.44	0.57	-	U, HS, FS, PO
1069	956	Half Sibs	0.26	-46.32	0.88	-	0.05	9999	U, HS, FS
1069	960	Unrelated	0	-46.49	-	1.68	4.47	9999	U, HS
1077	362	Half Sibs	0.16	-37.01	0.12	-	1.5	9999	U, HS, FS
1077	531	Half Sibs	0.32	-45.45	1.05	-	0.13	9999	U, HS, FS
1077	809	Unrelated	0.11	-40.17	-	0.09	1.85	9999	U, HS, FS
1077	956	Unrelated	0.12	-46.52	-	0.03	1.82	9999	U, HS, FS
1077	960	Unrelated	0.01	-45.81	-	0.67	2.65	9999	U, HS
1077	1069	Half Sibs	0.31	-39.74	0.71	-	0.28	0.35	U, HS, FS, PO
2426	362	Unrelated	0	-57.38	-	2.48	6.94	9999	U
2426	531	Unrelated	0	-66.75	-	3.59	8.54	9999	U
2426	809	Unrelated	0	-60.42	-	1.77	6.36	9999	U, HS
2426	956	Unrelated	0.1	-66.78	-	0.21	4.28	9999	U, HS
2426	960	Half Sibs	0.23	-63.67	2.4	-	2.3	9999	HS
2426	1069	Unrelated	0	-60.7	-	4.95	10.44	9999	U
2426	1077	Unrelated	0	-60.03	-	4.11	9.34	9999	U
2858	362	Unrelated	0.1	-36.23	-	0.09	1.77	9999	U, HS, FS
2858	531	Half Sibs	0.2	-45.46	0.13	-	0.69	9999	U, HS, FS
2858	809	Unrelated	0	-39.27	-	0.56	2.88	9999	U, HS
2858	956	Parent-Offspring	0.5	-42.2	3.42	0.98	1.49	-	HS, FS, PO
2858	960	Half Sibs	0.34	-43.26	1.65	-	0.3	9999	U, HS, FS
2858	1069	Half Sibs	0.37	-39.07	0.48	-	1.45	0	U, HS, FS, PO
2858	1077	Half Sibs	0.3	-37.89	0.99	-	0.14	0.06	U, HS, FS, PO
2858	2426	Unrelated	0.02	-59.13	-	3	8.41	9999	U