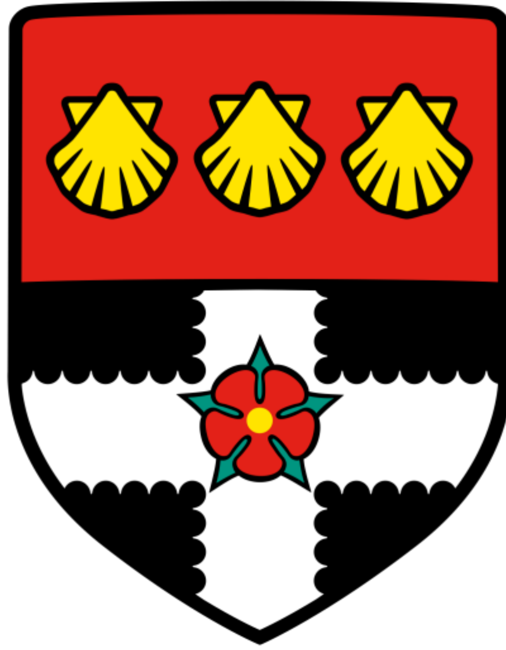


UNIVERSITY OF READING

MODELLING THE FUNCTIONAL CONNECTIVITY AND POPULATION DYNAMICS OF BUTTERFLIES USING POPULATION AND GENETIC DATA

A thesis submitted for the degree of
Doctor of Philosophy

Matthew Peter Greenwell

School of Biological Sciences

September 2020

Preface

This thesis includes one published manuscript (Chapter 2):
Greenwell, M., Brereton, T., Day, J., Roy, D. & Oliver, T. (2019). Predicting resilience of ecosystem functioning from co-varying species' responses to environmental change. *Ecol. Evol.*, 9, 11775–11790. https://doi.org/10.1002/ece3.5679

In this thesis Chapter 5 has also been submitted for publication as follows:
Greenwell, M., Botham, M., Bruford, M.W., Day, J.C., Evans, L.C., Gibbs, M., Roy, D., Watts, K. & Oliver, T. (2020). In review. The influence of chalk grasslands on butterfly phenology and ecology

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Matthew P. Greenwell

For published and submitted chapters, author contributions are detailed below.

Chapter 2 - Predicting resilience of ecosystem functioning from co-varying species' responses to environmental change:
THO conceived the study with input from MPG; DBR and TB collated and processed the data. MPG performed the analysis. All authors contributed to the writing of the manuscript.

Chapter 5 - The influence of chalk grasslands on the phenology and ecology of *Maniola jurtina* in the UK:
THO conceived the study with input from MPG, DBR, MB and MG. DBR and MB collated and processed the population monitoring data. MG conducted drought experiments on *M. jurtina* larvae. DBR, LCE, MB, MPG, KW and THO collected samples for genetic analysis. DNA extraction were conducted by MPG and MG. All genetic analysis was conducted by MPG with the help of JCD. MPG performed the analysis with the help of LCE. All authors contributed to the writing of the manuscript.

Acknowledgements

PhDs have a reputation of being overly stressful and all consuming. I can thankfully say that this hasn't been the case for me, largely due to the support that I've received over the past four years from everyone around me.

Firstly, I need to thank NERC and SCENARIO for funding my studentship, as well as their constant interest and encouragement. I'd also like to thank all of my collaborators across Europe who have been involved with the project: Anne Eskildsen, Chris van Swaay, Constanti Stefanescu, David Roy, Dirk Maes, Elisabeth Kühn, Irma Wynhoff, Josef Settele, Lars Pettersson, Marc Botham, Martin Musche, Matthias Plattner, Melanie Gibbs, Tiit Teder, Toke Thomas Høye, Xavier Mestdagh. Additionally, thanks to Nia Thomas for all her help with the landscape genetics.

You hear horror stories of students with multiple supervisors being pulled in different directions. This could not have been further from the truth. All four of my supervisors were immensely supportive and always had my best interests in mind. Tom, thank your support, guidance and input, all of which have been invaluable. You may well be the busiest man I have ever met, but you have constantly inspired me and shown what can be achieved if you are passionate about something. Kevin and Mike, thank you for your timely inputs. A conversation with either of you was never a wasted one and always left me energised, even if they did make my brain hurt from time to time. John, thank you for making this happen. Without your support, patience and sense of humour in the face of failure I would never have managed to keep smiling throughout the lab work. Microsatellites. Never again, until the next time.

I've been incredibly fortunate to be surrounded by friends from not one but two institutions. To all the staff and students at CEH, thank you for welcoming me into the group. I'd particularly like to thank Denny, Elmer and Holly for putting up with me turning up to the office unannounced at strange hours of the day. Thanks also to Anna and Lindsey for all their help in the lab, particularly every time I thought I'd broken the ABI machine. To all the staff and students at Reading, thank you for being there for me. Luke and Vicky, thanks for guiding the way. Alex and Chanida, thanks for being alongside me from the start. Andre, Joe, Lewis, Lisbeth and Tom, thanks for pushing me along.

To my parents, thank you for everything. Your support has been unwavering and never unappreciated. To my brother, thanks for sharing a sense of humour with me, it's good to have someone to laugh with. To Sophie, thanks for putting up with me for so long. Finally, to Simon Gathercole, thanks for setting me on this path. Who'd have known what one little butterfly book would lead to.

Table of Contents

Preface i
Declaration iii
Acknowledgements v
Table of Contents vi
Abstract 1
1.1 Biodiversity monitoring 3
1.2 Ecosystem services 5
1.3 How landscape configuration affects biodiversity 6
1.4 Landscape ecology 7
1.5 Genetic diversity 9
1.6 Butterflies as indicators of biodiversity 14
1.7 The meadow brown butterfly, *Maniola jurtina* 18
1.8 Overall project aims 26
Chapter 2. Predicting resilience of ecosystem functioning from co-varying species' responses to environmental change 29
2.1 Abstract 29
2.2 Introduction 30
2.3 Materials and methods 33
2.4 Results 44
2.5 Discussion 45
Chapter 3. Landscape genetics of *Maniola jurtina* in Southern England 49
3.1 Abstract 49
3.2 Introduction 50
3.3 Methods 53
3.4 Results 60
3.5 Discussion 79
Chapter 4. Quantifying the genetic diversity of *Maniola jurtina* across Europe 87
4.1 Abstract 87
4.2 Introduction 88
4.3 Methods 91
4.4 Results 97
4.5 Discussion 110
Chapter 5. The influence of chalk grasslands on the phenology and ecology of *Maniola jurtina* in the UK 115
5.1 Abstract 115
5.2 Introduction 116
5.3 Materials and methods 118
5.4 Results 123

5.5 Discussion 128
Chapter 6. General discussion 131
6.1 Thesis overview and implications 131
6.2 Limitations 133
6.3 Implications and applications 136
6.4 Conclusions 139
Appendices 141
Appendix A Chapter 3 – Sample capture and preparation 141
Appendix B Chapter 3 – Novel primer development and evaluation 142
Appendix C Chapter 3 – Pre-existing primer testing 150
Appendix D Chapter 4 – Effects of UK site removal on genetic analysis 156
Appendix E Chapter 4 – Bottleneck analysis results 164
Appendix F Chapter 5 – Site characteristics and correlations 167
Appendix G Chapter 5 – Phenology 169
Appendix H Chapter 5 – Population genetics 175
References 181

Abstract

Biodiversity monitoring underpins decision making in conservation. From small scale management, through to international policy decisions, biodiversity data are used to advise and influence those in positions of power. We live in an age of increasing data accumulation. In particular, an increase in citizen science schemes has led to a wealth of biodiversity information. Despite this increase in data collection, there are still large gaps in our knowledge, with questions that remain unanswered and subjects that continue to be neglected. This thesis focuses on biodiversity monitoring, investigating how data from current long-term monitoring schemes can be applied to novel questions, how new forms of biodiversity monitoring are desperately required and how different forms of biodiversity monitoring data can be combined to explain features of a species biology.

Firstly, long-term population monitoring data are used to overcome an impasse in functional ecology. The ability to predict ecosystem service stability has so far been an out of reach goal. However, analysing correlations between species population dynamics offers an achievable method of determining whether specific functions and services are at risk of declining due to changes in species abundances across a community. Secondly, the genetic diversity of the meadow brown butterfly, *Maniola jurtina*, is investigated at both the spatial and temporal scale. The genetic diversity of the species is found to be stable across the study area and over time. This represents an important contribution to the field of genetic diversity monitoring. Despite being acknowledged as an increasingly important measure of biodiversity, genetic diversity monitoring schemes are extremely rare outside of socioeconomic species. This study represents one of the first examples of the monitoring of a wild species that has no direct economic value. Next, the genetic diversity of *M. jurtina* is investigated at the continental scale, building upon the work in the previous chapter. Across the continent there appears to be distinct population structuring, with individuals in the UK belonging to a different genetic cluster to those on the mainland. Finally, the ability to combine monitoring data with genetic and experimental data is demonstrated with an investigation into the phenology of *M. jurtina.* Analysis of long-term monitoring data determined that *M. jurtina* display a protracted flight period on chalk sites. Genetic data are used to determine whether any genetic structuring of populations is associated with these differences, whilst experimental data are used to determine the effect of drought on phenology.

Overall, this thesis brings together three separate areas to demonstrate the wide range of studies that monitoring data can be applied to. Furthermore, the importance of genetic diversity monitoring is highlighted, along with a demonstration into the relative ease at which it can be accomplished. In the final chapter, the limitations of the work are discussed along with the wide range of future applications.

Chapter 1. Introduction

1.1 Biodiversity monitoring

Making informed, evidence-based decisions is an important factor in the creation and implementation of policies. However, the ability to make such decisions relies upon the availability of data. Policies relating to ecology and conservation biology are no exception to this as governments globally are coming under increasing pressure to address the effects of anthropogenic changes to biodiversity and climate (Sala *et al.* 2000; Butchart *et al.* 2010; McGill *et al.* 2015; Newbold *et al.* 2015; Geijzendorffer *et al.* 2016). This pressure has largely come about due to international conventions (Schmeller *et al.* 2009), such as the Convention on Biological Diversity (CBD) (Convention on Biological Diversity 2020a), rising public pressure in the wake of increasing mainstream media coverage (Legagneux *et al.* 2018) and landmark reports from the Intergovernmental Panel on Climate Change (IPCC 2018) and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES 2019).

The Convention on Biological Diversity is currently the largest political attempt at reducing and halting biodiversity loss. Whilst ambitious in scope the CBD is limited, as member states are not legally bound to implement the agreed policies and targets, meaning that action rests upon the will of participating governments (Laikre 2010). In an attempt to reduce the pressures on, and to improve the status of biodiversity the CBD developed the Aichi Biodiversity Targets to be achieved by 2020 (Convention on Biological Diversity 2011a; Proença *et al.* 2017). For these targets to be effective, regular updates on changes to biodiversity both at local and global scales are required (Proença *et al.* 2017). Therefore the need for reliable measures of biological change, allowing governments and policy makers to maximise the efficiency of conservation actions are vital (Isaac *et al.* 2014). The data required for such robust measures include an understanding of the distributions, habitat preferences, and movements over both time and space of species; knowledge of which is reliant upon human observations of biodiversity (Hochachka *et al.* 2012). Few of the 20 Aichi targets have been met even partially, including where monitoring data are sufficient to make estimations (Convention on Biological Diversity 2020b). For some aspects of biodiversity, such as genetic diversity, there is not even enough data to determine whether we are on course to meet these targets (Tittensor *et al.* 2014; Convention on Biological Diversity 2020b).

1.1.1 The value of biodiversity monitoring

Biological recording i.e. the reporting of the presence of a species at a certain place and time (Isaac & Pocock 2015), is nothing new. Locust outbreaks have been recorded in China for over 3500 years, and the first flowering dates of cherry blossom in Kyoto Japan have been recorded for over 1200 years (Miller-Rushing *et al.* 2012). Whilst biological recording has occurred throughout history, well-structured biological monitoring schemes involving members of the public are a more recent occurrence; the longest running being the Christmas Bird Count in the USA which dates back to 1900 (Silvertown 2009). The use of non-expert volunteers, or citizen scientists, has been crucial to the rise of successful monitoring schemes because, whilst there is a trade off in the accuracy of

recordings in some instances, this is outweighed by the geographic extent of these data, the level of spatial precision and the efficiency at which data can be collected (Schmeller *et al.* 2009; Powney & Isaac 2015; Sutherland *et al.* 2015b). One example of such a scheme is the UK Butterfly Monitoring Scheme (UKBMS) which has been ongoing since 1976 (Pollard 1977). The standardised methods used in the UKBMS, described in Pollard & Yates (1993), combined with the long-term nature of the dataset allow reliable estimates of trends in species abundances to be calculated (Van Swaay *et al.* 2008).

Long-term monitoring schemes are not without their problems. These include the accuracy of recordings, biases in the distribution of sites and habitat types both at national and international scales, and overrepresentation of certain species (Martin *et al.* 2012; Isaac & Pocock 2015). Even with these drawbacks, the vast amounts of data now available at relatively low costs, coupled with advances in recording technology (August *et al.* 2015) and more sophisticated modelling techniques, mean that the range of policy applications that long-term monitoring data can be used for is increasing (Isaac & Pocock 2015). Additionally, monitoring data can be used to evaluate the effectiveness of current policies already in place (Schmeller *et al.* 2009; Geijzendorffer & Roche 2013).

1.1.2 Addressing the gaps in biodiversity monitoring

Whilst long-term biological monitoring schemes provide high quality data, there are certain well known problems including the restricted distribution of monitoring schemes and the limited diversity of taxa included in current schemes that need to be addressed (Isaac *et al.* 2014; Geijzendorffer *et al.* 2016). There are, however, a number of gaps and opportunities regarding monitoring schemes that do not fall into these well explored categories. One such gap is a lack of monitoring data for aspects of biodiversity that are of great interest to conservation biology, such as genetic diversity. Another is the lack of application of monitoring data in solving current problems, such as predicting ecosystem service resilience. Using already available monitoring data to address this issue has the potential to save both time and resources in an area that has previously been data and computationally intensive. Finally, the combination of monitoring data with other datasets e.g. experimental or genetic could allow greater levels of understanding of ecological processes.

Attempting to address these gaps and to take advantage of these opportunities will be the focus of this thesis. Firstly, with an investigation into using monitoring data to predict the resilience of ecosystem services, followed by the development of a pilot genetic diversity monitoring scheme for the meadow brown butterfly, *Maniola jurtina* (Linnaeus 1758), over time in England. This will then be extended across Europe to measure the genetic diversity of the species at a much larger spatial scale. Finally, a combination of genetic, experimental and long-term monitoring data will be used in combination to investigate the phenology of *M. jurtina*.

1.2 Ecosystem services

Increasing recognition is being given to ecosystem services i.e. the benefits that people derive from ecosystems, in both scientific and political agendas (Millennium Ecosystem Assessment 2005; Geijzendorffer & Roche 2013). This is reflected in the CBD's strategic goal D:

> ***Strategic Goal D: Enhance the benefits to all from biodiversity and ecosystem services*** (Convention on Biological Diversity 2011a)

Ecological systems provide a multitude of services and functions that humanity depends upon (Díaz *et al.* 2013), including climate regulation, food and water provisioning, flood prevention and cultural well-being (Costanza *et al.* 1997). These services are currently predicted to have a value globally of over US$125 trillion per year (Costanza *et al.* 2014). Biodiversity is a crucial factor in the maintenance of ecosystem services (Hector & Bagchi 2007; Isbell *et al.* 2011; Cardinale *et al.* 2012; Harrison *et al.* 2014; Lefcheck *et al.* 2015), however as a result of human activities global biodiversity is reducing rapidly (Pimm *et al.* 2014; Tittensor *et al.* 2014; Newbold *et al.* 2015).

Understanding how ecosystem services will respond to changes in species assemblages is regarded as an urgent priority (Díaz et al., 2013; Oliver et al., 2015; De Palma et al., 2017). Without this understanding, mitigation methods cannot be put in place to reduce the negative impacts on society of ecosystem service loss (Oliver et al., 2015). Until now attempts to quantify the resilience of ecosystem services have relied upon trait data and changes in community composition (Díaz *et al.* 2013), which is both time consuming and results in statistically weak predictions. The use of monitoring data to address these concerns could provide the solution, allowing scientists to predict the resilience of specific ecosystem services, without the need for detailed trait data collection.

Interactions between organisms and their environments result in changes in population dynamics (Wallner 1987) and these changes can be observed using long-term monitoring data. Not all species respond to the environment in the same way; covariance in the population dynamics of any two species is determined by a multitude of factors including species interactions, responses to environmental change, and in the fundamental aspects governing population growth (e.g. intrinsic rate of population increase; Birch, 1948; Wallner, 1987; Walther et al., 2002; Loreau & de Mazancourt, 2013).

Some species will respond more similarly to environmental factors than others. Species that respond similarly to the environment will both increase and decrease in population abundances at similar times, as a result of responding to similar drivers. In instances where such similarly responding species carry out the same ecosystem services e.g. pollination, declines in species numbers could reduce the overall provision of the ecosystem service. In contrast species with asynchronous population dynamics performing the same ecosystem services are expected to result in more stable ecosystem functioning and subsequent ecosystem service provision (Ives *et al.* 1999; Yachi & Loreau 1999; Loreau & de Mazancourt 2013). Where biodiversity is high it is more likely ecosystem servicing will be stable (Lefcheck *et al.* 2015), as it increases the likelihood that multiple

species within an environment performing similar ecosystem services will respond differently to the environment.

Long-term population monitoring data offer a chance to determine firstly which species share similar population dynamics and secondly the vulnerability of specific ecosystem services. Using long-term monitoring data offers a new approach to tackle a problem that has been ongoing since the 1990s, with little progress (Díaz & Cabido 1997; Lavorel *et al.* 1997; Lavorel & Garnier 2002; Funk *et al.* 2017).

1.3 How landscape configuration affects biodiversity

Landscapes can be defined as ecological systems at the kilometre scale (Pickett & Cadenasso 1995) and the structure of a landscape directly affects the movement of organisms within it (Sutherland *et al.* 2015a). Loss of habitat within a landscape and the resultant habitat fragmentation are major drivers of species loss (Fischer & Lindenmayer 2007; Pimm *et al.* 2014), although there is some debate as to whether the fragmentation of habitats and landscape configuration is more or less important than the area of habitat within a landscape (Fahrig 2003, 2017; Hanski 2015; Crooks *et al.* 2017; Fletcher *et al.* 2018; Fahrig *et al.* 2019). A key feature of a landscape is the matrix; an area of non-habitat surrounding native habitat patches of interest (Ricketts 2001). Quantifying the impacts of the matrix is difficult, as the matrix is made up of patches with different levels of resistance to movements for different species (Ricketts 2001; Watts *et al.* 2010).

Habitat fragmentation is the reduction of areas of continuous habitat into smaller habitat patches surrounded by areas of hostile matrix (Young *et al.* 1996; Crooks *et al.* 2017). Changes to landscape configuration through fragmentation can reduce a landscape's level of connectivity (Delattre *et al.* 2013a) i.e. the degree to which a landscape facilitates or impedes the movement of an organism (Kadoya 2009). This can influence multiple ecological factors including species and population abundance, species distribution and dispersal, genetic diversity, species persistence and the ability of individuals to search for areas of suitable habitat (Fahrig 2003; Schooley & Branch 2011; Fletcher *et al.* 2013). These in turn can have knock on effects which can alter ecosystem processes and functions (Young *et al.* 1996).

The preservation and maintenance of heterogeneous habitat is a key aim in conservation biology. This includes the reduction of habitat fragmentation and increasing levels of connectivity between habitat patches (Oliver *et al.* 2010; Spear *et al.* 2010; Watts *et al.* 2010). Changing aspects of the matrix to increase species movement is one method of increasing connectivity, as described in Eycott *et al.* (2012). In that study movement between habitat patches was shown to be higher when surrounding matrix is similar to habitat. However, whilst a matrix with lower levels of resistance can promote connectivity (Baum et al., 2004), high resistances can also lead to increased connectivity as individuals are more likely to move rapidly in unfavourable habitat (Driscoll et al., 2013).

Habitat loss and fragmentation affect biodiversity in a number of ways. Firstly, reducing connectivity can lead to losses in genetic variation (Ferrari *et al.* 2007; Galpern *et al.* 2012). Where populations become increasingly isolated as a result of increasing distances between habitat patches, dispersal between the populations reduces due to increasing dispersal costs (Delattre *et*

al. 2013a). This results in a reduction in gene flow, which in turn can lead to increasing levels of inbreeding and a loss of alleles through genetic drift, resulting in reductions in genetic variation within populations and increasing genetic differentiation between populations (Honnay *et al.* 2006). In the short term this can reduce the fitness of individuals (Young *et al.* 1996), whilst in the long-term reductions in genetic variability can limit a species ability to adapt to changes in an environment and reduce the likelihood of a species to survive into the future (Honnay *et al.* 2006; Segelbacher *et al.* 2010). These effects are not exclusive to anthropomorphic habitat fragmentation. For example natural landscape features like rivers, forests and mountain ranges can act as barriers to gene flow for certain species, increasing the levels of genetic differentiation between populations (Storfer *et al.* 2007).

As well as affecting genetic diversity, habitat loss and fragmentation can affect biodiversity in other ways. Habitat loss reduces the size of habitat patches, reducing both the number of species and number of individuals the patch can support (Fahrig 2003). Decreasing population sizes can reduce densities of individuals to a point where numbers are too low to support breeding populations (Fischer & Lindenmayer 2007), resulting in local extinctions and biodiversity loss. As ecosystem service provision is enhanced by biodiversity (Lefcheck *et al.* 2015), these changes to the number of individuals and species within a landscape can dramatically affect the composition of ecosystems and therefore the functions and services that they provide (Crooks *et al.* 2017). Additionally, reductions in genetic diversity as a result of habitat loss can also have a negative impact on ecosystem services and ecosystem functioning (Hajjar *et al.* 2008; Bailey 2011).

Landscape configuration clearly has a major impact on biodiversity and genetic diversity, both of which contribute to the services we derive from ecosystems. The best way that landscapes can be managed for wildlife and therefore biodiversity, genetic diversity and ecosystem service provision is summed up by the phrase "more, bigger, better and joined" (Lawton *et al.* 2010). An increase in the number of good quality, large habitats with high levels of connectivity between them offers the best possible landscape configuration for the maintenance of nature and the services upon which humans depend.

1.4 Landscape ecology

The goal of landscape ecology is to understand gene flow, organism movement and population fluctuations with respect to the landscape and subsequently influence conservation management strategies (Bélisle 2005). First termed by Troll (1939), the field of landscape ecology developed in the 1960s with the combination of multiple disciplines including geography, ecology and land management (Wiens *et al.* 1993; Wu & Hobbs 2002). During the 1980s and 1990s landscape ecology grew rapidly, especially with the introduction of population genetics (Turner 2005; Lianyong & Eagles 2009).

1.4.1 Landscape connectivity

The concept of landscape connectivity was first used by Merriam (1984) and defined by Taylor *et al.* (1993) as "the degree to which the landscape facilitates or impedes movement among resources

patches". Landscape connectivity can be split into two areas: structural connectivity, the arrangement of habitat, and functional connectivity, how organisms move through a landscape (Ferrari *et al.* 2007).

Landscape connectivity allows individuals to move between patches if a path exists between them (Minor & Urban 2008; Matisziw *et al.* 2015), this includes landscape elements that act as stepping stones or corridors (Vogt *et al.* 2009). Landscape connectivity is influenced by habitat fragmentation, potentially resulting in species declines (Schumaker 1996) and reducing the likelihood of population recovery or recolonisation after local extinctions (Estrada & Bodin 2008).

One method of calculating structural connectivity is by calculating the effective distance between two patches. This method takes the cost of movement into account and determines the path of least resistance through a landscape (Adriaensen *et al.* 2003). A common method of doing this is by turning a landscape into a graph made up of nodes (habitat patches) and edges (connections between patches) (Urban & Keitt 2001; Minor & Urban 2008). The weight of the edge can be used to give an estimation of the connectivity of the landscape (Kadoya 2009; Matisziw *et al.* 2015). Using graph theory the least cost path can be calculated (Ferrari *et al.* 2007). An advantage of this method is that relatively little data is required, only species habitat requirements, dispersal distances and patch coordinates (Kadoya 2009). Estimations of resistance values can be poor using this method, although new techniques using genetic data are leading to improvements (Peterman 2018). Additionally a landscape's spatial heterogeneity may limit interpretation (Fortin *et al.* 2012).

1.4.2 Functional connectivity

Whilst structural connectivity looks at the effects of landscape on species movement, functional connectivity includes the attributes of the species of interest and their responses to an environment (Rodriguez Gonzalez *et al.* 2008; Watts & Handley 2010; Auffret *et al.* 2015). A key feature of functional connectivity is the incorporation of the effects of the matrix on species movement (Fitzgibbon *et al.* 2007; Stevenson-Holt *et al.* 2014). The primary reason for this is that different species respond differently to the same habitats and therefore the connectivity of a landscape does not have a fixed value (DiLeo & Wagner 2016). Landscapes with higher levels of functional connectivity have higher probabilities of individual movements between patches (Bélisle 2005).

It is possible for a landscape to be functionally, but not structurally connected. For example if two habitat patches are separated by a matrix, but organisms are able to move between patches, they are functionally connected (Tischendorf & Fahrig 2000). The functional connectivity of a landscape can be increased by changing the structure of the landscape to increase the amount of movement (Kadoya 2009). This can be done by increasing the permeability of the matrix between patches (Powney *et al.* 2011).

An additional method of increasing functional connectivity between patches would be to increase the population size within a patch. This would increase competition between individuals, leading to density dependent dispersal (Matthysen 2005), providing sufficient additional habitat requirements were available within an individual's dispersal range. Whilst this may be an option in bird and mammal species, this isn't always the case for butterfly species. Roland *et al.* (2000) showed

that alpine *Parnassius sp.* more readily emigrated from sites with small populations and individuals from sites with large populations were more likely to stay in the same habitat patch.

In order for habitat patches to function as a network, they must be connected (Lawton *et al.* 2010). For this reason the Lawton review; "*An Independent Review of England's Wildlife Sites and the Connections Between Them*", was commissioned by the UK government and published in 2010 (Department for Environment Food & Rural Affairs 2010). The report recommends the enhancement of connections between sites and suggests the use of both existing connections such as rivers as well as manmade structures such as roadsides and cycle ways (Lawton *et al.* 2010).

1.5 Genetic diversity

Genetic diversity, the amount of genetic variability within a population (Hughes *et al.* 2008), underpins all other levels of biodiversity including species, habitats, traits and ecosystem services (Bruford *et al.* 2017). Without genetic variation, populations are unable to adapt to changes in an environment due to a lack of evolutionary resilience (Boettcher *et al.* 2010; Bruford *et al.* 2017). Populations containing higher levels of allelic diversity present greater levels of phenotypic variation and these levels of variation give species a greater chance to adapt (McGill *et al.* 2015). Where genetic diversity is low, as with endangered species which have undergone population bottlenecks or prolonged inbreeding, reproductive fitness can decrease, resulting in a reduced ability of a species to persist (Hutchinson *et al.* 2003; Hughes *et al.* 2008). Therefore, genetic diversity should be a major aim of any conservation management project. The problem, however, is that for the majority of wild species very little is known about the levels of genetic diversity; let alone how to conserve it.

1.5.1 The current state of genetic diversity monitoring

The importance of genetic diversity in maintaining biodiversity has been acknowledged for decades, but is often overlooked when it comes to creating policies and management strategies (Laikre 2010; Laikre *et al.* 2020). This is in spite of the fact than an estimated US$2-4.5 trillion in genetic resources are lost annually, which could be catastrophic for both ecology and economics (Hoban *et al.* 2013). However, the contribution of genetic diversity to ecosystems is gaining increasing recognition as scientists monitor the effects of human activities on global ecosystems (Schwartz *et al.* 2007; Hoban *et al.* 2013). This increasing awareness can be seen with the CBD's Strategic Goal C and Aichi Target 13:

CBD Strategic Goal C: ***To improve the status of biodiversity by safeguarding ecosystems, species and genetic diversity*** (Convention on Biological Diversity 2011a).

Aichi Target 13: ***By 2020, the genetic diversity of cultivated plants and farmed and domesticated animals and of wild relatives, including other socio-economically as well as culturally valuable species, is maintained, and strategies have been***

developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity" (Convention on Biological Diversity 2011a).

A proposed update to this strategic goal can be seen in the recently published CDB post-2020 draft (Convention on Biological Diversity 2020d, c). This draft contains five key goals, of which the maintenance of genetic diversity is one:

Preliminary Draft Monitoring Framework for the 2030 and 2050 Goal 3: ***Genetic diversity is maintained or enhanced on average by 2030, and for [90%] of species by 2050*** (Convention on Biological Diversity 2020d).

Both the original strategic goal Aichi Target and proposed update have come under criticism, largely due to their focus on domesticated species (Laikre 2010; Laikre *et al.* 2020) and a lack of explicit commitments to the monitoring of genetic diversity (Convention on Biological Diversity 2020c; Laikre *et al.* 2020).

Genetic diversity monitoring schemes can be defined as schemes which 'quantify the temporal changes in the genetic metrics of a population' (Schwartz *et al.* 2007) and clearly the necessity for such schemes is widely recognised (Boettcher *et al.* 2010). However, because policy is more often built upon monitoring changes in population numbers rather than genetic diversity, genetic diversity is one of the least reported measures of biodiversity globally (Pereira *et al.* 2013; Geijzendorffer *et al.* 2016).

Despite a general lack of monitoring, there are cases where genetic diversity has been well studied. These examples tend to focus on domestic or socio-economically important species. This is largely due to the fact that the genetic diversity of crops and domestic animals will play an important role in maintaining and increasing food supplies in the face of climate change (Bruford *et al.* 2017). A major concern with this is that the global genetic diversity of livestock is decreasing, in part due to intensive selective breeding programmes (Boettcher *et al.* 2010). This decrease in genetic diversity could reduce the ability of breeds to survive in a changing climate.

Although rare, there are examples of non-crop/non-domestic species genetic diversity monitoring schemes. These include North Sea cod (*Gadus morhua*) (Hutchinson *et al.* 2003; Poulsen *et al.* 2006), leopard frog (*Rana pipiens*) (Hoffman & Blouin 2004) and red deer (*Cervus elaphus*) (Nussey *et al.* 2005). These studies found the genetic structure to be stable over time in both *G. mohua* and *R. pipiens* but declining in *C. elaphus*.

Recent advances in molecular technology such as next generation sequencing (NGS) have the potential to revolutionise biodiversity monitoring schemes (Creer *et al.* 2016) and conservationists are increasingly recognising the value of genomics in conservation biology (Corlett 2017). Whilst some of these techniques have limited application to genetic diversity monitoring, being able to tell little more than whether a species is present or absent in an environment, the use of such techniques could revolutionise traditional biodiversity monitoring (Lawson Handley 2015). When it comes to using molecular markers to monitor genetic diversity, an additional problem is

that whilst we have the technology and the statistical methods, there is no clear consensus on the best protocols or criteria to follow (Pereira *et al.* 2013), or even which markers are suitable for use and cost effective.

1.5.2 Molecular techniques for monitoring genetic diversity

Over the last fifty years genetic diversity monitoring techniques have developed markedly from simple statistical comparisons of soluble enzymes, to whole genome sequencing at increasingly reducing costs (Bruford *et al.* 2017). The first widely used molecular method for estimating genetic diversity in ecology was protein electrophoresis in the 1960s (Leberg 1996). Since then, many techniques have been developed, used and been superseded by advances in technology. These include Random Amplified Polymorphic DNAs (RAPDs), Restriction Fragment Length Polymorphisms (RFLPs) and Amplified Fragment Length Polymorphisms (AFLPs) (Williams *et al.* 1990; Vos *et al.* 1995; Leberg 1996; Freeland *et al.* 2011).

Another example, microsatellites, have been the molecular marker of choice for most ecological studies until recently, and are still extensively applied in genetic monitoring programs (Bruford *et al.* 2017). Compared to previous methods, fewer markers are required to gain the same amount of statistical information (Gerber *et al.* 2000; Selkoe & Toonen 2006). Microsatellites are tandemly repeated units of DNA, commonly one to six base pairs in length, distributed throughout the genome (Schlötterer 2000; Guichoux *et al.* 2011; Putman & Carbone 2014). In genetic studies di, tri and tetranucleotide sequences with between five and 40 repeats are most frequently used (Selkoe & Toonen 2006). Loci with five or fewer repeat units are not included as they are not considered polymorphic (Valdes *et al.* 1993). The discovery and first use of microsatellites arose in the 1980s (Litt & Luty 1989), along with the invention of the polymerase chain reaction (PCR) (Mullis & Faloona 1987).

Microsatellites possess a range of properties making them ideal for population genetic studies (Balloux & Lugon-Moulin 2002; Gardner *et al.* 2011). For example, microsatellites have mutation rates in the region of 10^{-3} to 10^{-5} events per locus per replication (compared to 10^{-9} to 10^{-10} for background point mutations) (Freeland *et al.* 2011). These mutation rates result in high levels of allelic diversity, allowing estimations of genetic diversity (Selkoe & Toonen 2006; Väli *et al.* 2008). Mutations in microsatellites commonly occur as a result of slipped-strand mispairing during DNA synthesis, causing the gain or loss of single repeat units in a step-wise fashion (Chambers & MacAvoy 2000).

Although microsatellites can occur within coding regions, most frequently with trinucleotide repeats that do not disrupt the open reading frame of the gene (Li *et al.* 2004), microsatellites are widely distributed throughout the genome and frequently occur within non-coding regions where there are less constraints on range expansion (Vieira *et al.* 2016). A further property of particular microsatellites in these non-coding regions is that they often show neutral genetic variation; whereby they have no effect on an individual's fitness (Holderegger *et al.* 2006). This makes them extremely useful for studying processes such as gene flow across a landscape. As a result of their high mutation rates, high abundance and selective neutrality, microsatellites have been used in

many studies to investigate genetic aspects of populations, including migration rates, population bottlenecks and the relatedness of individuals (Balloux & Lugon-Moulin 2002; Selkoe & Toonen 2006; Leclercq *et al.* 2010).

Until recently microsatellites were isolated by enrichment processes and screening enriched genomic DNA libraries with oligonucleotide primers, followed by Sanger sequencing and confirmation by PCR amplification of microsatellite containing fragments (Chambers & MacAvoy 2000). Recent technological advances have led to next generation sequencing of microsatellites taking over from enrichment methods. The identification of microsatellites in this way is faster and simpler, with reduced labour costs, resulting in far more loci being isolated (Gardner *et al.* 2011).

Once microsatellite regions have been identified within a genome, fluorescently labelled oligonucleotide PCR primers, or primers with fluoro-labelled tags ligated to them, can be designed for the flanking regions; sequences of base pairs either side of the microsatellite repeat, allowing amplification and detection (Selkoe & Toonen 2006).

Microsatellites have some limitations. Firstly, the relative specificity of microsatellites means that, more often than not, new primers have to be developed for most species (Glenn & Schable 2005), which can be both time and resource consuming. In some cases cross amplification can occur, but often only with closely related species (e.g. Cassel-Lundhagen *et al.* 2009; Jiang *et al.* 2014, 2016). Secondly, mutations in primer regions can result in amplification failure, a problem which affects some taxa more than others (Selkoe & Toonen 2006). A final problem with microsatellites is that datasets for the same species often cannot be combined due to differences in how the alleles are scored by different observers and computers (Bruford *et al.* 2015). These reasons, coupled with the decreasing costs of next generation sequencing profiling techniques such as single nucleotide polymorphisms (SNPs), mean that microsatellites are rapidly being replaced as the molecular marker of choice for genetic monitoring.

SNPs are a next generation sequencing technique which use codominant markers and allow the sequencing of millions of DNA fragments rapidly and cheaply compared to other techniques (Helyar *et al.* 2011; Creer *et al.* 2016). SNPs have rapidly taken over from microsatellites and are currently the preferred molecular marker for many landscape genetics projects because they are abundant throughout the genome and many can be identified for the cost of developing a much smaller number of microsatellites (Bruford *et al.* 2017). Whilst microsatellites often have greater levels of allelic diversity, SNPs are still able to be used to segregate between populations (Helyar *et al.* 2011). Additionally SNPs can be more directly comparable between labs (Corlett 2017), which can be a problem with microsatellite datasets.

Finding SNPs can also be both financially costly and time consuming to analyse, as it requires many genes within a genome to be sequenced. However, as an increasing number of DNA sequences are being made available in open access DNA databases, the time and financial resources required for SNP detection is reducing (Holderegger & Wagner 2008). SNPs are not without their problems, as they are not yet available for all species and can be subject to ascertainment bias (Bruford *et al.* 2015). Additional problems with SNPs include biases in the coverage of certain parts of genomes and a lack of a reference genome. However technological advances are reducing the

number of problems caused by such issues (Helyar *et al.* 2011). Furthermore, the rapidly reducing costs of whole genome sequencing could result in SNPs also becoming outdated. A challenge of molecular ecology will be to incorporate previous data from older techniques such as microsatellites into monitoring schemes using SNPs or whole genome sequencing, so as not to waste previous effort (Bruford *et al.* 2015).

Another next generation sequencing technique with huge potential for population genetics studies is Restriction site-Associated DNA sequencing (RADSeq) (Baird *et al.* 2008). RADSeq allows the simultaneous discovery and scoring of hundreds of thousands of SNPs in many individuals for relatively low investment (Etter *et al.* 2011). RADSeq results in the amplification of a subsample of specific sites across the genome, defined by restriction endonucleases (Davey & Blaxter 2010). The combination of only a small percentage of a target genome being covered and the ability to be used without a reference genome makes RADSeq an extremely useful tool for the discovery and genotyping of SNPs in non-model organisms (van Dijk *et al.* 2014).

Finally, the ability to sequence entire genomes offers a number of advantages to genetic monitoring (Hudson 2008). A major benefit of whole genome sequencing is that it provides far more data, allowing hugely powerful statistical techniques to be applied. This improves the reliability of studies and increases the number of potential investigations (Ekblom & Wolf 2014). Furthermore the process is relatively simple and can be automated (Bruford *et al.* 2015). A current issue with whole genome sequencing is that it remains costly compared to other molecular techniques (Bruford *et al.* 2017). However the costs are reducing rapidly, as is the practicality, with new technologies becoming portable and able to be used in the field (Ekblom & Wolf 2014; Bruford *et al.* 2015).

1.5.3 Landscape genetics

Landscape genetics combines population genetics and landscape ecology (Manel *et al.* 2003) and is defined as "research that explicitly quantifies the effects of landscape composition, configuration and matrix quality on gene flow and spatial genetic variation" (Storfer *et al.* 2007). It does this by assessing how a landscape facilitates an organism's movement in relation to gene flow and landscape structure, including how anthropogenic elements act as barriers to gene flow (Storfer *et al.* 2007; Galpern *et al.* 2012). This can be directly applied to conservation management (Manel & Holderegger 2013). For example, once a barrier to gene flow has been identified, landscape elements, such as corridors, can be introduced to increase individual movement and gene flow. This benefits populations by potentially increasing genetic variation (Segelbacher *et al.* 2010), thereby reducing the impacts that a loss of genetic variation through genetic drift and inbreeding depression would have on individual fitness (Lacy 1997).

Genetic connectivity, the movement of genes between populations, is maintained by the movement of individuals between populations (Sork & Smouse 2006). Gene flow i.e. the movement of genes across a landscape, helps to ensure the persistence of small populations and is influenced by the makeup of a landscape and permeability of the surrounding matrix (DiLeo & Wagner 2016). Gene flow is dependent upon successful reproduction after migration and can be transferred over

multiple generations, meaning that spatially distant habitat patches can be connected temporarily (Spear *et al.* 2010).

To study genetic connectivity, data are collected from individuals within landscapes, most commonly, to date, using microsatellite molecular markers (Storfer *et al.* 2010; Prunier *et al.* 2013). The highly variable nature of microsatellites makes them ideal for studies investigating current landscape changes (Anderson *et al.* 2010). However, as genetic data are a result of the influence of both historic and current landscape features, the current landscape may not accurately explain genetic results (Balkenhol *et al.* 2009a). Similarly, landscapes can undergo rapid change. This can result in a lag between landscape data and genetic data (Anderson *et al.* 2010). For example, species with low dispersal rates and long generation times may not show the genetic effects of a new landscape barrier instantly (Landguth *et al.* 2010). Therefore lag time must be taken into account when selecting molecular markers (Holderegger & Wagner 2008; Balkenhol *et al.* 2009b). Landscape genetics can be used to calculate genetic distances (Storfer *et al.* 2010; Manel & Holderegger 2013) and quantify the functional connectivity of a landscape by investigating features that enhance or inhibit gene flow (DiLeo & Wagner 2016).

The development of Geographic Information Systems (GIS) has greatly benefited all aspects of landscape ecology. With GIS, landscapes are converted into raster grids where each pixel is assigned a value of resistance depending upon the permeability of the habitat type that the pixel represents (Etherington 2011; Stevenson-Holt *et al.* 2014). Resistance values are hypothesised reductions in gene flow, most commonly determined by expert opinion (Spear *et al.* 2010; Galpern *et al.* 2012; Zeller *et al.* 2012).

Paths of least resistance for individuals or gene flow are calculated using these grids (Rodriguez Gonzalez *et al.* 2008; Hanks & Hooten 2013). Under least cost paths movement/gene flow occurs along a single path (Spear *et al.* 2010) with species specific resistances attributed to features of the landscape (Holderegger & Wagner 2008). An alternative to least cost path analysis is circuit theory (Hanks & Hooten 2013), which builds upon graph theory and uses concepts from electrical systems (McRae *et al.* 2008). Circuit theory looks at all possible paths across a raster grid as gene flow is not limited to a single pathway in a landscape (Spear *et al.* 2010). This is more applicable to natural situations as multiple, wider habitat corridors allow greater levels of gene flow (McRae & Beier 2007).

1.6 Butterflies as indicators of biodiversity

Butterflies possess a number of traits that make them useful indicators of insect biodiversity. Firstly, they are a popular taxonomic group which can be easily identified by members of the public (Van Swaay *et al.* 2008). Secondly, they have been shown to be suitable indicators for other terrestrial insects (Thomas 2005). Finally, and perhaps most importantly, there are robust, long-term monitoring data available for statistical analysis, with the number of monitoring sites and monitoring schemes increasing globally (Van Swaay *et al.* 2008). These three factors mean that butterflies are ideal biological indicators of biodiversity. Therefore, it makes sense to use butterflies as a pilot taxon

to develop a genetic monitoring scheme and take advantage of the vast amount of data already available.

1.6.1 Microsatellite isolation in Lepidoptera

The greater the number of microsatellite markers used in a study, the more reliable the results are (Selkoe & Toonen 2006; Bruford *et al.* 2015). Since the development of microsatellites in the 1980s, the isolation and characterisation of microsatellites in Lepidoptera has been historically problematic. This has resulted in a paucity of loci detected compared to related taxa (Harper *et al.* 2000, 2003; Anthony *et al.* 2001; Keyghobadi *et al.* 2002; Meglécz *et al.* 2004; Zhang 2004; Van't Hof *et al.* 2007; Beldade *et al.* 2009; Mikheyev *et al.* 2010; Sinama *et al.* 2011). The isolation of two microsatellites from the checkerspot butterfly (*Melitaea cinxia*) by Palo *et al.* (1995) was the first case of microsatellite isolation in Lepidoptera. After a decade of study, isolation rates remained low with 80% of attempts to isolate microsatellites in Lepidoptera yielding less than five microsatellites per study (Zhang 2004). Up to 2011, only economically important pests or model species had more than ten microsatellites reported (Sinama *et al.* 2011). Recent advances and refinements in microsatellite isolation techniques have led to an increase in Lepidoptera microsatellite discoveries, however isolation still remains difficult compared to other taxa (Aarnes *et al.* 2015; Cao *et al.* 2015; Jiang *et al.* 2016).

The problems with microsatellite isolation and characterisation in Lepidoptera are the result of multiple factors. Firstly compared to other taxonomic groups Lepidoptera have a lower frequency of microsatellite repeat motifs (Meglécz & Solignac 1998; Zhang 2004; Van't Hof *et al.* 2007). However a recent study by Cao *et al.* (2015) found that this low frequency was not universal across the entire order, something that had been suggested a decade ago (Ibrahim *et al.* 2004).

Another factor that makes Lepidoptera microsatellite development difficult are high levels of null alleles that do not amplify and result in high levels of PCR failure (Bogdanowicz *et al.* 1997; Meglécz & Solignac 1998; Harper *et al.* 2003; Meglécz *et al.* 2004; Tay *et al.* 2010; Sinama *et al.* 2011). Isolation of microsatellites requires PCR amplification using PCR primers to bind to flanking regions (Callen *et al.* 1993). Null alleles are the result of mutations in the flanking regions, preventing primer binding and subsequent amplification (Dakin & Avise 2004). This frequently results in false homozygotes, leading to deviations from the Hardy Weinberg Equilibrium (Van Oosterhout *et al.* 2004). Null alleles leading to PCR failure, homozygote excess and large deviations from Hardy Weinberg Equilibrium, have been frequently recorded in Lepidoptera studies (Bogdanowicz *et al.* 1997; Anthony *et al.* 2001; Keyghobadi, Roland & Strobeck 2002; Harper, Maclean & Goulson 2003; Ji, Wu & Zhang 2005; Anderson, Dawson & Freeland 2006; Mikheyev *et al.* 2010).

A third reason for difficulties in Lepidoptera microsatellite isolation are repetitive sequences in flanking regions, something that appears to be prevalent in Lepidoptera microsatellites (Zhang 2004). Lepidoptera have highly conserved flanking regions, with many microsatellite motifs sharing similar flanking regions (Mikheyev *et al.* 2010), or microsatellites occurring within minisatellite repeats (Meglécz *et al.* 2004). This makes primer design for specific flanking regions extremely

difficult, as one primer design may result in multiple sequences amplified during PCR, creating multiple, uninterpretable banding patterns during electrophoresis (Tay *et al.* 2010).

A final difficulty in Lepidoptera microsatellite isolation is the association of microsatellites with transposable elements (TE) (Tay *et al.* 2010), which are able to multiply and insert themselves throughout the genome (Coates *et al.* 2010). Propagation of microsatellites through the genome via TE insertion can result in microsatellites at different locations having similar flanking regions. In *Bombyx mori*, a lepidopteran species with an unusually large number of identified microsatellites (Miao *et al.* 2005), 12% are associated with TEs (Tay *et al.* 2010).

The issues of isolation and characterisation are still problematic and result in relatively few usable microsatellites (Table 1.1). However, the versatility of usable microsatellites makes them a hugely useful tool in population and genetic studies, making the effort worthwhile if an end product can be achieved.

Table 1.1 Selected microsatellite studies of Lepidoptera from which microsatellites have been successfully characterised for population studies. References for each study are given at the bottom of the table. Microsatellites characterised per study are shown in square brackets.

Family / Species
Bombycidae
Bombyx mori (Silkworm) Refs 1, 2, 3 [15, 36, 518]
Carposinidae
Carposina sasakii (Peach fruit moth) Ref 4 [35]
Crambidae
Chilo suppressalis (Rice stem borer) Refs 5, 6 [4, 12]
Diatraea saccharalis (Sugarcane borer) Ref 7 [16]
Ostrinia nubilalis (European corn borer) Refs 8, 9, 10 [10, 13, 5]
Saucrobotys futilalis (Panic moth) Ref 11 [12]
Erebidae
Arctica caja (Garden tiger moth) Ref 12 [7]
Lymantria dispar (Gypsy moth) Ref 13 [4]
Parasemia plantaginis (Wood tiger moth) Ref 14 [10]
Utetheisa ornatrix (Bella moth) Ref 15 [250]
Zale galbanata (Maple Zale) Ref 16 [5]
Hyphantria cune (Fall webworm) Ref 17 [48]
Gelechiidae
Pectinophora gossypiella (Pink bollworm) Ref 18 [13]
Tuta absoluta (Tomato leaf miner) Ref 19 [8]
Geometridae
Biston betularia (Peppered moth) Ref 20 [14]
Chiasmia assimilis Ref 21 [12]
Epirrita autumnata (Autumnal moth) Ref 22 [21]
Gracillariidae
Cameraria ohridella (Horse chestnut leaf mining moth) Ref 23 [6]
Hesperiidae
Erynnis propertius (Propertius duskywing) Ref 24 [15]
Polytremis fukia Ref 25 [11]
Polytremis nascens Ref 26 [12]
Lasiocampidae
Dendrolimus pini (Pine-tree lappet moth) Ref 27 [10]
Dendrolimus punctatus (Masson pine moth) Ref 28 [10]

Gonometa postica (African wild silk moth) Ref 29 [6]
Lycaenidae
Arhopala epimuta (Common disc oakblue) Ref 30 [5]
Cyclargus thomasi bethunebakeri (Miami blue butterfly) Ref 31 [12]
Drupadia theda (Dark posy) Ref 30 [5]
Maculinea alcon (Alcon large blue) Ref 32 [1]
Maculinea nausithous (Dusky large blue) Ref 32 [11]
Plebejus melissa samuelis (Karner blue) Ref 33 [4]
Polyommatus bellargus (Adonis blue) Ref 34 [5]
Lycaena helle (Violet copper) Ref 35 [6]
Noctuidae
Busseola fusca (Maize stalk borer) Ref 36 [8]
Chrysodeixis includes (Soybean Looper) Ref 37 [13]
Helicoverpa armigera (Cotton bollworm) Refs 38, 39, 40, 41, 42 [5, 5, 5, 5, 30]
Helicoverpa zea (Corn earworm) Ref 43 [13]
Heliothis virescens (Tobacco budworm) Ref 44 [15]
Spodoptera exigua (Beet armyworm) Ref 45 [10]
Spodoptera frugiperda (Fall armyworm) Refs 46, 47 [174, 6]
Spodoptera litura (Common Cutworm) Ref 48 [9]
Notodontidae
Thaumetopoea pinivora (Northern pine processionary moth) Ref 49 [13]
Thaumetopoea pityocampa (Pine processionary moth) Refs 50, 51 [5, 17]
Nymphalidae
Argynnis Niobe (Niobe fritillary) Ref 52[10]
Bicyclus anynana (Squinting bush brown) Ref 53 [28]
Boloria aquilonaris (Cranberry fritillary) Ref 54 [15]
Boloria Eunomia (Bog fritillary) Ref 55 [15]
Brenthis ino (Lesser marbled fritillary) Refs 56, 57 [11, 16]
Coenonympha hero (Scarce heath butterfly) Ref 58 [7]
Dione moneta (Passionflower butterfly) Ref 59 [19]
Erebia palarica (Chapman's ringlet) Ref 60 [10]
Euphydras aurinia (Marsh fritillary) Refs 61, 62[5, 12]
Euphydryas edutha (Edith's checker) Ref 63 [10]
Heliconius erato (Red postman) Ref 64 [15]
Maniola jurtina (Meadow brown) Ref 65 [15]
Melitaea ambigua Ref 66 [9]
Melitaea cinxia (Glanville fritillary) Refs 67, 68 [2, 5]
Melitaea protomedia Ref 66 [9]
Speyeria idalia (Regal fritillary) Ref 69 [4]
Papilionidae
Papilio zelicaon (Anise swallowtail) Ref 70 [17]
Parnassius apollo (The Apollo butterfly) Refs 71, 72 [6, 26]
Parnassius Mnemosyne (Clouded Apollo) Refs 73, 74 [3, 5]
Parnassius smintheus (Rocky mountain Apollo) Refs 75, 76 [4, 4]
Pieridae
Mylothris jacksoni knutsoni Ref 77 [8]
Psychidae
Dahlica fennicella Ref 78 [11]
Pyralidae
Plodia interpunctella (Indian meal moth) Ref 79 [25]
Saturniidae
Antheraea assama (Indian golden silkmoth) Ref 80 [87]
Graellsia isabelae (Spanish moon moth) Ref 81[10]

Sphingidae

Hyles tithymali (Spurge hawkmoth) Refs 82, 83 [11, 8]

Tineidae

Pringleophaga marioni (Marion flightless moth) Ref 84 [15]

Tortricidae

Choristoneura fumiferana (Spruce budworm) Ref 85 [8]

Choristoneura occidentalis Ref 85 [8]

Cydia pomonella (Codling moth) Refs 86, 87 [24, 17]

Lobesia botrana (European grape vine moth) Refs 88, 89 [7, 11]

Rhyacionia leptotubula (Pine tip moth) Ref 90 [9]

Tortix viridana (Green oak leaf roller) Ref 91 [8]

Zeiraphera diniana (Larch tortrix) Ref 92 [6]

Yponomeutidae

Yponomeuta padellus (Orchard ermine) Ref 93 [9]

Zygaenidae

Reissita simonyi (Arabian burnet moth) Ref 94 [14]

Table References: 1. (Reddy *et al.* 1999) 2. (Prasad *et al.* 2005) 3. (Miao *et al.* 2005) 4. (Wang *et al.* 2016) 5. (Ishiguro & Tsuchida 2006) 6. (Liu *et al.* 2009) 7. (Pavinato *et al.* 2013b) 8. (Kim *et al.* 2008) 9. (Dalecky *et al.* 2006) 10. (Kim *et al.* 2008) 11. (Grant & Bogdanowicz 2006) 12. (Anderson *et al.* 2006) 13. (Bogdanowicz *et al.* 1997) 14. (Galarza *et al.* 2011) 15. (Bezzerides *et al.* 2004) 16. (Caldas *et al.* 2002) 17. (Cao *et al.* 2015) 18.(Liu *et al.* 2006) 19. (Bettaibi *et al.* 2013) 20. (Daly *et al.* 2004) 21. (Wardill *et al.* 2004) 22. (Aarnes *et al.* 2015) 23. (Marí-Mena *et al.* 2008) 24. (Zakharov *et al.* 2007) 25. (Jiang *et al.* 2016) 26. (Jiang *et al.* 2014) 27. (A'Hara & Cottrell 2013) 28. (Ji *et al.* 2005a) 29. (Delport *et al.* 2005) 30. (Fauvelot 2005) 31. (Saarinen *et al.* 2009) 32. (Zeisset *et al.* 2005) 33. (Anthony *et al.* 2001) 34. (Harper *et al.* 2000) 35. (Habel *et al.* 2008) 36. (Faure & Silvain 2005) 37. (Silva *et al.* 2019) 38. (Tan *et al.* 2001) 39. (Ji *et al.* 2003) 40. (Scott *et al.* 2004) 41. (Ji *et al.* 2005b) 42. (Venkatesan *et al.* 2016) 43. (Perera *et al.* 2007) 44. (Perera *et al.* 2011) 45. (Kim *et al.* 2012) 46. (Arias *et al.* 2011) 47. (Pavinato *et al.* 2013a) 48. (Wu *et al.* 2019) 49. (Cassel-Lundhagen *et al.* 2009) 50. (Rousselet *et al.* 2004) 51. (Sauné *et al.* 2015) 52. (Zima *et al.* 2013) 53. (Van't Hof *et al.* 2005) 54. (Vandewoestijne *et al.* 2012) 55. (Legrand *et al.* 2014) 56. (Abello *et al.* 2012) 57. (Lebigre *et al.* 2015) 58. (Cassel 2002) 59. (Massardo *et al.* 2012) 60. (Vila *et al.* 2009b) 61. (Petenian *et al.* 2005) 62. (Sinama *et al.* 2011) 63. (Mikheyev *et al.* 2010) 64. (Flanagan *et al.* 2002) 65. (Richard *et al.* 2015) 66. (Nakahama *et al.* 2015) 67. (Palo *et al.* 1995) 68. (Sarhan 2006) 69. (Williams *et al.* 2002) 70. (Zakharov & Hellmann 2007) 71. (Petenian *et al.* 2005) 72. (Mira *et al.* 2014) 73. (Meglécz & Solignac 1998) 74. (Gratton & Sbordoni 2009) 75. (Keyghobadi *et al.* 1999) 76. (Keyghobadi *et al.* 2002) 77. (Zima *et al.* 2014) 78. (Chevasco *et al.* 2012) 79. (Grace *et al.* 2005) 80. (Arunkumar *et al.* 2009) 81. (Vila *et al.* 2010) 82. (Hundsdoerfer *et al.* 2010) 83. (Mende *et al.* 2011) 84. (Groenewald *et al.* 2011) 85. (Lumley *et al.* 2009) 86. (Franck *et al.* 2005) 87. (Zhou *et al.* 2005) 88. (Amsellem *et al.* 2003) 89. (Reineke *et al.* 2015) 90. (Zhu *et al.* 2011) 91. (Schroeder *et al.* 2009) 92. (Delamaire *et al.* 2010) 93. (Voetdijk *et al.* 2007) 94. (Klütsch *et al.* 2003)

1.6.2 Justification for using butterflies

While there are considerable issues with microsatellite isolation in many Lepidoptera, using a taxonomic group with such robust monitoring data means that the benefits outweigh the negatives, providing microsatellites can be characterised. In this study we use the meadow brown butterfly, *M. jurtina*, as a study species for three reasons. Firstly, long running population data are available. Secondly, *M. jurtina* is a much-studied species, with considerable literature already published. Thirdly, microsatellite isolation has already been conducted resulting in the characterisation of fifteen loci (Richard *et al.* 2015). Consequently, investigations into genetic diversity are already underway, with Villemey *et al.* (2016) using these loci to investigate the effects of specific landscape factors on *M. jurtina* gene flow in France (see below).

1.7 The meadow brown butterfly, *Maniola jurtina*

1.7.1 Life history and background

M. jurtina is a common satyrine butterfly distributed across much of the western Palearctic, spanning from the Canary Islands to Kazakhstan, and Algeria to southern Scandinavia (Tolman & Lewington 1997). *M. jurtina* has undergone various declines across Europe over the past 20 years (Van Swaay *et al.* 2013), most severely in Finland, Malta, Luxembourg, Sweden and Russia (Asher *et al.* 2001). These declines are largely considered to be the result of changes in land use, such as the removal of hay meadows and increasing agricultural intensification (Asher *et al.* 2001), resulting in fragmented landscapes (Delattre *et al.* 2010). Despite these recent declines *M. jurtina* populations are now considered stable across much of Europe (Van Swaay *et al.* 2019) and it remains one of

the most abundant butterfly species in the United Kingdom (UK) (Fox *et al.* 2006), classified as least concern under the Red List of British butterflies (Fox *et al.* 2011).

Over the past 70 years *M. jurtina* has been the subject of a great deal of study and as a result much is already known about many aspects of its biology, including life history and ecology (Scali & Masetti 1973; Dowdeswell 1981; Brakefield 1982b, a; Haeler *et al.* 2014; Lebeau *et al.* 2016, 2018), historical distributions (Thomson 1987; Schmitt *et al.* 2005; Habel *et al.* 2009; Dapporto *et al.* 2011, 2014; Kreuzinger *et al.* 2015), evolutionary biology (Dowdeswell & Ford 1952; Creed *et al.* 1959; Forman *et al.* 1959; Dowdeswell *et al.* 1960; Dowdeswell 1962), and dispersal and movement behaviours (Conradt *et al.* 2000; Schneider *et al.* 2003; Delattre *et al.* 2010; Evans *et al.* 2019, 2020).

M. jurtina can be split into four subspecies in the British Isles based upon size, colour and distribution: the large and brightly coloured *M. jurtina splendida* occurs in Western Scotland, the largest, *M. jurtina iernes,* in Ireland, *M. jurtina cassileridum* is smaller, with clear striations on the underside of the hind wings and is found on the Isles of Scilly, finally *M. jurtina insularis* is found throughout England, Wales and southern Scotland (Dowdeswell 1981).

All subspecies are found in open grassland habitats (Schmitt *et al.* 2005), with eggs deposited upon a range of species within the grass family (*Poaceae*) (Delattre *et al.* 2010). After hatching the larvae feed on a range of grasses before overwintering (Ouin *et al.* 2008), emerging again in March. After re-emerging, the larvae feed at night until pupation occurs between May and August (Brakefield 1987). Adults typically fly from June to August, peaking in late July, (Delattre *et al.* 2010). A second wave of emergence is present in southern England during October on chalk grasslands, however no reproductive isolation between the two emergence peaks occurs (Goulson 1993b; Thomas & Lewington 2010).

Adult lifespans in the British Isles are estimated at between five and 12 days (Brakefield 1982b), during which time males and females exhibit different behaviours to the extent that they occupy different ecological niches (Dowdeswell 1981; Brakefield 1982a). Males show increased flight activity, keeping low to the vegetation, no higher than two meters from the ground (Ouin *et al.* 2008), in search of females to mate with (Brakefield 1982a). Alternatively, males will perch in the vegetation in order to intercept and mate with passing females (Lebeau *et al.* 2017). In contrast, females mate once, usually within the first 24 hours of emergence (Brakefield 1982a), and then spend the majority of their time feeding or depositing eggs (Delattre *et al.* 2010; Lebeau *et al.* 2016). Both sexes feed upon a range of nectar providing flower species, however they will preferentially feed upon knapweeds (*Centaurea sp.*) and thistles (*Cirsium sp.*) if available within an environment (Brakefield 1982a; Lebeau *et al.* 2018).

M. jurtina is a relatively sedentary species, with individuals typically remaining within closed populations (Ouin *et al.* 2008). On average adult *M. jurtina* individuals move around an area with a radius of 320 meters. However in mark-release-recapture studies, individuals have been found up to 2.1km away from where they were released (Schneider *et al.* 2003). Rather than moving randomly throughout a landscape *M. jurtina* have been shown to recognise boundaries (Conradt & Roper 2006) between areas of habitat e.g. meadows, and non-habitat e.g. harvested wheat fields (Delattre *et al.* 2010). Individuals display two distinct dispersal strategies. Firstly 'foray searching' in which

individuals fly in loops around a point of departure (Conradt *et al.* 2003; Delattre *et al.* 2013b); and secondly 'direct flight' where individuals fly in straight lines over longer distances. This behaviour is seen twice as often in females and is believed to be an adaptation for dispersal across areas of hostile habitats e.g. arable crops (Ouin *et al.* 2008; Delattre *et al.* 2010, 2013a).

1.7.2 *Maniola jurtina* as a model for evolutionary biology

In the 1940s *M. jurtina* became a key model organism in early ecological genetics research, with particular interest in variations in the number of hindwing spots on individuals within populations, both spatially and temporally (Dowdeswell 1961; Dowdeswell & McWhirter 1967; Brooke *et al.* 1985). Spotting patterns were first studied on the Isles of Scilly in 1946 by Dowdeswell *et al.* (1949). Here it was found that individuals on large islands had similar wing spot patterns, but differences occurred between individuals on smaller islands, such as the Isle of Tean. Additionally, ecological changes such as the removal of cattle on Tean, were found to alter previously stable wing spot pattern distributions (Dowdeswell *et al.* 1957). It was then concluded that wing spot variation was largely a result of island specific selection and that migration between the islands was therefore minimal. This was supported by mark-release-recapture results which found no migration between the islands of Tean and St Martins over a distance of less than 300 meters (Dowdeswell *et al.* 1949).

Subsequent investigations on mainland populations found that male spot distributions across eight southern British counties matched those found on Tean, but female distributions did not (Dowdeswell & Ford 1952). An exception to this was found at a boundary region in West Cornwall, with females south of this boundary exhibiting the same spot distribution seen on Tean (McWhirter 1957). This was suggested to be a hybridisation zone between two British races of *M. jurtina* (*M. jurtina cassiteridum* and *M. jurtina insularis*) (Clarke 1970).

Further investigation into *M. jurtina* spot patterns determined that spot number was controlled by polygenes (Brakefield 1984), groups of genes which also control other aspects of development. An example being that the genes involved in the maturation speed of the larvae also result in a high number of wing spots in the adult butterfly (Thomas & Lewington 2010). Furthermore differences in spot patterns between males and females were linked to their different behaviours and resource requirements (Brakefield 1982a). Studies also found that hindwing and forewing eyespot patterning were highly significantly correlated, with small forewing spots resulting in fewer hindwing spots. The combination of these factors led to the hypothesis that avian predation acted as a selective factor (Bengston 1981), with more active males selected for higher spots to distract predators, and less active females selected for lower spots and greater camouflage (Brakefield 1984). Habitat was also shown to account for the variation in spot number (Brakefield 1984), with uniform grassland sites selecting for fewer or no spots (Thomas & Lewington 2010).

More recently Baxter *et al.* (2017) investigated whether the distribution of hindwing spot patterns first observed by Dowdeswell *et al.* (1949) and then later by Handford (1973a) had remained constant. In addition to confirming that spot pattern distribution has remained relatively unchanged since the 1960s, they also carried out empirical tests to confirm whether the original assumptions

of minimal gene flow between islands was valid. By analysing 176 AFLPs they determined significant genome wide differentiation between islands (Baxter *et al.* 2017) (see below for further details).

1.7.3 *Maniola jurtina* as a model species for phylogeography

As well as being a model organism for population genetics studies, *M. jurtina* has also been the subject of a great deal of study regarding its phylogeography (Tauber 1970; Thomson 1973, 1987; Goulson 1993a; Schmitt *et al.* 2005; Habel *et al.* 2009; Richard *et al.* 2015; Villemey *et al.* 2016; Baxter *et al.* 2017). This is a result of contradictory patterns in both morphological and allozyme studies (Dapporto *et al.* 2011)

Early work on the subject was conducted by Tauber (1970), using morphological measurements and paleo-ecological theory. Tauber suggested that *M. jurtina* originated in Africa and spread into Europe by two dispersal routes, an eastern and a western one. Although the idea of an African origin was not widely accepted, subsequent morphological (Thomson 1973) and allozyme distribution studies (Thomson 1987) also suggested that two distinct lineages of *M. jurtina* occurred. As such it is generally accepted that *M. jurtina* can be split into a western Atlantic-Mediterranean lineage (*Maniola jurtina jurtina*) and an eastern-Mediterranean-Asian lineage (*Maniola jurtina janira*), with a hybridisation zone between the two in central Europe (Thomson 1987) (Fig. 1.1).

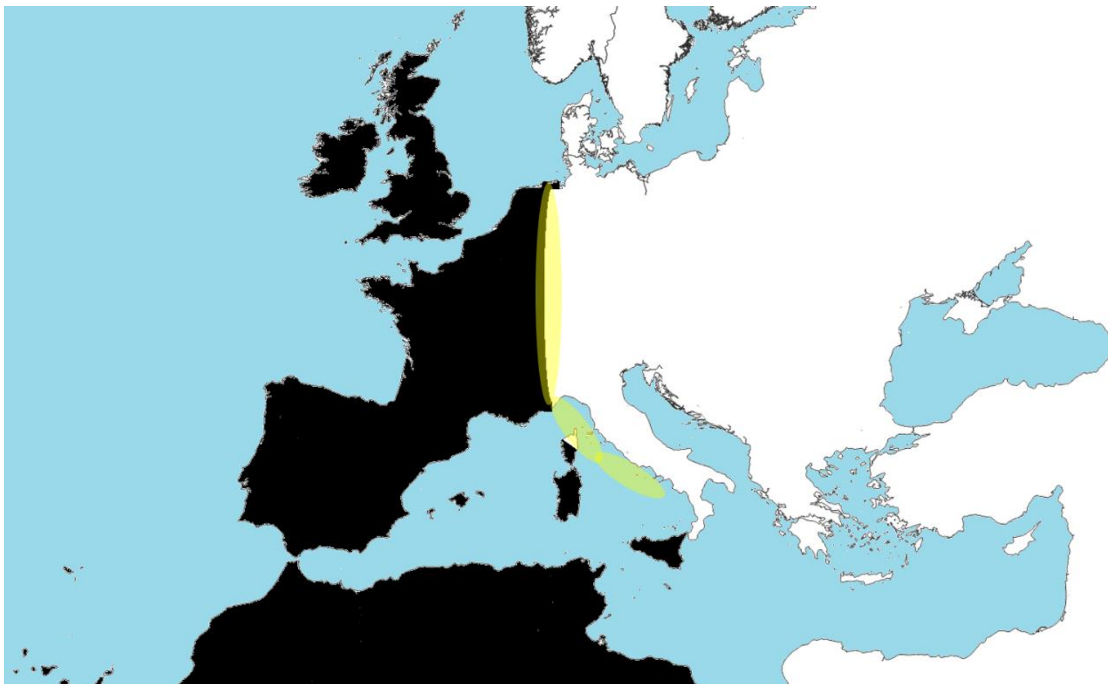


Figure 1.1 Current distributions of the two genetic lineages of *M. jurtina* with the western lineage (*M. jurtina jurtina*) in black and the eastern lineage (*M. jurtina janira*) in white. Yellow indicates areas of hybridisation where the two lineages overlap as shown in (Vodă 2015).

Further allozyme work by Schmitt *et al.* (2005) supported these findings and determined that the two lineages likely diverged at the beginning of the last glacial maximum, around 40,000 years ago, via the occupation of two separate glacial refugia (Fig. 1.2).

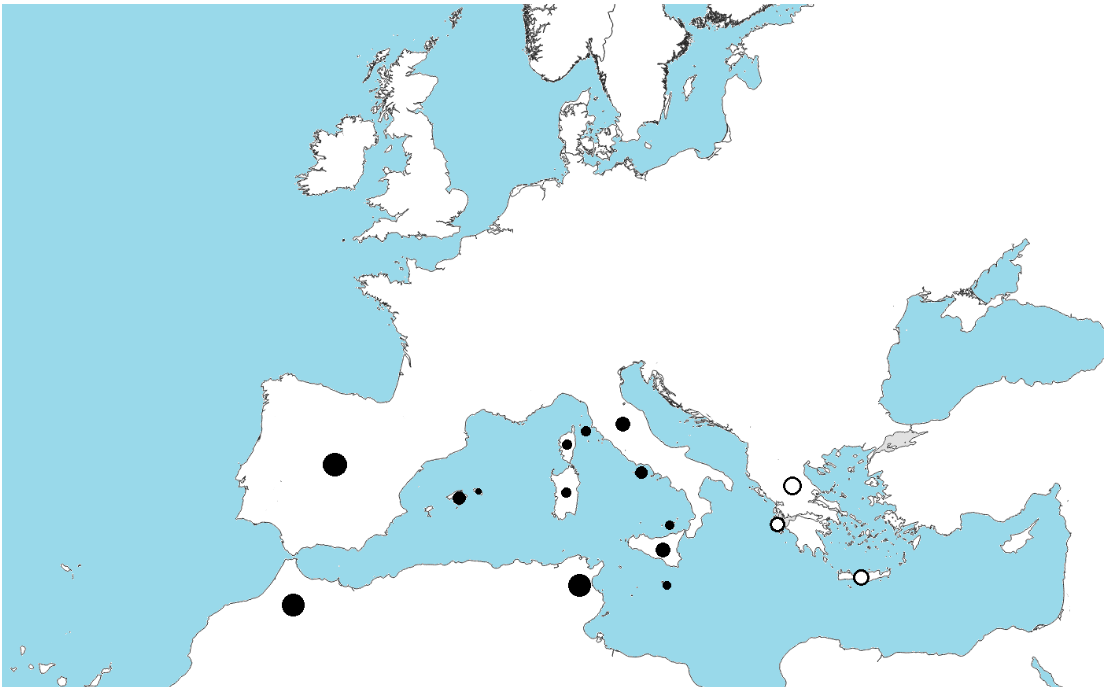


Figure 1.2 Hypothesised glacial distribution of *M. jurtina* reproduced from (Dapporto *et al.* 2011). Black dots indicate the western lineage *M. jurtina jurtina,* white dots indicate the eastern lineage *M. jurtina janira*.

A subsequent allozyme study by Habel *et al.* (2009) found three distinct genetic groupings across Europe, in contrast to the two suggested previously: a western European grouping, a central/eastern European grouping and an Italian/Maghreb grouping. Furthermore, they found that the Moroccan populations studied were more genetically similar to the Italian lineage, despite being geographically closer to the western European one. These results led to the conclusion that Africa was in fact colonised by the Italian lineage of *M. jurtina* (Habel *et al.* 2009)*,* contrary to Tauber (1970).

Using morphometrics Dapporto *et al.* (2009) contradict Habel *et al.* (2009). Again, their study only supported the two lineages of *M. jurtina*. They also found that populations on western Italian islands showed genital morphologies most similar to the western lineage, despite mainland populations belonging to the eastern lineage. Their findings led to the conclusion that Mediterranean islands also acted as a glacial refuge as well as the mainland peninsulas (Dapporto *et al.* 2009).

Dapporto *et al.* (2011) then attempted to clarify the situation, by using both allozyme and morphometric data. They determined that recent, postglacial gene flow had resulted in the discrepancies between the morphometric and allozyme data. This was attributed to postglacial range contractions and expansions (Dapporto *et al.* 2011). This work was continued by Dapporto & Bruschini (2012), who suggested that the entire Mediterranean was originally occupied by the

western lineage of *M. jurtina* and subsequently colonised by the eastern lineage from the Balkans (Fig. 1.3).

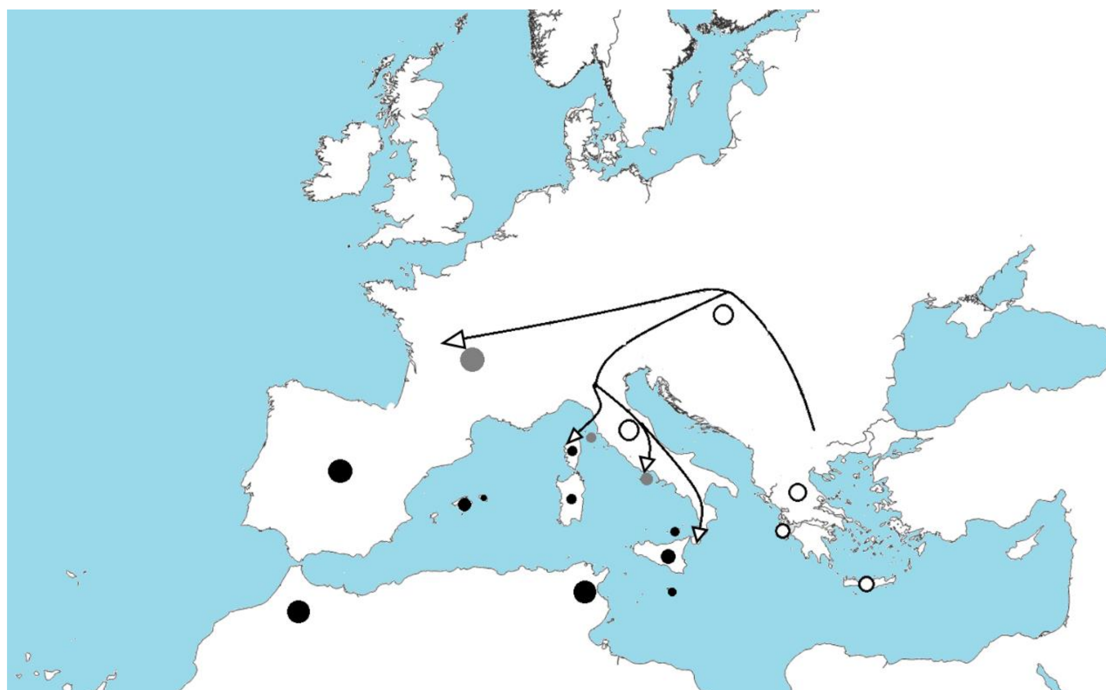


Figure 1.3 Hypothesised recolonization and invasions of *M. jurtina* after the last glacial maximum, reproduced from (Dapporto *et al.* 2011). Black dots indicate the western lineage *M. jurtina jurtina,* white dots indicate the eastern lineage *M. jurtina janira*. Grey dots indicate areas of hybridisation between the two lineages. Arrows show direction of recolonization by the eastern lineage.

A recent study by Kreuzinger *et al.* (2015) using both nuclear and mitochondrial genetic markers does not support the above findings. In their study the authors investigated all seven species within the *Maniola* genus and determined them all to belong to a single, highly variable "super species" *Maniola jurtina.* They also found only moderate genetic differentiation between two lineages of *M. jurtina* (when including all seven *Maniola* species) and refer back to the original "out of Africa" hypothesis suggested by Tauber (1970) almost fifty years ago.

Clearly further research is required to fully determine the phylogeography of *M. jurtina*. However, despite no clear-cut conclusions these studies do provide valuable information regarding the contemporary genetic diversity of the species across Europe.

1.7.4 *Maniola jurtina* landscape genetics studies and microsatellite isolation

M. jurtina use systematic search methods and looping flight when searching for suitable habitat (Conradt *et al.* 2000). As a result of modern farming practices landscapes have become increasingly fragmented, resulting in declines in *M. jurtina* (Delattre *et al.* 2010). Therefore, the ability of individuals to disperse across these fragmented landscapes is of increasing interest.

Previous studies into the landscape genetics of *M. jurtina* have been carried out using allozymes (Goulson 1993a; Wood & Pullin 2002; Schmitt *et al.* 2005; Habel *et al.* 2009). All of these

studies found high levels of genetic diversity but only weak levels of genetic differentiation between populations. These levels of variation were considered higher than expected for a sedentary butterfly (Goulson, 1993a), however Schmitt *et al.*, (2005) ascribed the results to a high number of individuals with a great ability to adapt to environmental change. The levels of heterozygosity found in their study were more than double those found in Goulson (1993a) using some of the same loci, however the different lab techniques employed mean the two studies cannot be directly compared. These results were further supported by Habel *et al.* (2009) who also found high levels of genetic diversity and low differentiation, using some of the loci used in (Schmitt *et al.* 2005).

The next major investigation into the landscape genetics of *M. jurtina* occurred with the isolation and development of fifteen microsatellites (Table 1.2) (Richard *et al.,* 2015). DNA from six male and six female individuals was used and enriched for microsatellites. In total 646 microsatellite containing regions were found, with 374 suitable for primer design. Fifteen were selected. These were then tested on 96 individuals from six populations across France (Richard *et al.* 2015). One locus was found to be sex-linked and discarded. Although moderate frequencies of null alleles were found among seven loci, all fourteen remaining loci were used to estimate pairwise F_{ST} values. Pairwise F_{ST} values were all below 0.02, similar to those reported by Goulson (1993a) and Wood & Pullin (2002), suggesting that isolation by distance was negligible at the study's scale (~600km) (Richard *et al.* 2015). The authors suggest that the low F_{ST} values reported in this and previous studies could be due to either previously unreported long distance dispersal events, or stepping stones between populations allowing gene flow (Richard *et al.* 2015), with contradicting evidence supporting both options (Schneider *et al.* 2003; Delattre *et al.* 2010).

The microsatellites isolated in Richard *et al.*, (2015) were used by Villemey *et al.* (2016) to determine whether certain landscape features have an effect on *M. jurtina* gene flow. The authors sampled a total of 1681 samples across 18-30 locations within three regions in France, over a total distance of 600km. Using both circuit theory and least cost paths, they determined that grasslands and linear grass elements enhance gene flow, whilst arable landscapes and forests limit gene flow (Villemey *et al.* 2016). Their analysis only explained five percent of the genetic variation observed, suggesting that either certain unmeasured landscape elements may have an effect on gene flow or there are other, non-landscape, factors causing larger changes in gene flow (Villemey *et al.* 2016). As with previous studies they also found high levels of gene flow across regions, with low F_{ST} values. They agreed with Schmitt *et al.*, (2005) in that the high abundance of *M. jurtina* is linked to the low levels of genetic differentiation.

Table 1.2 Fifteen microsatellite loci isolated from *M. jurtina.* Reproduced from (Richard *et al.* 2015). * sex-linked locus, # high null allele frequency.

Locus	Primer sequence (5'-3')	Repeat motif
Mj0008	F: PET-CGTGTCGCCTAAACCACATC R: TGGCAACCCTAAACCCTACG	$(ACAT)^7$
Mj3956	F: PET-CAACATCGGGAGTCGAAACG R: CTCAGCCAGGATACCCACTC	$(GATA)^7$

Locus	Primer sequence (5'-3')	Repeat motif
Mj5331	F: PET-TTAGACCGTGATCCCACTGC R: ATTTCGATAGGCAACGAGGC	$(TATC)^{10}$
Mj5287	F: 6FAM-GCTAGCTCGTGGGTACTCTG R: CTCCAAGCAATAAGACCGCC	$(GATA)^{11}$
Mj7232	F: 6FAM-AAGTTACAAGAGCGTTGGCG R: GCGGGAACTCTTGGGTTTTC	$(CTGT)^{7}$
Mj4870	F: 6FAM-ATGATCCATAGCTGCGTTGC R: CTCCTTAGCGCTTACACGTC	$(ATGT)^{7}$
Mj7132	F: NED-ATCTGCGGATTTGCAGTTGG R: CACTATTGAGCACGTGTGTCC	$(TATG)^{13}$
Mj5522	F: NED-TGATCTTTGCCAGCAGGAAC R: AGTGTAAGCTGGCCCTAAAC	$(GATA)^{8}$
Mj3637	F: NED-CTTCCGCAAAATAACGTCTGC R: AGATACTCCATTGACCCGGC	$(TCTA)^{7}$
Mj5647	F: PET-GCGTTCTGATTACCACCCTG R: GCGACAGTCCCCTAAGATCG	$(TATG)^{13}$
Mj0247	F: PET-ATTCCACAAACGAGCCAACG R: ACTCCGATGGTAAGAGGTGC	$(GATG)^{8}$
Mj2410*	F: PET-TAATTAGAGTTTGCGCGGGG R: CGCACACCGCAGTATAAGTG	$(TGTA)^{7}$
Mj5563	F: VIC-CGGTTTTGCCGATAGCGTAG R: CGCAAGGCAATAGACCACTC	$(ATCT)^{7}$
Mj0272#	F: VIC-GTTGCATTGGCACACTCCTC R: CAGCTGCACACTACGACAAG	$(AGAT)^{7}$
Mj0283	F: VIC-CCCTTAGAATAAGAACTCGGCTC R: TGTTCGCACATGCTTAGTCC	$(AGAT)^{9}$

The most recent investigation into the genetics of *M. jurtina* was by Baxter *et al.* (2017). In this study DNA from 196 individuals was collected from samples across five islands on the Isles of Scilly. Twelve AFLP primers were used, resulting in 176 polymorphic bands. They then used a permutation test to determine that significant levels of genetic structure were present between islands, but found no evidence of isolation by distance (Baxter *et al.* 2017). They suggested that these results confirm the assumptions Dowdeswell *et al.* (1949) made, based upon their mark-release-recapture studies, i.e. that little migration occurs between islands.

All previous studies into the genetic diversity of *M. jurtina* show high levels of genetic diversity within populations. Below the level of postglacial lineages (Schmitt *et al.* 2005; Habel *et al.* 2009) all studies show little genetic differentiation between populations. Baxter *et al.* (2017) are the exception to this, finding distinct population structure between islands. These results are very interesting not only because of the apparent structure that they show, but also because of the

required assumptions regarding open water barriers (Baxter *et al.* 2017), which are in direct contrast to observations by Dennis & Shreeve (1996).

Despite being one of the best studied butterflies of the past century it is clear that there are still many unanswered questions about the biology of *M. jurtina* to be investigated. In this study we will investigate a number of aspects of *M. jurtina* biology using combinations of population, experimental and genetic data.

1.8 Overall project aims

Biological monitoring schemes are an extremely valuable tool in nature conservation. Whilst the current range of species and biological aspects covered is already sizable, as is the number of questions that monitoring data can be used to answer, there will always be room for more monitoring and for a wider range of applications of recording data. This thesis will focus on a new use for monitoring data, as well as describe a new biological monitoring scheme for genetic diversity.

The application of monitoring data to predict the resilience of ecosystem services has the potential to save huge amounts of time and effort, by reducing the need to collect specific trait-based datasets. As reductions to biodiversity put increasing pressure on ecosystem services, the use of already available data to determine the resilience of specific services offers a new, more efficient solution to the problem.

Currently there is a severe lack of monitoring data on the genetic diversity of many plant and animal species. This recording black hole is most severe in regard to wild species, with no direct socio-economic value. This is in spite of the growing recognition of the importance of genetic diversity to maintaining biodiversity and ecosystem services. A major part of this thesis will cover the work in developing a pilot genetic monitoring scheme for *M. jurtina.* In theory the methods used should be applicable to a range of other species, increasing the possibilities for other genetic monitoring schemes.

The follow topics will be addressed, each forming a separate chapter that will help to address gaps in biodiversity monitoring highlighted at the start of the introduction, i.e. novel applications of monitoring data to existing problems, a lack of monitoring data for genetic diversity, and the combination of monitoring data with experimental and genetic data to investigate species biology.

Chapter 2 – Predicting resilience of ecosystem functioning from co-varying species' responses to environmental change

In this chapter I use long-term monitoring data to determine the similarities of species' population dynamics via correlations. Using a hierarchical clustering algorithm, I then create a dendrogram showing which species respond most similarly to environmental change. I then create proxies for ecosystem services and using Mantel tests investigate whether they are spread evenly across response guilds, or clustered into specific groups of species.

Chapter 3 – Monitoring the genetic diversity of *Maniola jurtina*

In this chapter I use microsatellite markers to measure levels of genetic diversity and divergence across fifteen populations of *M. jurtina* in southern England over time. The dataset used comprises eight continuous years of samples.

The following hypotheses are tested in this chapter:

1) *High levels of gene flow are present in M. jurtina populations across the south of England.*
2) *Levels of genetic diversity among M. jurtina populations in the south of England do not change over time.*

Chapter 4 – The genetic diversity of *Maniola jurtina* across Europe

In this chapter I use microsatellite markers to measure the genetic diversity of *M. jurtina* across Europe. Using a total of 810 samples from eleven countries spanning 2525 kilometres I investigate whether population structuring occurs at the continental scale.

The following hypotheses are tested in this chapter:

1) *Greater levels of genetic differentiation are present among populations of M. jurtina across Europe than compared to just the UK.*
2) *M. jurtina populations across Europe exhibit significant isolation by distance effects.*

Chapter 5 – The influence of chalk grasslands on the phenology and ecology of *Maniola jurtina* in the UK

In this chapter I use a combination of long-term monitoring, genetic and experimental data to investigate the effects of geology and topology on *M. jurtina* phenology. Using a number of statistical approaches including linear, general linear and mixed effects modelling I determine to what extent flight periods are protracted on chalk sites and whether these are associated with any levels of population structuring or drought tolerance. Overall, I test the following:

1) To what extent are M. jurtina population flight periods protracted on chalk grasslands in the UK.?

2) Are populations of M. jurtina clustered into genetically structured populations based upon the same habitat conditions?

3) Are populations of M. jurtina on chalk grasslands more drought tolerant than populations in other habitats?

Chapter 2. Predicting resilience of ecosystem functioning from co-varying species' responses to environmental change

Published as: Greenwell MP, Brereton T, Day JC, Roy DB, Oliver TH. Predicting resilience of ecosystem functioning from co-varying species' responses to environmental change. Ecol Evol. 2019; 00:1–16. https ://doi.org/10.1002/ece3.5679

2.1 Abstract

Understanding how environmental change affects ecosystem function delivery is of primary importance for fundamental and applied ecology. Current approaches focus on single environmental driver effects on communities, mediated by individual response traits. Data limitations present constraints in scaling up this approach to predict the impacts of multivariate environmental change on ecosystem functioning.

We present a more holistic approach to determine ecosystem function resilience, using long-term monitoring data to analyse the aggregate impact of multiple historic environmental drivers on species' population dynamics. By assessing covariation in population dynamics between pairs of species, we identify which species respond most synchronously to environmental change and allocate species into 'response guilds'. We then use 'production functions' combining trait data to estimate the relative roles of species to ecosystem functions. We quantify the correlation between response guilds and production functions, assessing the resilience of ecosystem functioning to environmental change, with asynchronous dynamics of species in the same functional guild expected to lead to more stable ecosystem functioning.

Testing this method using data for butterflies collected over four decades in the United Kingdom, we find three ecosystem functions (resource provisioning, wildflower pollination, and aesthetic cultural value) appear relatively robust, with functionally important species dispersed across response guilds, suggesting more stable ecosystem functioning. Additionally, by relating genetic distances to response guilds we assess the heritability of responses to environmental change. Our results suggest it may be feasible to infer population responses of butterflies to environmental change based on phylogeny - a useful insight for conservation management of rare species with limited population monitoring data.

Our approach holds promise for overcoming the impasse in predicting the responses of ecosystem functions to environmental change. Quantifying co-varying species' responses to multivariate environmental change should enable us to significantly advance our predictions of ecosystem function resilience and enable proactive ecosystem management.

2.2 Introduction

Ecological systems are essential to human society for many reasons, including the provision of ecosystem functions and services (Díaz *et al.* 2013). These services include regulation of climate, prevention of flooding, provision of resources and cultural well-being (Costanza *et al.* 1997). A rapidly rising global population is leading to an increasing demand for ecosystem services (Biggs *et al.* 2012), however consequent anthropogenic drivers degrading ecosystems means that their ability to deliver these services is increasingly at risk (Millennium Ecosystem Assessment 2005; UK National Ecosystem Assesment 2011). A key factor in the maintenance of ecosystem functions and services is biodiversity (Hector & Bagchi 2007; Isbell *et al.* 2011; Cardinale *et al.* 2012; Harrison *et al.* 2014; Lefcheck *et al.* 2015). Human activities, including habitat fragmentation, pollution and climate change have led to declines in both species richness and abundance, as well as increasing levels of extinction risk (Pimm *et al.* 2014; Tittensor *et al.* 2014; Newbold *et al.* 2015).

Understanding how ecosystem services will respond to changes in species assemblages is regarded as an urgent priority for informing ecosystem management (Díaz et al., 2013; Oliver et al., 2015; De Palma et al., 2017). Indeed the ability to predict ecological functions from species' traits has been hailed as the 'Holy Grail' of functional ecology (Lavorel & Garnier 2002; Suding & Goldstein 2008; Funk *et al.* 2017). Yet, after decades of research, there is still limited ability to make predictions of multiple environmental drivers on ecosystem functioning for multiple species in real-world situations. Previous attempts to predict the impact of environmental changes on ecosystem functions and services have focused on a 'reductionist' approach, attempting to determine how ecological traits ('response traits') mediate community responses to environmental change, and how altered community composition then leads to changes in ecosystem function delivery (mediated by species' 'effect' traits; Díaz et al. 2013).

Since its introduction into ecological literature by Holling (1973), the use of the term resilience has encompassed a number of different definitions, leading to confusion and no clear consensus within the literature (Walker *et al.* 2004). A key reason for this is that resilience can be split into ecological resilience i.e. the magnitude of disturbance that a system can experience before shifting into a different state, including the ability of a system to maintain its functioning, structure and identity (Berkes *et al.* 2003; Folke *et al.* 2004; Elmqvist *et al.* 2007; Suding *et al.* 2008; Chappin *et al.* 2009; Gunderson & Allen 2010), aspects that are sometimes termed 'resistance' (Donohue *et al.* 2013), and engineering resilience i.e. the time taken for a system to return to equilibrium after a perturbation (Pimm 1984; Holling 1996). Whilst engineering resilience draws from a more classical use of the term outside of ecology, stemming from the etymology of the word (Gunderson & Allen 2010), it should not be considered as the definitive term for resilience in ecology (Walker *et al.* 2004). It should also be noted that resilience, along with constancy and persistence are factors that contribute to the overall stability of an ecosystem (Grimm & Wissel 1997), which also encompasses a number of other factors including robustness and variability (Donohue *et al.* 2013). In this study, we focus specifically on the ability of an ecosystem function to be maintained in the face of environmental perturbations, therefore integrating aspects of resistance and adaptive capacity from Holling's (1973) definition of ecological resilience, and recovery from Pimm's (1984) engineering

resilience definition. Sometimes, the same underlying mechanisms can be responsible for both resistance and recovery, and rapid recovery can appear as resistance depending on the time window of measurement (Oliver et al., 2015). Therefore, using resilience as an umbrella term for resistance and recovery makes good sense, and is increasingly widely used by others (e.g. Kohler et al., 2017; Beller et al., 2019). Specifically, the term resilience hereon refers to 'the degree to which an ecosystem function can resist or recover rapidly from environmental perturbations, thereby maintaining function above a socially acceptable level' (Oliver et al., 2015).

The resilience of any particular ecosystem function to a certain environmental driver is related to the correlation between response and effects traits (Suding *et al.* 2008; Díaz *et al.* 2013; Oliver *et al.* 2015). For example, if all species which are important pollinators of a certain crop are highly susceptible to warmer winters (i.e. positive correlation between response and effects traits) then crop pollination would have a low resilience to that aspect of environmental change. In contrast a lack of correlation would lead to the maximum resilience of the ecosystem function (Larsen *et al.* 2005; Díaz *et al.* 2013).

There are, however, a number of significant limitations with this approach that constrain its applicability. Firstly, the number of species for which accurate trait data are available is severely limited, typically belonging to plant species (e.g. Kattge et al. 2011). Where trait data are available for other taxa they tend to be 'soft traits' such as body size, with tenuous or unknown correlations to environmental change and/or ecosystem functioning. There can also be significant disagreements regarding trait measurements between different datasets for the same species (Middleton-Welling *et al.* 2018). Importantly, even where accurate trait data are available, trait-based analyses cannot always be reliably transferred to different regions (Powney *et al.* 2014) and in many cases the goodness of fit of the relationships between putative response traits and environmental change or between putative effect traits and ecosystem function are too low to be used predictively (Lavorel & Garnier 2002; Luck *et al.* 2012).

In some case, the same trait can be used as both the response and effect trait. For example, body size can be used as a response trait when investigating the effects of agricultural intensification on pollinators and can also be used as an effect trait to predict pollination efficiency (Larsen, Williams, & Kremen, 2005). Here, the ability to predict the effects of agricultural intensification on pollinators depends on two relationships: a regression of agricultural intensification on body size, and a regression of body size on pollination. Unfortunately the goodness of fit for such relationships is often low (Lavorel & Garnier 2002; Luck *et al.* 2012). Furthermore, in the majority of cases, a different effect trait must be used from the response trait meaning an additional relationship between the two traits must be calculated, adding further uncertainty and reducing the predictive power of the models.

The substantial sources of uncertainty severely constrain our ability to predict the delivery of ecosystem functions under any particular aspect of environmental change. It may explain why the few successful demonstrations have been limited to studying plant communities (Lavorel *et al.* 2011), with most focussing on single ecosystem functions (primary regulating services), and only 11% of studies considering more than two ecosystem functions (Hevia *et al.* 2017). Furthermore only 4% of

trait-based approaches consider the simultaneous effects of multiple environmental drivers (Hevia *et al.* 2017), even though we know that drivers such as climate and land use change strongly interact in their impacts on biodiversity (Brook *et al.* 2008; Oliver & Morecroft 2014). We expect the environment to change across multiple variables (e.g. multiple different aspects of climate and land use change), therefore additively combining predictions of the effects of single drivers in order to understand the effects of multiple drivers on general resilience of ecosystem functioning makes the overall uncertainty in these reductionist predictive frameworks untenable.

These problems may explain the apparent impasse in functional ecology whereby attempts to develop a predictive framework using a reductionist 'Holy Grail' approach have been ongoing since the late 1990s (Díaz & Cabido 1997; Lavorel *et al.* 1997), with revisits in the early 2000s (Lavorel & Garnier 2002), and again more recently (Funk *et al.* 2017). After three decades of methodological development with only limited application (e.g. see Gross et al. 2008 and Suding & Goldstein 2008), new methods are urgently needed to predict the resilience of ecosystem functioning under environmental change.

Here, we propose a more holistic approach, utilising long-term population monitoring data that reflect the aggregate effects of multivariate environmental change on species' population dynamics. Using this method, groups of species with similar responses to multiple historic environmental drivers, identified through more synchronous population dynamics, can be allocated into 'response guilds'. The distribution of effects traits across these response guilds can then inform on the resilience of ecosystem functioning.

Changes in population dynamics are due to the interactions between organisms and the combined biotic and abiotic effects of their environments (Wallner 1987). Covariance in the population dynamics of any two species is determined by a number of factors including direct and indirect species interactions (e.g. competition effects), similarity in responses to environmental change (e.g. population responses to weather), and in the fundamental aspects governing population growth (e.g. intrinsic rate of population increase and density dependence; Birch, 1948; Wallner, 1987; Walther et al., 2002; Loreau & de Mazancourt, 2013).

If multiple species perform the same ecosystem function and decline synchronously (e.g. through strong positive correlations between response and effect traits; Suding & Goldstein, 2008) then the overall ecosystem function delivered by the species community is likely to decline, albeit just temporarily. This may lead to levels of functioning falling below some threshold that causes a socially unacceptable deficit in ecosystem services (e.g. yield deficits due to a loss of pollination function). Conversely, asynchronous dynamics of species in the same functional guild are expected to lead to more stable ecosystem functioning and subsequent ecosystem service provision (Ives *et al.* 1999; Yachi & Loreau 1999; Loreau & de Mazancourt 2013).

To explore these risks to ecosystem function, in this study, we map ecosystem functions onto species 'response guilds' identified through analysis of the covariance between species' historical responses to environmental change. We also explore how phylogenetic relationships between species can be related to response guilds (Díaz *et al.* 2013), which will lend additional understanding to species conservation and ecosystem management.

To demonstrate our method we use butterfly time series data. Butterflies are often used as indicators for other taxonomic groups (Thomas 2005). They perform a range of ecosystem functions that underpin supporting, regulating and cultural services and have excellent population time series data available. Three ecosystem functions were selected to demonstrate how this new method can be used to examine the resilience of ecosystem functioning: 1) the provision of food to higher trophic levels, as lepidopteran larvae are a key food source for many bird species during chick development (Visser *et al.* 2006); 2) outcrossing pollination function, comprising the important role that butterflies play in dispersing wildflower pollen over large distances (Courtney *et al.* 1982); 3) aesthetic cultural function, through members of the public experiencing culturally important taxonomic groups, which underpin cultural ecosystem services that support wellbeing (Clark *et al.* 2014).

2.3 Materials and methods

2.3.1 Creating a population dynamics correlation matrix of inter-annual changes in abundance

UK-wide annual abundance indices for 54 UK butterfly species from 1976 to 2014 were available from the UK Butterfly Monitoring Scheme (UKBMS). UKBMS data were collected by volunteers using the 'Pollard walk' method (Pollard & Yates 1993). Collated indices were calculated by the UKBMS in a two-step method. First, site abundance indices were calculated by fitting a Generalised Additive Model to count data from each site, in order to estimate missing data values within a year (Rothery & Roy, 2001; further description can be found in Botham et al., 2013). Second, the site abundance indices were used to calculate national collated indices, as with other European species monitoring schemes (ter Braak *et al.* 1994). This was achieved using a log-linear Poisson regression model to calculate expected counts each year, with a site factor to take into account differences between sites (UKBMS 2016) and a year factor to account for missing years. These national-level abundance time series reflect aggregate changes of UK populations to broad environmental conditions, such as weather effects (e.g. Roy, Rothery, Moss, Pollard, & Thomas, 2001), as well as density dependence (Pollard *et al.* 1987).

Using these national abundance time-series, for each species inter-annual changes were calculated by subtracting the standardised log abundance index from that of the year preceding it, creating a dataset containing the yearly changes in species abundance for all species from 1977 to 2014. Using the base R function *cor* (R Core Team 2016), a population dynamics correlation matrix was created using Pearson's correlation coefficient, for the inter-annual changes in species abundance between each pair of species (Fig. 2.1). Only complete pairs of observations were included in the correlations. The population dynamics correlation matrix was then transformed by multiplying by -1, resulting in the pairs of species with least synchronised population dynamics having positive values (i.e. creating a distance matrix). After this transformation, all values were increased by +1. This was necessary as the methods used to perform a hierarchical cluster analysis do so using Euclidean distances between variables, therefore negative values cannot be included. All future references to the population dynamics correlation matrix refer to this newly transformed

matrix, where a value of zero indicates perfectly positively correlated interannual dynamics between species, a value of 1 indicates no correlation and a value of 2 indicates perfect negative correlation (i.e. opposite dynamics).

A hierarchical cluster analysis was performed using this transformed population dynamics correlation matrix, using the *hclust* function in the program R (R Core Team 2016). Species were grouped sequentially into clusters based upon their similarity until all species were grouped into a single cluster (R Core Team 2016). Response guilds were then defined by plotting a dendrogram and allocating all species on a branch below a threshold into guilds (Fig. 2.1, Table 2.1).

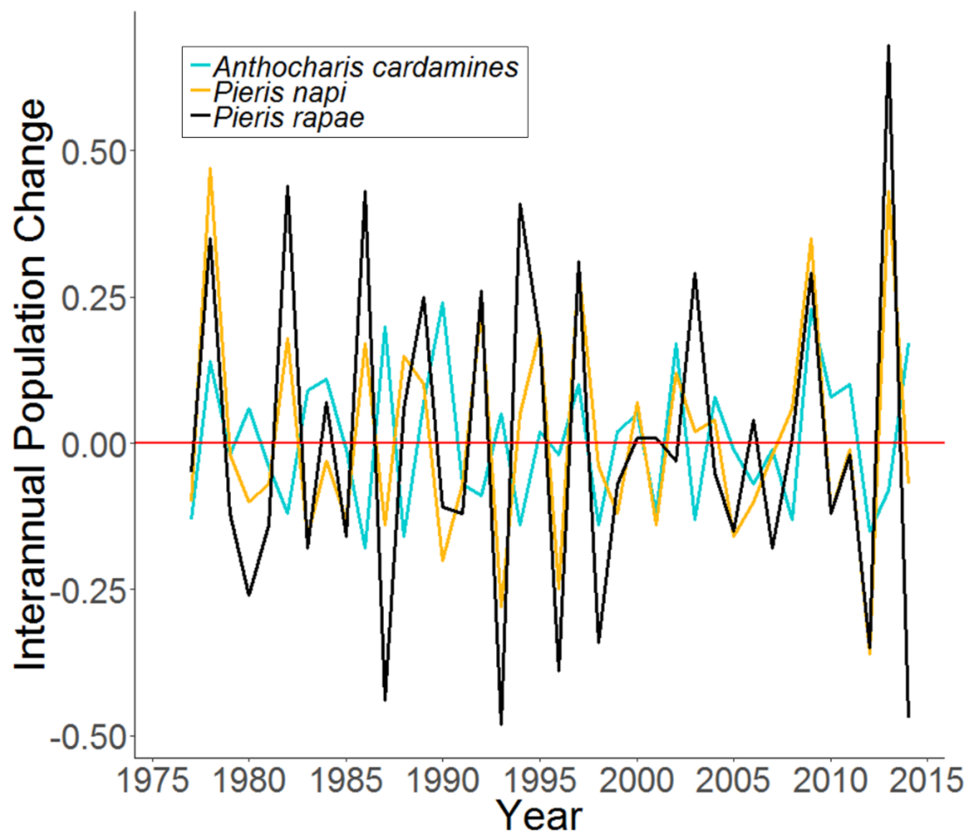


Figure 2.1 Comparison of interannual population changes for three butterfly species. Green-veined white *Pieris napi* and small white *Pieris rapae* have highly correlated population dynamics (Pearson's r = 0.81), indicating they have responded to past environmental change in the same way. Green-veined white *P. napi* and orange tip *Anthocharis cardamines* have much less correlated population dynamics (r = 0.05), indicating they respond differently to changes in the environment; that is, the same environmental drivers have different effects on the overall populations.

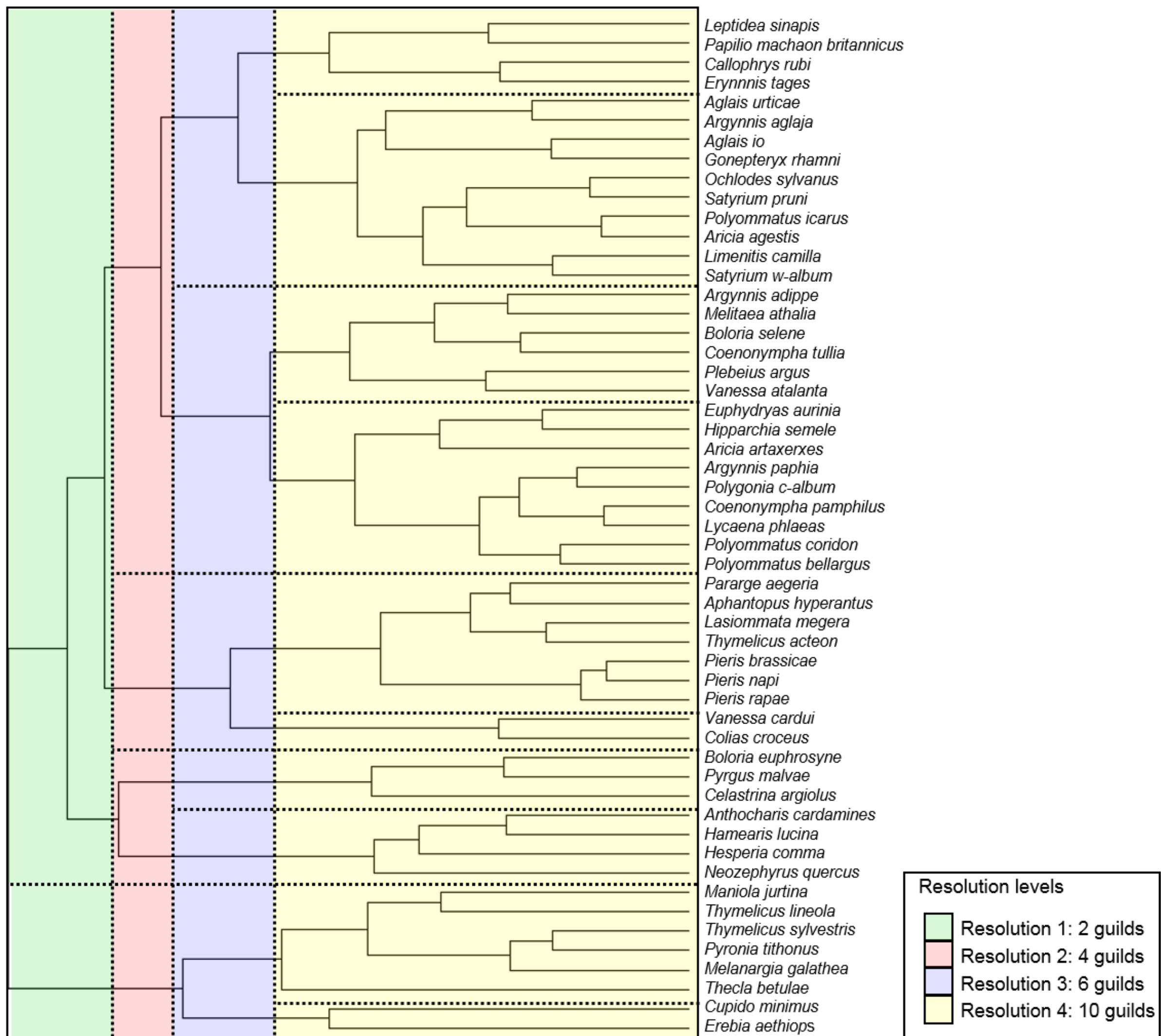


Figure 2.2 Population dynamics dendrogram showing "response guilds," which are groups of species with similar population dynamics. Species with more correlated population dynamics join further to the right-hand side of the dendrogram. Here, four resolutions of response guild are shown (also see Table 2.1), but further grouping is possible.

Table 2.1 Allocation of species into response guilds at different levels of resolution. Different resolutions are achieved by plotting all species onto a dendrogram and selecting species on a branch below a threshold point (see Fig. 2.2). Species with the same number in the table are in the same response guild, meaning they tend to have more similar population dynamics (i.e. have responded to past environmental change in similar ways).

Species	Species allocation into guilds at:			
	Resolution 1	Resolution 2	Resolution 3	Resolution 4
Erebia aethiops	1	1	1	1
Cupido minimus	1	1	1	1
Thecla betulae	1	1	1	2
Melanargia galathea	1	1	1	2
Pyronia tithonus	1	1	1	2

Species	Species allocation into guilds at:			
	Resolution 1	Resolution 2	Resolution 3	Resolution 4
Thymelicus sylvestris	1	1	1	2
Thymelicus lineola	1	1	1	2
Maniola jurtina	1	1	1	2
Neozephyrus quercus	2	2	2	3
Hesperia comma	2	2	2	3
Hamearis lucina	2	2	2	3
Anthocharis cardamines	2	2	2	3
Celastrina argiolus	2	2	3	4
Pyrgus malvae	2	2	3	4
Boloria euphrosyne	2	2	3	4
Colias croceus	2	3	4	5
Vanessa cardui	2	3	4	5
Pieris rapae	2	3	4	6
Pieris napi	2	3	4	6
Pieris brassicae	2	3	4	6
Thymelicus acteon	2	3	4	6
Lasiommata megera	2	3	4	6
Aphantopus hyperantus	2	3	4	6
Pararge aegeria	2	3	4	6
Polyommatus bellargus	2	4	5	7
Polyommatus coridon	2	4	5	7
Lycaena phlaeas	2	4	5	7
Coenonympha pamphilus	2	4	5	7
Polygonia c-album	2	4	5	7
Argynnis paphia	2	4	5	7
Aricia artaxerxes	2	4	5	7
Hipparchia semele	2	4	5	7
Euphydryas aurinia	2	4	5	7
Vanessa atalanta	2	4	5	8
Plebeius argus	2	4	5	8
Coenonympha tullia	2	4	5	8
Boloria selene	2	4	5	8
Melitaea athalia	2	4	5	8
Argynnis adippe	2	4	5	8
Satyrium w-album	2	4	6	9
Limenitis camilla	2	4	6	9
Aricia agestis	2	4	6	9
Polyommatus icarus	2	4	6	9
Satyrium pruni	2	4	6	9
Ochlodes sylvanus	2	4	6	9
Gonepteryx rhamni	2	4	6	9
Aglais io	2	4	6	9
Argynnis aglaja	2	4	6	9
Aglais urticae	2	4	6	9
Erynnnis tages	2	4	6	10
Callophrys rubi	2	4	6	10
Papilio machaon britannicus	2	4	6	10
Leptidea sinapis	2	4	6	10
Carterocephalus palaemon	2	4	6	10

2.3.2 Comparison of inter-annual population dynamics with phylogenetic relationships

In order to determine whether similarities in species population dynamics are related to the genetic relatedness of species (Fig. 2.3) a Mantel test was carried out using a matrix of genetic distances and the population dynamics correlation matrix. Using 1000 possible phylogenies of British butterflies created by Roy *et al.* (2015), for each phylogeny we extracted branch lengths between all pairs of UK butterfly species using the *cophenetic* function from the ape package in R (Paradis *et al.* 2004). Average branch lengths between each pair of species across all trees were then calculated and inputted into a matrix of phylogenetic distances. The phylogenetic and population dynamics correlation matrices were then trimmed to include only species occurring in both (n = 43 species in total). The similarity of the two matrices was determined via a Mantel test with 9999 permutations, using the *mantel* function from the ecodist package in R (Goslee & Urban 2007). P-values were determined by comparing the sum of the distance values between the two matrices to the sums of randomised permutations of the matrices. Under the assumption that if the two matrices are related, the sum of their values will be high and randomisation of the matrices will result in the sums being lower. P-values are calculated by dividing the number of times that the sum of the matrices is higher than the original non-randomised matrices by the number of permutations plus the number of times the sum was higher. Further details can be found in Mantel (1967) and explained in (Diniz-Filho *et al.* 2013).

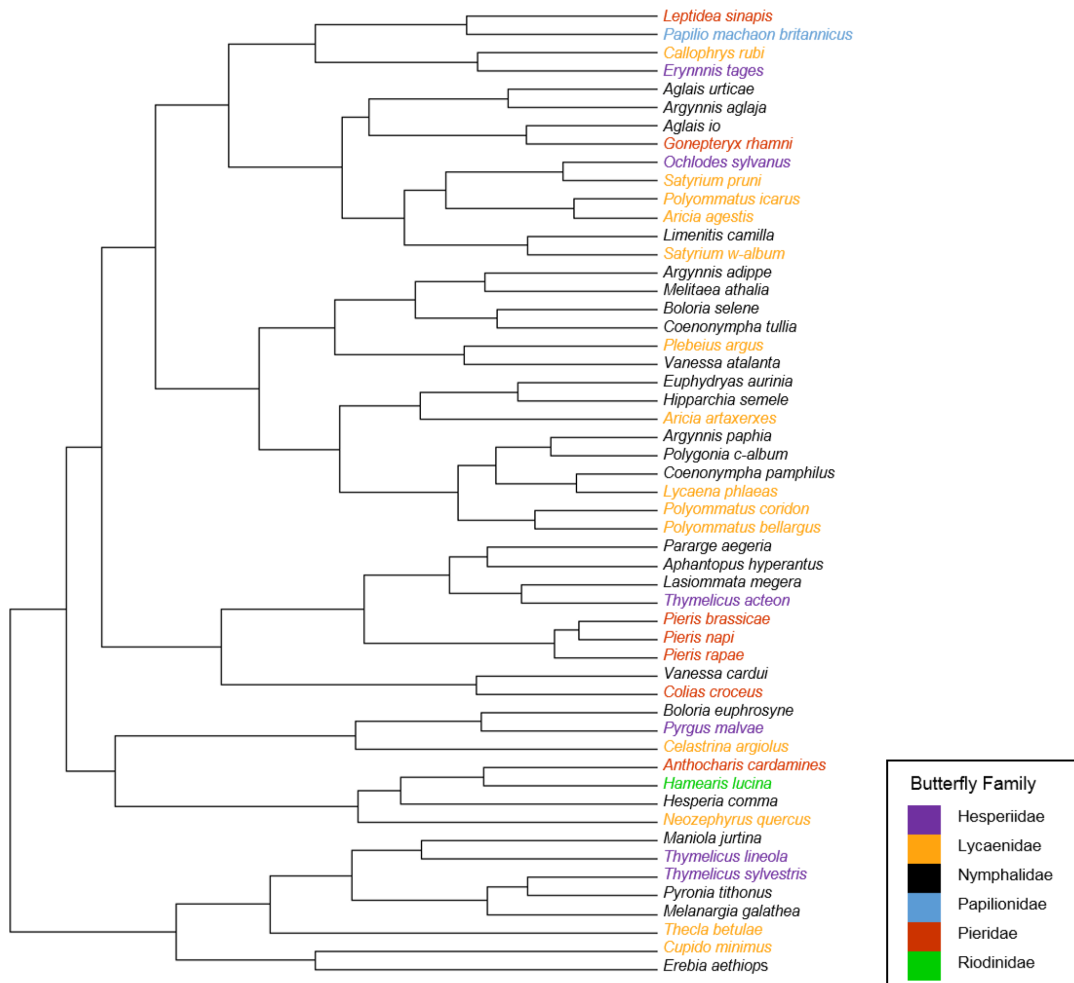


Figure 2.3 Population dynamics dendrogram with butterfly species names coloured by family to show phylogenetic patterning of population dynamics. Species with more correlated population dynamics join further to the right-hand side of the dendrogram.

2.3.3 Calculating proxies of species' roles in ecosystem functioning

We combined ecological theory with published trait datasets to develop new proxies for the relative roles of UK butterfly species in delivering three broad types ecosystem functions: 1) the provision of food to higher trophic levels, 2) wildflower pollination (outcrossing) function, and, 3) aesthetic cultural function. Our basic approach is to develop 'production functions' which combine relevant trait data to estimate relative roles of species in a community in contributing to ecosystem function. Beyond these broad functions we can also calculate several 'sub-functions', (e.g. wildflower pollination function is assessed for different plant families). This approach is an extension of traditional community functional ecology approaches that often use a single trait or functional grouping as a proxy for ecosystem functioning (e.g. Luck et al., 2012; Funk et al., 2017). It allows better incorporation of basic ecological process understanding into our predictions of species' functional roles (e.g. outcrossing pollination can be a function of both insect mobility and plant association). The approach can also be extended further in light of new understanding and available

data (e.g. outcrossing pollination is also likely affected by amount of pollen carried on an insect's body and the likelihood of pollen transfer during flower visitation). Thus, we see our method as a provisional approach towards more nuanced investigation of ecosystem functioning, beginning with the basic production functions below. Standardised trait values for all species can be found in Table 2.2.

Table 2.2 Standardised trait scores for five example traits; larval biomass, cultural function and three levels of pollination outcrossing function. Trait scores scaled between zero and one by dividing all scores by the maximum value for that trait across all species. See main text for data sources.

Species	Biomass Index (B)	Cultural Function Index (C)	General wildflower pollination Index (P)	Brassicaceae pollination Index ($P_{Brassicaceae}$)	Caryophyllaceae pollination Index ($P_{Caryophyllaceae}$)
Aglais io	0.125	0.699	0.116	0.074	0
Aglais urticae	0.121	0.396	0.21	0.138	0
Anthocharis cardamines	<0.001	0	<0.001	>0.001	0
Aphantopus hyperantus	0.25	0.326	0.19	0	0
Argynnis adippe	NA	0	NA	NA	NA
Argynnis aglaja	<0.001	0	<0.001	0	0
Argynnis paphia	0.002	0	0.002	0	0
Aricia agestis	<0.001	0	<0.001	0	0.001
Aricia artaxerxes	<0.001	0	<0.001	0	0
Boloria euphrosyne	<0.001	0	<0.001	0	>0.001
Boloria selene	<0.001	0	<0.001	0	>0.001
Callophrys rubi	<0.001	0	<0.001	0	>0.001
Carterocephalus palaemon	NA	0	NA	NA	NA
Celastrina argiolus	0.002	0.067	0.004	0	0
Coenonympha pamphilus	0.006	0	0.005	0	0.008
Coenonympha tullia	<0.001	0	<0.001	0	0
Colias croceus	<0.001	0	<0.001	0	0
Cupido minimus	<0.001	0	<0.001	0	0
Erebia aethiops	<0.001	0	<0.001	0	0
Erynnnis tages	<0.001	0	<0.001	0	>0.001
Euphydryas aurinia	NA	0	NA	NA	NA
Gonepteryx rhamni	0.005	0.062	0.005	0.003	0
Hamearis lucina	NA	0	NA	NA	NA
Hesperia comma	<0.001	0	<0.001	0	0
Hipparchia semele	<0.001	0	<0.001	0	0
Lasiommata megera	0.001	0	0.001	0	0
Leptidea sinapis	<0.001	0	<0.001	0	0

Species	Biomass Index (B)	Cultural Function Index (C)	General wildflower pollination Index (P)	Brassicaceae pollination Index ($P_{Brassicaceae}$)	Caryophyllaceae pollination Index ($P_{Caryophyllaceae}$)
Limenitis camilla	<0.001	0	NA	0	0
Lycaena phlaeas	0.003	0.059	0.005	0	0
Maniola jurtina	1	0.911	1	0	0
Melanargia galathea	0.009	0.099	0.008	0	0
Melitaea athalia	NA	0	NA	NA	NA
Neozephyrus quercus	<0.001	0	NA	0	>0.001
Ochlodes sylvanus	0.011	0.106	0.008	0	0.010
Papilio machaon britannicus	<0.001	0	<0.001	0	>0.001
Pararge aegeria	0.13	0.177	0.11	0	0
Pieris napi	0.25	0.26	0.35	0.383	0
Pieris brassicae	0.612	0.923	0.627	0.250	0
Pieris rapae	0.561	0.985	0.898	0.561	0
Plebeius argus	<0.001	0	<0.001	>0.001	0
Polygonia c-album	0.031	0.18	0.029	0	0
Polyommatus bellargus	<0.001	0	NA	0	0
Polyommatus coridon	<0.001	0	NA	0	0
Polyommatus icarus	0.017	0.173	0.027	0	0
Pyrgus malvae	NA	0	NA	NA	NA
Pyronia tithonus	0.355	1	0.325	0	0
Satyrium pruni	NA	0	NA	NA	NA
Satyrium w-album	<0.001	0	<0.001	0	0
Thecla betulae	<0.001	0	<0.001	0	0
Thymelicus acteon	<0.001	0	<0.001	0	0
Thymelicus lineola	NA	0	NA	0	0
Thymelicus sylvestris	0.018	0	0.017	0	0
Vanessa atalanta	0.068	0.396	0.081	0	0
Vanessa cardui	0.013	0.071	NA	0	0

2.3.3.1 Provision of food to higher trophic levels

We aimed to create an index of total butterfly larval biomass which reflects the provision of food to higher trophic levels i.e. as a food source for many bird species during chick development (Visser *et al.* 2006). Using updated 10km resolution butterfly occupancy data provided by Butterfly Conservation (Asher *et al.* 2001; Fox *et al.* 2015) and abundance data from the stratified-sampling UK Wider Countryside Butterfly Survey (WCBS), described in Brereton *et al.* (2011), we calculated an estimate for the relative average expected density of individuals across the UK. These relative

national density scores were calculated using Equation 1 below, where D = relative national density of individuals, O = average number of $10km^2$ grid squares across the UK occupied by a species between 2009 to 2017, A = average number of observations for a species between 2009 to 2017 from the WCBS survey, and OA_{max} = maximum O.A score across all species. Thus, the index is standardised to scale between zero and one, with a relative national density of one for the most widely occurring species- the meadow brown *Maniola jurtina.*

$$D = (O.A) / OA_{max} \quad [1]$$

This index of relative national density was then combined with larval length data (L; in mm) described in Carter and Hargreaves (1986), to estimate the relative total butterfly biomass across the UK, under the assumptions that a) larval length is proportionally related to larval biomass with a constant scaling factor, and b) species with high adult abundances also have a high larval abundances and, therefore, provide more food biomass to higher trophic levels. Using Equation 2 below, a relative larval biomass score for each species was calculated, where B = total larval biomass index and DL_{max} = highest D.L score of all species of all species (*M. jurtina*).

$$B = D.L / DL_{max} \quad [2]$$

2.3.3.2 Wildflower pollination (outcrossing) function

Pollination by butterfly species is an important source of outcrossing and maintenance of the genetic diversity of wild flowers, as many species travel further distances than other pollinators (Courtney *et al.* 1982). The relative national density (D), combined with species' mobility scores, was used as a proxy for wildflower outcrossing pollination function (P), under the assumption that species with a greater number of individuals, and higher levels of movement provide a greater function. Mobility indices (M) were taken from Cowley et al. (2001). To standardise the index between zero and one, all values were divided by the highest D.M. score (DM_{max}).

$$P = (D.M) / DM_{max} \quad [3a]$$

Additionally we estimated pollination function for each plant family individually (P_x), where X = 1 if a butterfly species visited the plant family or X = 0 if the species did not (data from Dennis, 2010; Equation 3b below). To standardise the index between zero and one, the denominator DMX_{max} reflects the maximum D.M.X score across all butterfly species for any given plant family X.

$$P_x = (D.M.X) / DMX_{max} \quad [3b]$$

For this case study we present results for two plant families, Brassicaceae and Caryophyllaceae, chosen because each are visited by similar numbers of butterfly species (eight and nine species

respectively; Dennis, 2010), which are clustered differently across the population dynamics dendrogram (Fig. 2.4).

2.3.3.3 Aesthetic cultural function

Butterflies are a culturally important taxonomic group, constituting a major part of the general public's engagement with nature (Clark *et al.* 2014). By determining which species the general public have the highest awareness of, it is possible to estimate the level to which people may notice declines in species. For butterflies, large amounts of data are collected by skilled volunteers on UKBMS sites or WCBS squares across the wider countryside. Unlike UKBMS or WCBS transects, the Big Butterfly Count (BBC) encourages data collection by members of the general public in short 15 minute surveys over a one month period in summer (Dennis, Morgan, Brereton, Roy, & Fox, 2017). As a result, the survey is a better measure of which species members of the public see most often in their local environment. Using published results from the BBC described in Dennis et al. (2017), the mean average number of recordings for the 18 most recorded UK butterfly species between 2011 and 2017 were calculated. Relative cultural function scores were calculated using Equation 4, where C = relative cultural function score, Y = individual species average score from the BBC survey, and Y_{max} = highest species average BBC score (gatekeeper *Pyronia tithonus*). Species that did not occur in the top 18 species in the BBC had negligible occurrence in local environments and were given a score of zero.

$$C = Y / Y_{max} \qquad [4]$$

2.3.3.4 Associations between ecosystem function proxies and species' response guilds

Species' scores for their relative role in providing different ecosystem functions were mapped onto the population dynamics dendrogram, showing which species provided the highest levels of functioning and where they clustered (Figs. 2.4 & 2.5). In order to determine whether functionally important species were distributed non-randomly across the population dynamics dendrogram, the differences in scaled (unit variance and zero mean) ecosystem function scores between all pairs of UK butterfly species were calculated and absolute values were inputted into a matrix of Euclidean distance. Each ecosystem function score matrix then underwent a Mantel test, as described previously, with the transformed population dynamics correlation matrix to determine whether the two showed significant associations.

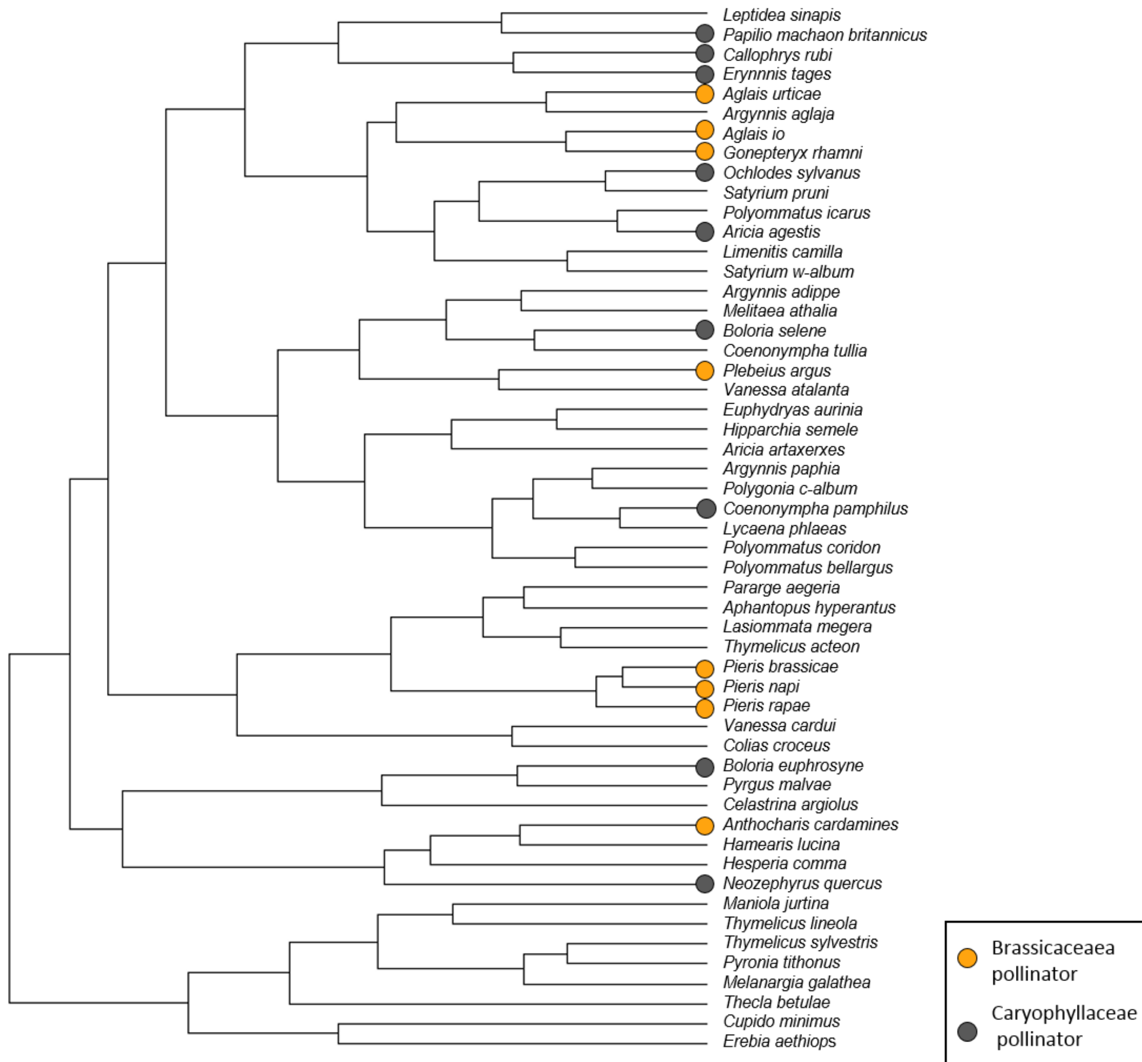


Figure 2.4 Standardized Brassicaceae and Caryophyllaceae pollination scores (*Px*) mapped onto the population dynamics dendrogram. Species proposed to provide a higher level of outcrossing pollination function for Brassicaceae and Caryophyllaceae are indicated by circles.

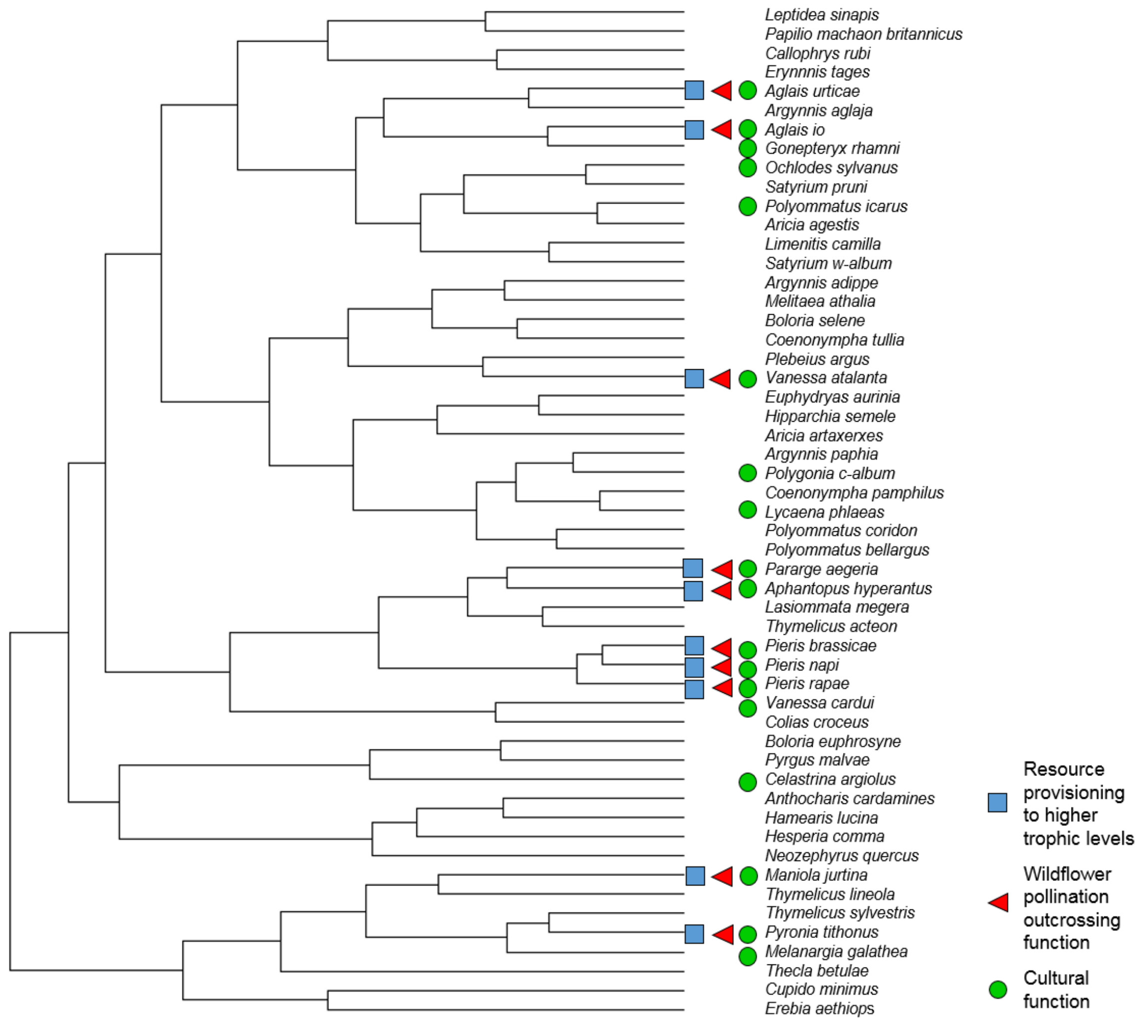


Figure 2.5 Resource provisioning to higher trophic levels, general wildflower outcrossing pollination, and cultural function scores mapped onto the population dynamics dendrogram. For resource provisioning and pollination, the ten species with the highest index scores have been mapped and are indicated by coloured squares and triangles, respectively. For cultural functioning, all species with a score greater than zero have been mapped and are indicated by green circles.

2.4 Results

2.4.1 Comparison of inter-annual population dynamics with phylogenetic relatedness

The results of the Mantel test show that increasing values in the transformed population dynamics correlation matrix are significantly positively associated with increasing genetic distances between species ($p < 0.05$, Table 2.3). Therefore, the greater the genetic distance between two species, the greater the difference in their population dynamics, suggesting that closely related species respond more similarly to environmental change than more distantly related species ($r = 0.151$; Table. 2.3), i.e. in UK butterflies we find there to be sigificant heritability in species' population dynamics.

Table 2.3 Mantel test results relating differences in butterfly population dynamics, genetic distances matrix and all trait matrices.

Matrix 1	Matrix 2	Observed correlation (Mantel r)	Significance (Simulated p-value)	Lower confidence limit (2.5%)	Upper confidence limit (97.5%)
Population dynamics	Phylogenetic tree	0.143	0.003	0.100	0.185
Population dynamics	Larval biomass	-0.279	0.868	-0.567	0.089
Population dynamics	Cultural function	0.086	0.141	-0.006	0.157
Population dynamics	General wildflower pollination score	-0.162	0.665	-0.517	0.198
Population dynamics	Brassicaceae pollination score	-0.232	0.663	-0.419	0.000
Population dynamics	Caryophyllaceae pollination score	0.489	0.163	0.000	0.780

2.4.2 Comparing trait distributions with population dynamics

There were no significant associations between the transformed population dynamics correlation matrix and either the larval biomass or cultural function matrices (p = 0.868 and p = 0.141 respectively (Table 2.3)). Additionally, none of the matrices of pollination functioning (general wildflower pollination, Brassicaceae or Caryophyllaceae) showed any significant associations with the population dynamics correlations (p = 0.665, p = 0.663 and p = 0.163 respectively (Table 2.3)). Therefore, functionally important species are not patterned across the dendrogram in a manner significantly different from random for any of the traits investigated, i.e. they are not significantly clustered within response guilds.

2.5 Discussion

The need to predict the effects of environmental change on ecosystem services remains an urgent priority (Díaz *et al.* 2013; Oliver *et al.* 2015; De Palma *et al.* 2017). Previous methods have so far failed to adequately address this priority and a fresh perspective is required to overcome the decades-long impasse (Díaz & Cabido 1997; Lavorel & Garnier 2002; Funk *et al.* 2017). In this paper, we have demonstrated an alternative method that begins to overcome some of the previous constraints, by using long-term monitoring data to inform on overall species' responses to past environmental change (i.e. integrated across multiple aspects of historic environmental change). This eliminates the need to ascertain relationships between individual response and effects traits, and combine these additively in order to understand overall responses to multivariate environmental change and the subsequent effects on function. Using long-term monitoring data, we show that correlations between species' population dynamics can be used to determine whether functionally

important species respond to historic environmental drivers in the same way, which according to theory should inform on the resilience of ecosystem functioning (Lavorel & Garnier 2002; Loreau & de Mazancourt 2013; Oliver *et al.* 2015). Essentially, rather than considering the correlations between individual response and effect traits, we consider the correlation between ecosystem function proxies and 'response guilds', in order to predict ecosystem service resilience.

Applying this approach for three types of ecosystem functions that underpin supporting, regulating and cultural services provided by UK butterflies, we found that provision of food for higher trophic levels, wildflower pollination function and aesthetic cultural function appear relatively resilient to environmental change. These functional traits were spread across a number of response guilds, suggesting uncorrelated or even asynchronous responses of functionally important species, which should lead to more stable ecosystem functioning (Loreau & de Mazancourt 2013; Mori *et al.* 2013) and lower levels of ecosystem function deficit (Allan *et al.* 2011; Oliver *et al.* 2015). The investigation into the stability of wildflower pollination function showed that butterfly species that visit the family Caryophyllaceae showed more clustering into response guilds than those that are important for Brassicaceae pollination, perhaps suggesting a greater resilience of pollination of the latter, although in both cases the overall correlation between ecosystem function and population dynamics matrices was not significant.

We propose that a higher number of functionally important species across multiple response guilds leads to more resilient ecosystem functioning. Therefore any species which is the sole representative of a response guild should be more important for resilience, as these species have asynchronous dynamics compared with others and so will have more influence on the statistical averaging ('portfolio') effect that results in an overall more stable ecosystem function from a community (Ives *et al.* 1999; Tilman 1999; Yachi & Loreau 1999). Using cultural function in UK butterflies as an example, we find that in some cases, multiple functionally important species are aggregated into the same response guild, e.g. *Pieris rapae, Pieris napi, Pieris brassicae, Aphantopus hyperantus,* and *Pararge aegeria* (Fig. 2.5, Table 2.1). In other cases, however, important functional species are isolated in their own response guilds, for example, the holly blue butterfly *Celastrina argiolus* (Fig. 2.5, Table 2.1). We suggest that this species is particularly important because in years when the other species are in synchronised decline, this may be one of the few remaining species apparent in gardens, ensuring at least some butterflies are seen and providing the maintenance of cultural services. Populations of this species appear to respond to an interacting set of drivers related to weather and parasitoids in a unique way (Oliver & Roy 2015).

In our analysis of UK butterflies, we found that population dynamics show some degree of heritability, with species more closely related more likely to respond to environmental drivers in the same way (Fig. 2.3). This fits with the niche conservatism theory proposed by Harvey & Pagel (1991), whereby closely related species are more likely to be ecologically similar (Ackerly 2009). Interestingly, it contrasts with results from Diamond, Frame, Martin, & Buckley (2011) who found little evidence of a phylogenetic signal in UK butterflies phenological responses. Our findings of a phylogenetic patterning in population dynamics suggest there might be potential opportunity for conservationists

to infer how rarer, data-sparse species respond to environmental change based on the responses of related species for which population dynamics data are available.

Although we believe our methodology offers significant advances over previous reductionist approaches for predicting resilience of ecosystem functioning in real world situations, it has several limitations. First, our method is most applicable to species for which long-term monitoring data are available e.g. in the UK this primarily comprises groups such as plants, butterflies, birds, aphids, moths and ground beetles e.g. Morecroft et al. (2009). Other spatially replicated standardised recording schemes, such as for pollinators, are still in their infancy, although should produce usable data for this method in due course (Pocock *et al.* 2015; Hayhow *et al.* 2016). Furthermore, as well as an expansion in population monitoring schemes, there has also been a recent increase in the taxonomic coverage and participation in citizen science distribution recording schemes (Pocock *et al.* 2017a). In some cases, yearly changes in the total number of biological records (georeferenced records of a species presence at a particular time) can be used as a proxy for yearly changes in species' abundance, as shown by Mason et al., (2018). Using such proxies for time series data would open up this method to a far greater range of species and ecosystem functions, greatly increasing its potential implementation.

Second, using our approach to predict resilience of ecosystem functioning in the future requires the assumption that patterns of species' covariance will remain similar over time. This is a reasonable assumption to some degree, since morphological and physiological traits determine responses to environmental change (supported by our result reflecting significant heritability), and such traits can only change relatively slowly through evolution. However, it remains feasible that newly arising environmental drivers of change could affect individual species idiosyncratically, e.g. a newly arriving pathogen which is species-specific. Therefore, some deliberation is needed with regards to the appropriate level of uncertainty when making predictions, as in any ecological forecasting attempt (Oliver & Roy 2015).

Finally, there are still constraints in applying these methods based on the availability of functional 'effect' traits. To demonstrate the applicability of the method, we used three basic proxies for ecosystem functions delivered by butterflies. Uncertainty remains in the appropriateness of these proxies; for example, we assume that all species found in urban gardens have equal cultural value, with total cultural function scaling proportionally with relative butterfly density. However, certain species might be more culturally important than others (e.g. see Hiron, Pärt, Siriwardena, & Whittingham, 2018), and there may be diminishing marginal returns of cultural value with increasing butterfly abundance. Whilst such concerns are not critical in demonstrating the applicability of the method, further refinement of trait selection and calculation will be necessary for using this method for conservation strategies and in predictive frameworks. Nevertheless, our approach needs far less trait specific information than previous reductionist approaches because we bypass the need to assess response traits for every species and for multiple different aspects of environmental change. Finally, in this study, we have not proposed levels of asynchrony in population dynamics below which 'safe' thresholds of ecosystem function resilience are passed, and further work is necessary, incorporating social science research into levels of acceptable environmental risk.

In summary, whilst there remains uncertainty in the links between species traits, population changes and ecosystem function, our method is more practical and feasible than previous reductionist approaches. It uses long-term monitoring data based on co-varying species' responses to multiple aspects of environmental change, and we hope it offers a significant advancement in our ability to predict ecosystem function resilience.

Chapter 3. Landscape genetics of *Maniola jurtina* in Southern England

3.1 Abstract

The effects of habitat loss and fragmentation on biodiversity are well studied. However, whilst measures such as species abundance or diversity are easily quantifiable, understanding the effects of habitat loss on more cryptic measures of biodiversity, such as genetic diversity, are more complicated. Whilst the importance of genetic diversity has been recognised by the Convention of Biological Diversity, attempts at monitoring or improvement have been minimal, with very few cases of genetic monitoring occurring outside of domesticated or socioeconomically important species. Using microsatellite markers, we present the results of a pilot genetic monitoring scheme for the meadow brown butterfly *Maniola jurtina*. By collecting and analysing samples yearly from 15 sites across the south of England we show that all populations sampled appear to belong to a single, large population with high levels of gene flow and genetic diversity, and low levels of genetic differentiation. Our results also show that for the populations studied there has been very little change in genetic diversity over time. These results provide further information regarding the biodiversity of a well-studied species and also show that the monitoring of a wild species' genetic diversity is an achievable aim, and one that could be carried out for many species.

3.2 Introduction

3.2.1 The effects of landscape on species movements and gene flow

The structure and configuration of a landscape has a direct effect on the movement of organisms within it (Sutherland *et al.* 2015a). Landscapes that facilitate the movement of individuals between habitat patches are more connected than those that impede movement (Taylor *et al.* 1993). However, the degree of connectivity of a landscape can be specific to individual species (Watts & Handley 2010), especially when considering the attributes and responses of the species being studied i.e. functional connectivity (Auffret *et al.* 2015).

Habitat loss and the fragmentation of habitats within a landscape can be major drivers of species loss (Pimm *et al.* 2014), however there is debate within the literature about which of these two factors has the greater effect on species declines, and this too, is likely to be species-specific, with sedentary species disproportionately affected by fragmentation (Fahrig 2003, 2017; Hanski 2015; Crooks *et al.* 2017; Fletcher *et al.* 2018; Fahrig *et al.* 2019). Habitat fragmentation, the reduction of areas of continuous habitat into smaller patches surrounded by areas of dissimilar non-habitat (matrix) (Crooks *et al.* 2017), can lead to subsequent reductions in connectivity and the degree to which a landscape facilitates the movement of an organism (Kadoya 2009; Delattre *et al.* 2013b). In turn this can result in a reduction of gene flow between populations (Keyghobadi 2007), potentially raising extinction risks due to factors such as increasing genetic drift and inbreeding depression (Keyghobadi 2007; Frankham 2010; Shirk *et al.* 2010).

Reductions to genetic diversity as a result of habitat fragmentation can be lessened by connecting fragmented habitats via corridors and stepping stones (Hale *et al.* 2001; Baum *et al.* 2004). However there are relatively few recorded cases where gene flow has been re-established after such measures (Frankham 2010). It would be better therefore, to ensure that habitats remain connected and gene flow does not become disrupted between populations. To measure this connectivity often requires monitoring of the genetic diversity of populations, where direct observation of individual movements is unfeasible. By tracking changes in genetic diversity conservationists can identify isolated populations and begin mitigation attempts to improve gene flow. The data required for such hands-on management are lacking for the majority of species impacted by landscape perturbations.

3.2.2 Landscape effects on *Maniola jurtina* ecology

The meadow brown butterfly, *Maniola jurtina* (Linnaeus 1758), is a common satyrine butterfly species, classified as 'Least Concern' under the red list of British butterfly species (Fox *et al.* 2011). However, the species has undergone declines across Europe over the past 20 years, due to changes in agricultural practices and increasing habitat fragmentation (Delattre *et al.* 2010; Van Swaay *et al.* 2013). *M. jurtina* are generalists, feeding on a range of nectar plants and using multiple grass species (Poaceae) as host plants (Lebeau *et al.* 2016). Adult flight movement through a landscape is non-random, with individuals exhibiting 'foray search' and 'direct flight' movements. Additionally individuals are able to recognise boundaries between areas of habitat and non-habitat (Conradt *et al.* 2000, 2003; Delattre *et al.* 2010, 2013b). Typical individuals are relatively sedentary compared to

other butterfly species (see silver studded blue *Plebejus argus* and painted lady *Vanessa cardui* in Thomas & Lewington (2010), moving around an area with a radius generally less than 500m (although longer distance movements of up to 2.1km have been recorded (Schneider *et al.* 2003)).

Since the mid twentieth century, much research has been conducted on *M. jurtina* covering a wide range of biological topics. These include evolutionary genetics (e.g. Creed *et al.*, 1959; Dowdeswell, 1961; Brakefield & van Noordwijkl, 1985; Brakefield & Shreeve, 1992 and references within), ecology and life history (e.g. Brakefield, 1982a, 1982b; Lebeau et al., 2018), dispersal and movement (e.g. Conradt *et al.* 2000; Wood & Pullin 2002; Schneider *et al.* 2003; Ouin *et al.* 2008; Delattre *et al.* 2013; Evans *et al.* 2019, 2020), and historical distributions (e.g. Dapporto *et al.*, 2011 and references within). However, despite this large body of information, relatively few studies specifically investigate its genetic diversity.

Studies that have investigated the genetic diversity of *M. jurtina* have found a wide range of heterozygosity within populations across Europe (see Thomson, 1987; Goulson, 1993; Schmitt *et al.*, 2005; Habel *et al.*, 2009; Richard *et al.*, 2015). However, as suggested by Schmitt *et al.* (2005), these differing estimates are likely the result of differing molecular techniques and therefore not comparable. These techniques include allozyme analysis using starch gel electrophoresis by Thomson, (1987) and Goulson (1993), allozyme analysis using cellulose acetate electrophoresis by Schmitt *et al.*, (2005) and more recently microsatellites by Richard *et al.* (2015).

A recent study by Richard *et al.* (2015) using microsatellite markers and *M. jurtina* populations in France determined that, at the scale of the study, distance had little effect on pairwise differentiation between populations (F_{ST} = -0.008 to 0.016, maximum distance Aquitaine to Lorraine ~ 670km), with low pairwise F_{ST} scores (<0.02) calculated between all pairs of sites (Richard *et al.* 2015). The authors conclude that these results suggest high levels of gene flow could be down to either previously unrecorded long range dispersal, or stepping stones between populations, functionally connecting areas of habitat (Richard *et al.* 2015).

Further investigations using the same microsatellite markers determined that linear grassland elements enhance *M. jurtina* gene flow, whilst woodlands and arable landscapes limit gene flow (Villemey *et al.* 2016). Their findings support the known ecology of *M. jurtina* and its utilisation of a range of grass species as host plants. As well as grasslands being a common habitat type across Europe, an additional reason for the high levels of genetic diversity in *M. jurtina* may be due to its high abundance and widespread distribution (Schmitt *et al.* 2005). Evidence to support this comes from another widespread, abundant European butterfly species (*Pieris napi*), which has also been found to have high levels of genetic diversity and low levels of genetic differentiation (Schmitt & Hewitt 2004). These results were attributed to *P. napi* belonging to large, continuous populations.

The most recent investigation into the genetic diversity of *M. jurtina* used AFLPs on populations of *M. jurtina* on the Isles of Scilly, UK (Baxter *et al.* 2017). The authors found genome wide differentiation and population structuring among five islands, suggesting restricted migration between islands and the suggestion that open water can inhibit *M. jurtina* gene flow. This work supports mark-release-recapture studies conducted by Dowdeswell *et al.,* (1949) which found

restricted dispersal both within and between the same islands used by Baxter *et al.* (2017). However these studies contradict direct observations of sea crossings, recorded in Dennis & Shreeve (1996) and mark-release-recapture work by Shreeve *et al.*, (1995) that show *M. jurtina* are capable of crossing open water in numbers (Dennis & Shreeve 1996; Dapporto *et al.* 2009).

3.2.3 Monitoring genetic diversity

Previous studies specifically investigating the genetic diversity of *M. jurtina* have been for single time points. For example Goulson (1993a) sampled individuals across a single flight season in 1990, whilst Villemey *et al.* (2016) sampled individuals in 2013. Differences and advances in the molecular technologies used mean the results are not directly comparable, therefore it is not possible to infer changes in genetic diversity from multiple independent studies over time. Specific long-term studies are required to monitor genetic diversity, using a consistent, comparable methodology.

Such studies, quantifying the temporal changes in the genetic metrics of a population (Schwartz *et al.* 2007), are rare outside of socio-economically important species (Hutchinson *et al.* 2003; Hoffman & Blouin 2004; Nussey *et al.* 2005; Poulsen *et al.* 2006). This is despite the need for such schemes being increasingly recognised globally (Boettcher *et al.* 2010), and genetic diversity regarded as a key measure of biodiversity (Pereira *et al.* 2013). Monitoring has become a necessity because genetic diversity is the foundation for all other levels of biodiversity (Bruford *et al.* 2017), and populations with high levels of genetic diversity are more able to adapt to environmental change (McGill *et al.* 2015). The Convention on Biological Diversity (CBD) determined that by 2020 we will have developed and implemented methods for maintaining and minimising genetic erosion of cultivated plants, farmed and domesticated animals, and of wild relatives, including other socio-economically and culturally valuable species (Convention on Biological Diversity 2011b). A recent proposed update to this target expanded this aim to maintain and enhance genetic diversity by 2030 and for 90% of species by 2050 (Convention on Biological Diversity 2020c, d). These aims have come under criticism, as the vast majority of species globally, including *M. jurtina*, do not fall into these narrow categories highlighted by the CBD and such knowledge is currently lacking for the vast majority of wild species (Laikre 2010; Laikre *et al.* 2020).

3.2.4 Chapter aims

Previous studies into the genetic diversity of *M. jurtina* have resulted in a range of estimates as to the levels of genetic diversity, heterozygosity and population differentiation of this species (Goulson 1993a; Schmitt *et al.* 2005; Richard *et al.* 2015). In this chapter microsatellite markers were used to determine levels of genetic diversity and population structure for *M. jurtina* populations across southern England. It was expected that the results would support the previous studies finding high levels of gene flow.

This chapter will also investigate whether the genetic diversity of *M. jurtina* is variable over time, comparing the results from fourteen populations before and after a time gap of five years. Yearly comparisons were also undertaken with three populations to see if there were any annual changes in the levels of genetic diversity.

Finally, this chapter assesses the future utility of long-term monitoring for the genetic diversity of British populations of *M. jurtina* and the practicalities and problems associated with it.

3.3 Methods

3.3.1 Sample collection and preparation

Over the eight-year study period a total of 1024 individual *Maniola jurtina* samples were collected from 15 sites in the South of England (Fig. 3.1). Sites encompassed a range of *M. jurtina* habitat types including lowland calcareous grassland, non-chalk grassland, woodland and open heathland. All sites also host a UKBMS transect, a 2-4km route where volunteer observers record butterflies at specific times of day and during specific weather conditions throughout the summer (Pollard & Yates 1993). Further information on sample collection and preparation can be found in Appendix A.

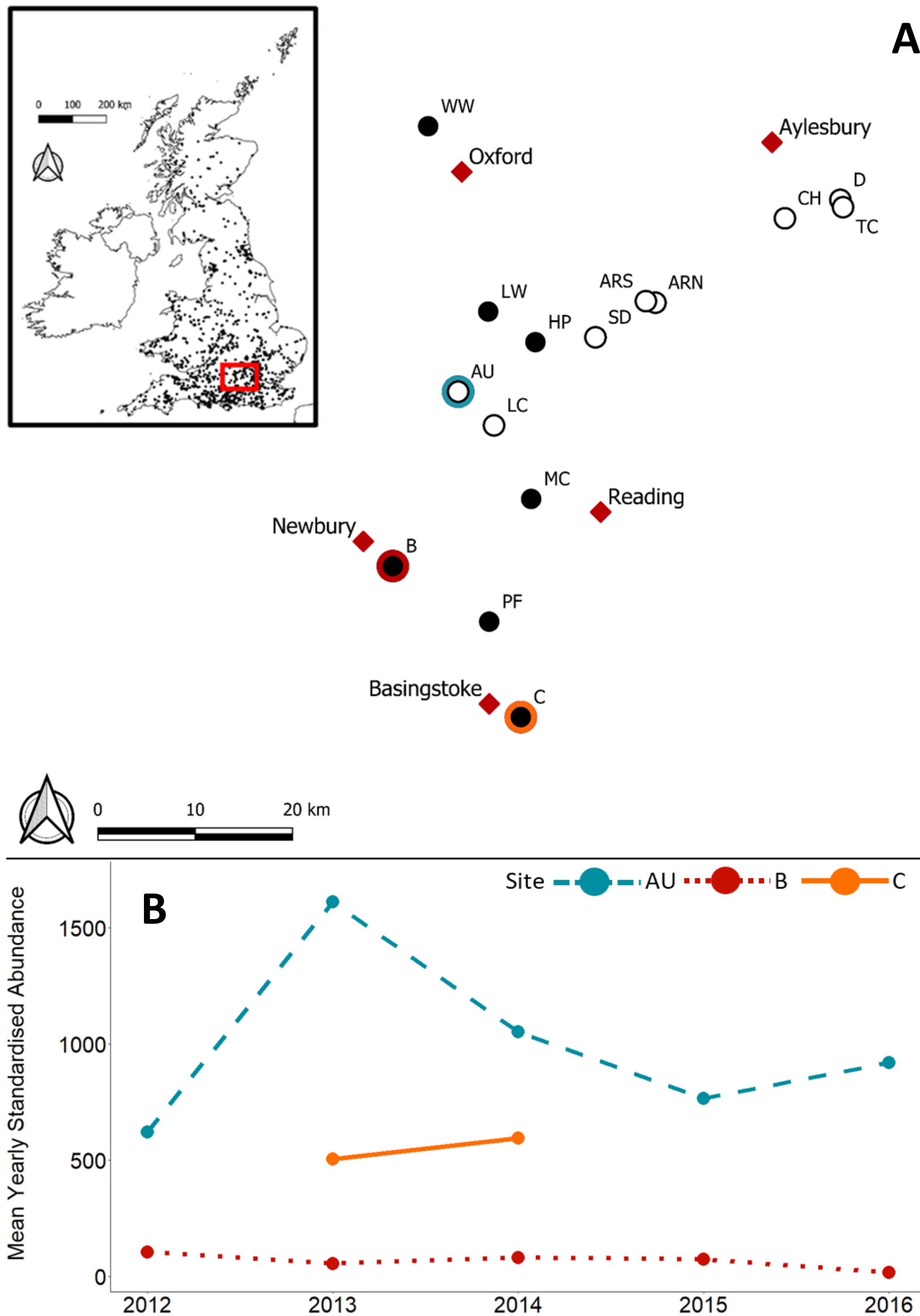


Figure 3.1 A) Fifteen sites around the Chiltern Hills from which *Maniola jurtina* samples were collected for genetic analysis. Large black circles signify non-chalk sites (n = 7), large white circles signify chalk sites (n = 8). Main towns are marked with red diamonds. Inset map shows the locations of 540 UKBMS transect sites from which the 15 used in this analysis are a subsample. Sites with coloured rings are the three sites used for the time series analysis: Blue = Aston Upthorpe (AU), Red = Bowdown (B) and Orange = Crabtree plantation (C). Yearly abundance for these sites shown in plot B. Data for 2012, 2015 & 2016 for C missing due to no recording transects being carried out in those years.

3.3.2 Molecular analysis

All samples were archived as taxonomic vouchers at -20°C at the University of Reading. DNA extractions were carried out using leg tissue. DNA from samples collected in 2012 and 2013 was extracted prior to the project starting by Melanie Gibbs from the UK Centre for Ecology & Hydrology (UKCEH), using DNeasy Blood and Tissue Kits (Qiagen), following the manufacturers guidelines. DNA from samples from Aston Upthorpe (AU), Bowdown (B) and Crabtree plantation (C) from 2014 onwards and samples from all sites in 2017 were extracted using prepGEM Universal (MicroGEM), following the manufacturer's guidelines, optimised by halving the reaction volumes resulting in 20µL eluted DNA. Genotyping was carried out using six microsatellite markers isolated by Richard *et al.* (2015). For further information regarding novel primer development and evaluation, and pre-existing primer evaluation please see Appendices B and C respectively.

Polymerase Chain Reactions (PCRs) were conducted in a total volume of 11µL containing 1µL template DNA, 6.25µL QIAGEN multiplex PCR master mix (3mM $MgCl_2$), 0.625µL tagged forward primer (5' labelled with 6-FAM, NED or PET), 0.625µL reverse primer, 1.25µL QIAGEN Q solution, 2.25µL RNase-free water. Further details of PCRs, including primer concentrations can be found in Appendix C. PCRs were carried out in an Eppendorf Mastercycler nexus eco with an initial denaturation for 15:00 at 95°C, followed by 40 cycles of 00:30 at 94°C, 01:30 at 56°C & 01:00 at 72°C, and a final extension 10:00 at 72°C. All PCR reactions underwent fragment analysis using an Applied Biosystems 3730 DNA Analyser. The results were then scored using GeneMarker® version 1.5 (SoftGenetics) using the standard default settings for animal fragments. Any individual for which there were more than two loci with missing data were removed from the analysis (5.96% of samples). Further samples were then removed to ensure that all populations had a maximum of 5% missing data per locus. This ensured that no loci were dropped in the analysis due to insufficient data.

3.3.3 Statistical analysis

Each set of analyses were undertaken separately for all samples in 2012 (totalling 252 individuals across 14 sites), all samples in 2017 (totalling 287 individuals across 15 sites) and a subset of all samples from three sites (AU, B, C) from 2012 to 2019 (totalling 432 individuals). These datasets are herein referred to as 2012, 2017 and All Years.

3.3.3.1 Site characteristics

Euclidean distances between sites were calculated from the GPS coordinates of each site using the function *distm* in the R package Geosphere (Hijmans 2019). Renkonen's habitat percentage similarity was estimated for each pair of sites (Renkonen 1938; Jost *et al.* 2011), following the methods outlined in Powney *et al.* (2014), Only habitat data within a 500m radius about the centroid of each site was included in the calculation of habitat similarity, using land use classifications from the UKCEH Land Cover Map 2000 (Fuller *et al.* 2002)..

3.3.3.2 Microsatellite analysis

Linkage disequilibrium among pairs of loci was tested for in Genepop v4.7 (Rousset 2008), to determine whether genotypes at each loci were independent from those at other loci. This was carried out via a test of composite linkage disequilibrium as described by Weir (1996). Null allele frequencies i.e. the frequency of alleles for which amplification did not occur, were also calculated. The observed and expected heterozygosities (H_o and H_e respectively) across the whole dataset, for individual loci and averaged across all loci, were calculated using the R package PopGenReport (Adamack & Gruber 2014; Gruber & Adamack 2015), as were deviations from Hardy Weinberg equilibrium (HWE). Significant deviations from HWE indicate population processes such as inbreeding, drift or genetic substructure and were calculated via a Chi^2 test in PopGenReport.

3.3.3.3 Genetic diversity, divergence and structure

For each locus Wright's F statistics (Wright 1965) (defined in Excoffier (2001) as F_{IT}: the correlation between genes within individuals (I) relative to the genes of the total population (T), F_{ST}: the correlation between genes within a subdivision (S) relative to the genes of the total population (T), and F_{IS} (inbreeding coefficient): the correlation between genes within individuals (I) relative to those within a subdivision of the population (S)) were calculated across all sampling locations using Genepop. An estimate of R_{ST}, a measure of differentiation that accounts for allele size variance in genetic markers that undergo stepwise mutation, such as microsatellites (Slatkin 1995), was calculated in FSTAT 2.9.4 (Goudet 1994). Additionally, for each site the allelic richness i.e. the average number of alleles per locus was calculated using the rarefaction procedure, based upon minimum samples sizes of 11, 14 and nine for the three datasets (2012, 2017 and all years) respectively. The number of private alleles, those occurring within a single population, per site were calculated using PopGenReport, as were the expected (H_e) and observed (H_o) heterozygosity for each locus and the percentage differences between H_e and H_o. Expected heterozygosities for each site and across all sites were calculated using Arlequin v 3.5.2.2 (Excoffier & Lischer 2010).

Effective population sizes (N_e) were estimated using two methods, the bias corrected linkage disequilibrium method (Hill 1981; Waples 2006; Waples & Do 2010), and the heterozygote excess method (Pudovkin *et al.* 1996; Zhdanova & Pudovkin 2008) both implemented in NeEstimator V2 (Do *et al.* 2014). The bias corrected linkage disequilibrium method is based upon the random linkage disequilibrium due to chance in each generation of a finite population (Hill 1981) and has a greater precision than alternate, temporal methods to estimate N_e when using microsatellite data (Waples & Do 2010). In contrast the heterozygote excess method indirectly estimates N_e based upon an increase in the observed proportion of heterozygotes away from proportion expected at HWE, within a finite population (Pudovkin *et al.* 1996).

Population structure was investigated using STRUCTURE v.2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2007). Structure uses a Bayesian, model-based clustering method to assign individuals into K number of populations, which are characterized by the allele frequencies per locus. Samples are assigned to populations based upon whether their genotypes are admixed. Separate models are run using Markov Chain Monte Carlo (MCMC) iterations, with increasing values of K up to a

maximum K i.e. the number of sampling locations (Pritchard *et al.* 2000; Falush *et al.* 2003). For all models in this study the parameter set used an admixture model and correlated allele frequencies with a 100,000 burn-in and 1,000,000 MCMC replications per chain. For each possible value of K, twenty chains were run. The most likely value of K within the sample sets was estimated using the programme STRUCTURE Harvester (Earl & VonHoldt 2012), which estimates an *ad hoc* statistic, Delta K, which is used to determine the most likely value of K based upon the rate of change in the log probability of data between successive values of K (Evanno *et al.* 2005). Where K = 1 the use of Delta K is not applicable (Evanno *et al.* 2005), instead the STRUCTURE output bar plots showing the estimated membership coefficients were used to confirm K = 1. To aid visualisation of the STRUCTURE outputs the full results of each run were uploaded to CLUMPAK (Kopelman *et al.* 2015) and run through the main pipeline to identify the optimal cluster label alignments across the different values of K tested.

Contemporary migration rates between sites were estimated using BayesAss v.3.04 (Wilson & Rannala 2003). A burn-in of 10,000 and 10,000,000 MCMC repetitions were used with a thinning interval of 800, and mixing parameters of 0.8 for migration rates, 0.4 for allele frequencies and 0.8 for inbreeding coefficients. MCMC convergence was monitored for each set of samples by plotting the profile of the likelihood and prior values over time using Tracer v1.7.1 (Rambaut *et al.* 2018). For each run the MCMC did not reach convergence. The number of MCMC repetitions was increased to 100,000,000, however convergence still did not occur and as a result BayesAss was abandoned from the study. Instead Slatkin's private allele method for estimating migrants, which estimates average gene flow between populations by measuring the average number of migrants exchanged (Slatkin 1985), was calculated in Genepop. This was estimated across the entire 2012 and 2017 datasets and done on a yearly basis for the three sites across eight years to provide an annual estimate of migration.

For measures of genetic distance, individuals were pooled on a per site basis. Weir and Cockerham's pairwise F_{ST} values were calculated using Fstat v. 2.9.4 (Goudet 1994), which calculates the multilocus estimator of F_{ST} between each pair of populations using a weighted ANOVA (Weir & Cockerham 1984). Estimates of isolation by distance (IBD) were calculated by plotting pairwise F_{ST} values against (log) Euclidean distances between sites, with a Mantel R test using the *mantel* function from the R package Ecodist (Goslee & Urban 2007). Pairwise F_{ST} values were also plotted against the level of habitat similarity between each site pair in a Mantel R test.

In addition to the yearly estimates of migration rates calculated for each year within the All Years dataset described above, changes in pairwise F_{ST} between sites, the effective population sizes and census population sizes were also plotted. Census population sizes were calculated from a relative population abundance score for each site from UKBMS monitoring data.

3.3.3.4 Landscape genetics

The effects of landscape features on *M. jurtina* geneflow were investigated using the R package ResistanceGA (Peterman 2018). ResistanceGA optimizes resistance surfaces, "spatial layers that assign a value to each landscape or environmental feature that represents the degree to which that

feature impedes or facilitates connectivity for an organism of interest" (Spear *et al.* 2010), using a genetic algorithm (GA) from the R package GA (Scrucca 2013). This method removes the subjectivity and bias in the parameterisation of resistance surfaces when using expert opinion or trial and error (Peterman 2018; Peterman *et al.* 2019; Winiarski *et al.* 2020). Detailed methods for ResistanceGA can be found in Peterman (2018).

Initial landscape surfaces were derived from the UKCEH Land Cover Map 25m Raster (LCM) (Rowland *et al.* 2017). Three different parameterisations were considered (Fig. 3.2). Firstly, a binary surface with pixels allocated either as grass (*M. jurtina* habitat) or non-grass (non-habitat). Secondly, each pixel was either defined as non-grass or as one of the four non-heathland grass types classified in the LCM (Improved, Neutral, Calcareous and Acid). Finally, pixels were classified into seven broad habitat classifications: Woodland (Broadleaf and Coniferous woodland), Arable (Arable and Horticulture), Improved Grassland, Semi-Natural Grassland (Neutral, Calcareous and Acid Grassland), Heathland (Heather, Heather Grassland, Inland Rock), Freshwater, and Built-up Areas (Urban and Sub-urban). No other habitat types occurred in the sample landscape. Due to computational and time limitations, initial runs were carried out at a 250m resolution, using the modal 25m pixel value within each 250m pixel. For the binary landscapes 100m resolution surfaces were also generated and tested.

In total five competing models were tested: a null model i.e. geographic structure has no effect on geneflow, an IBD model i.e. distance between populations rather than landscape structure affects geneflow, and the three isolation by resistance (IBR) models: binary grass/non-grass, multiple grasses and multiple variables. Two measures of genetic distance were used in the analysis. F_{ST} and the proportion of different alleles between populations (PODA). PODA was obtained by subtracting the proportion of shared alleles between populations (POSA) from one i.e. PODA = 1 – POSA. POSA was calculated using the *pairwise.propShared* function from PopGenReport,.

Using ResistanceGA an initial random population was generated and assigned resistance values using the parameter values being tested. Pairwise effective distances were then calculated across the landscape using random walk commute times with the *gdist.prep* function from the R package gdistance (van Etten 2017). A linear mixed effects model, with genetic distance as the response variable and scaled and centred pairwise effective distance as the predictor variable, was fitted to the data with a maximum likelihood population effects (MLPE) parameterization to account for non-independence within the pairwise data (Peterman 2018). Log-likelihood values were obtained from the model. After sufficient repetitions to create populations that were 15 times the number of parameters, the GA began selection, with the 5% of individuals with the best log-likelihood values carried on to the next generation. A new population was then created through crossover and mutation. The process was repeated until 40 generations had passed with no improvement to the log-likelihood results (Peterman 2018).

For each resistance surface the process was run twice to confirm convergence and parameter estimates (Peterman 2018). Akaike information criterion (AIC) were used to assess model fits. For each optimized resistance surface, the replicate with the greatest log-likelihood value was used in a bootstrap analysis, to determine the best supported resistance surface using the *Resist.boot* function from ResistanceGA*,* with 75% population resampling and 1000 iterations. The pairwise responses and distance matrices from each optimized surface were subsampled without replacement. Fit statistics were then calculated by refitting the MLPE model, with the frequency that a model was top ranked providing support for the accuracy of optimization (Peterman 2018).

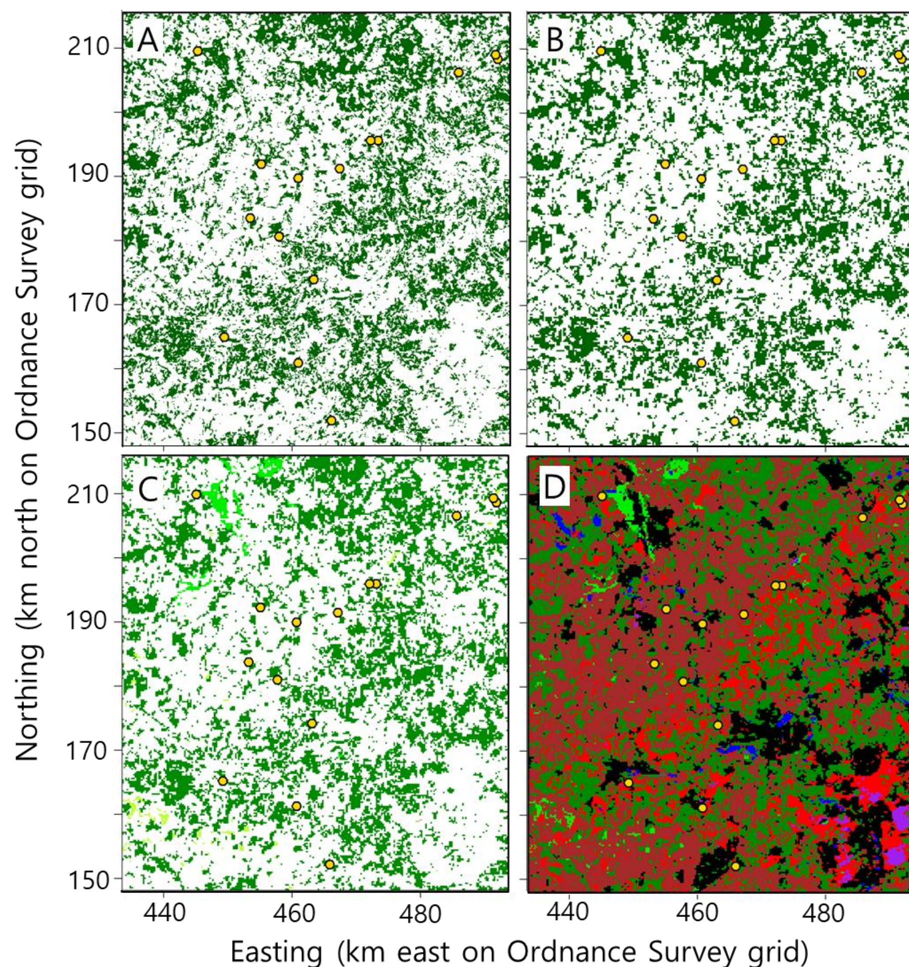


Figure 3.2 Maps showing the four landscape surfaces used in the ResistanceGA analysis.
A) Study landscape at the 100m pixel resolution. Binary landscape grass and non-grass. Dark green = grass, white = non-grass. B) Study landscape at the 250m pixel resolution. Binary landscape grass and non-grass. Dark green = grass, white = non-grass. C) Study landscape at the 250m pixel resolution, multiple grass types included. Green = improved grassland, light green = neutral grassland, lime green = calcareous grassland, brown = acid grassland. D) Study landscape at the 250m pixel resolution, multiple habitat types included. Brown = arable, red = woodland, green = improved grassland, light green = semi-natural grassland, blue = freshwater, black = built up areas, purple = heathland. Site locations marked by yellow points.

3.4 Results

3.4.1 Microsatellites, HWE and GE

There was no evidence of linkage disequilibrium between pairs of loci for any of the three datasets (Table 3.1). With the exception of a few loci at single locations and time points, site locus combinations had a low frequency (<0.2) of null alleles (Tables 3.2, A-C). In all three datasets the microsatellites displayed a high level of variability (Table 3.3), with the highest values of H_o occurring at Mj5331 (0.877, 0.902 and 0.905 for 2012, 2017 and All Years respectively) and the lowest occurring at Mj4870 (0.262, 0.279 and 0.277). Across each dataset all loci showed a non-significant level of heterozygote deficit across the population as a whole, the greatest occurring at Mj4870 with deficits of 31.444% (2012), 24.596% (2017) and 24.337% (All Years) (Table 3.3). The high heterozygote deficit of this locus led to all analyses being re-done with the locus Mj4870 removed. This had no significant effect on the overall results therefore the locus was included. No F_{ST} values were significantly greater than zero at any locus within any of the three datasets. However, F_{IS} values were significantly greater than zero at the following loci (Table 3.3), indicating potential deviations from HWE:

- 2012 Mj7232, Mj4870, Mj0247
- 2017 Mj7232, Mj4870, Mj0247, Mj5522
- All years Mj7232, Mj4870, Mj0247, Mj5522

Table 3.1 Composite linkage disequilibrium test outputs for all locus pair combinations for the datasets 2012, 2017 and All Years.

Locus pair		2012			2017			All Years		
		Chi^2	df	p-Value	Chi^2	df	p-Value	Chi^2	df	p-Value
Mj7232	Mj5522	33.498	28	0.218	30.528	30	0.439	19.476	44	1.000
Mj7232	Mj0247	17.833	22	0.716	23.029	26	0.631	3.754	34	1.000
Mj5522	Mj0247	17.482	22	0.736	7.734	26	1.000	28.455	36	0.811
Mj7232	Mj4870	25.673	28	0.591	30.842	30	0.423	35.302	44	0.822
Mj5522	Mj4870	19.728	28	0.874	15.522	30	0.986	35.110	46	0.879
Mj0247	Mj4870	8.877	22	0.994	20.015	26	0.791	24.946	36	0.917
Mj7232	Mj7132	27.608	28	0.485	16.554	30	0.978	14.614	42	1.000
Mj5522	Mj7132	18.644	28	0.909	23.531	30	0.793	15.838	44	1.000
Mj0247	Mj7132	5.156	22	1.000	13.529	26	0.979	11.729	34	1.000
Mj4870	Mj7132	21.817	28	0.790	23.315	30	0.802	27.830	44	0.973
Mj7232	Mj5331	6.305	24	1.000	24.854	26	0.527	9.982	40	1.000
Mj5522	Mj5331	11.532	24	0.985	14.212	26	0.970	11.567	42	1.000
Mj0247	Mj5331	12.538	20	0.896	13.037	22	0.932	19.544	34	0.978
Mj4870	Mj5331	11.300	22	0.970	17.472	26	0.894	21.974	42	0.995
Mj7132	Mj5331	24.511	24	0.433	11.868	26	0.992	37.668	42	0.661

Table 3.2 Locus by populations estimated null allele frequencies for A) 2012 B) 2017 and C) Subset of three sites for 2012-2019 inclusive. Values in bold exceed 0.2 frequency of null alleles. NA = no information available. This is due to only 9 individuals occurring in the dataset for the site year combination B-14 and all individuals being homozygous at 173 for the locus Mj4870.

3.2.A 2012 null allele frequencies

	ARN	ARS	AU	B	C	CH	D	HP	LC	LW	MC	SD	TC	WW	Mean
Mj7232	0.000	0.032	0.087	0.012	0.037	0.061	0.165	0.114	0.151	0.074	0.067	0.126	0.127	0.076	0.081
Mj5522	0.000	0.036	0.025	0.049	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.025	0.018
Mj0247	0.058	**0.222**	0.032	0.000	0.040	0.038	0.060	0.033	0.035	0.083	0.010	0.082	0.092	0.000	0.056
Mj4870	0.159	0.169	0.094	0.000	0.092	0.151	0.000	**0.202**	0.192	0.135	0.161	0.113	0.114	0.107	0.121
Mj7132	0.000	0.000	0.000	0.040	0.031	0.022	0.000	0.000	0.010	0.000	0.000	0.113	0.000	0.000	0.015
Mj5331	0.000	0.049	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.006

3.2.B 2017 null allele frequencies

	ARN	ARS	AU	B	C	CH	D	HP	LC	LW	MC	PF	SD	TC	WW	Mean
Mj7232	0.028	0.145	0.000	0.137	0.079	0.096	0.022	0.032	0.086	0.000	0.171	0.050	0.048	0.013	0.032	0.063
Mj5522	0.000	0.055	0.048	0.000	0.000	0.000	0.030	0.055	0.006	0.013	0.029	0.021	0.076	0.074	0.000	0.027
Mj0247	0.042	0.030	0.000	0.085	0.018	0.000	0.041	0.000	0.000	0.037	0.111	0.064	0.068	0.000	0.048	0.036
Mj4870	0.075	**0.205**	0.000	0.000	0.000	0.064	0.086	0.003	0.159	0.151	0.000	0.113	0.150	0.189	0.128	0.088
Mj7132	0.022	0.065	0.000	0.045	0.000	0.000	0.091	0.000	0.000	0.028	0.000	0.000	0.047	0.002	0.000	0.020
Mj5331	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.030	0.028	0.028	0.027	0.000	0.000	0.000	0.010

3.2.C Null allele frequencies across all years

	2012			2013			2014			2015			2016		
	AU	B	C	AU	B	C	AU	B	C	AU	B	C	AU	B	C
Mj7232	0.087	0.012	0.037	0.083	**0.208**	0.032	0.000	0.094	0.027	0.104	0.081	**0.215**	0.154	0.108	0.020
Mj5522	0.000	0.049	0.044	0.017	0.176	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.064
Mj0247	0.032	0.000	0.040	0.017	0.000	0.178	0.106	0.000	0.008	0.112	0.090	0.000	0.118	0.102	0.101
Mj4870	0.094	0.000	0.092	0.000	0.125	**0.263**	0.000	NA	0.103	0.000	0.053	0.000	0.093	0.102	0.153
Mj7132	0.000	0.040	0.031	0.047	0.000	0.060	0.099	0.096	0.000	0.051	0.000	0.000	0.000	0.000	0.024
Mj5331	0.000	0.000	0.000	0.000	0.000	0.000	0.064	0.000	0.000	0.002	0.000	0.000	0.000	0.017	0.000

3.2.C continued

	2017			2018			2019			Mean
	AU	B	C	AU	B	C	AU	B	C	
Mj7232	0.000	0.137	0.079	0.045	0.060	0.051	0.000	0.050	**0.300**	0.083
Mj5522	0.048	0.000	0.000	0.000	0.023	0.011	0.000	0.003	0.108	0.024
Mj0247	0.000	0.085	0.018	0.056	0.126	0.137	0.181	0.073	0.000	0.066
Mj4870	0.000	0.000	0.000	0.096	0.196	0.000	0.143	0.081	0.098	0.073
Mj7132	0.000	0.045	0.000	0.060	0.000	0.000	0.000	0.030	0.025	0.025
Mj5331	0.000	0.032	0.000	0.026	0.000	0.012	0.000	0.000	0.000	0.006

Table 3.3 Population-wide expected and observed heterozygosity, and percent difference ((E - O)/E*100), F_{IT}, F_{IS}, F_{ST} and R_{ST} at each locus. Bartlett's K-squared: 2012 = 0.049, df = 1, p-value = 0.8248. 2017 = 0.03, df = 1, p-value = 0.8618. All years = 0.006, df = 1, p-value = 0.9403.

	Locus	Number of samples	Number of alleles	H_e	H_o	H_e vs H_o % difference	F_{IT} (p-value)	F_{ST} (p-value)	F_{IS} (p-value)	R_{ST}
2012	Mj7232	251	13	0.804	0.733	-8.824	0.091 (0.001)	0.006 (0.967)	0.085 (0.002)	-0.014
	Mj5522	251	11	0.860	0.813	-5.548	0.058 (0.007)	0.012 (0.700)	0.046 (0.057)	-0.013
	Mj0247	250	29	0.941	0.808	-14.179	0.144 (0.000)	0.003 (1.000)	0.141 (0.000)	0.002
	Mj4870	248	5	0.382	0.262	-31.444	0.316 (0.000)	-0.015 (1.000)	0.326 (0.000)	-0.022
	Mj7132	248	10	0.768	0.758	-1.291	0.015 (0.342)	0.002 (0.997)	0.014 (0.380)	0.005
	Mj5331	252	26	0.894	0.877	-1.901	0.021 (0.186)	0.002 (1.000)	0.019 (0.211)	-0.009
	Mean	250	15.7	0.775	0.709	-10.531	0.107 (-)	0.002 (-)	0.105 (-)	-0.009
2017	Mj7232	285	12	0.798	0.762	-4.488	0.048 (0.035)	0.002 (0.999)	0.046 (0.049)	-0.013
	Mj5522	281	12	0.862	0.809	-6.226	0.064 (0.008)	0.000 (1.000)	0.064 (0.006)	0.011
	Mj0247	283	31	0.941	0.842	-10.614	0.105 (0.000)	0.000 (1.000)	0.105 (0.000)	-0.007
	Mj4870	282	6	0.370	0.279	-24.596	0.252 (0.000)	0.000 (0.982)	0.252 (0.000)	0.020
	Mj7132	282	10	0.741	0.752	1.500	-0.013 (0.692)	0.007 (0.911)	-0.020 (0.770)	0.005
	Mj5331	286	22	0.894	0.902	0.913	-0.007 (0.667)	0.002 (1.000)	-0.009 (0.692)	-0.010
	Mean	283	15.5	0.768	0.724	-7.252	0.075 (-)	0.002 (-)	0.073 (-)	0.001
All Years	Mj7232	426	14	0.819	0.719	-12.234	0.125 (0.000)	0.016 (0.429)	0.111 (0.000)	0.004
	Mj5522	426	12	0.865	0.806	-6.888	0.071 (0.000)	0.012 (0.739)	0.059 (0.001)	0.002
	Mj0247	424	31	0.936	0.758	-19.083	0.190 (0.000)	0.002 (1.000)	0.188 (0.000)	-0.010
	Mj4870	428	6	0.367	0.277	-24.337	0.247 (0.000)	0.006 (0.949)	0.243 (0.000)	0.001
	Mj7132	427	10	0.737	0.707	-4.055	0.042 (0.048)	0.003 (0.999)	0.039 (0.064)	-0.009
	Mj5331	430	24	0.898	0.905	0.757	-0.006 (0.680)	0.008 (0.988)	-0.014 (0.808)	-0.008
	Mean	427	16.2	0.770	0.695	-10.973	0.112 (-)	0.008 (-)	0.104 (-)	-0.003

3.4.2 Population genetic diversity

The genetic diversity, relatedness and effective population sizes of *M. jurtina* for each dataset can be found in Table 3.4. H_e across the whole datasets was 0.772 in 2012, 0.764 in 2017 and 0.764 across All Years. In 2012 genetic diversity (H_e and A_r) was similar across all populations with H_e ranging from 0.729 (C) to 0.825 (WW) and A_r ranging from 6.337 (C) to 7.726 (AU). Two private alleles (A_P) occurred at ARS, AU and SD and TC. CH, D, LW, MC and WW each had a single private allele and all other sites had no private alleles.

In 2017 H_e and A_r were similar across all populations with H_e ranging from 0.713 (AU) to 0.805 (CH) and A_r ranging from 7.239 (ARS) to 8.695 (PF). Five private alleles (A_P) occurred at PF, with two at LC. MC and TC each had a single private allele and all other sites had no private alleles.

Across all years H_e and A_r were similar across all populations (Fig. 3.3), with H_e ranging from 0.704 (B-14) to 0.814 (AU-13) and A_r ranging from 5.931 (C-12) to 7.135 (AU-12). Two private alleles (A_P) occurred at AU-12 and AU-19, with one occurring at C-12, AU-13, B-13, AU-15, C-16, AU-18 and C-18. All other sites had no private alleles.

Across all three datasets using both the heterozygote excess and the linkage disequilibrium method effective population sizes were estimated at infinity. In some case the point estimates were estimated at infinity, in others the confidence limits included infinity.

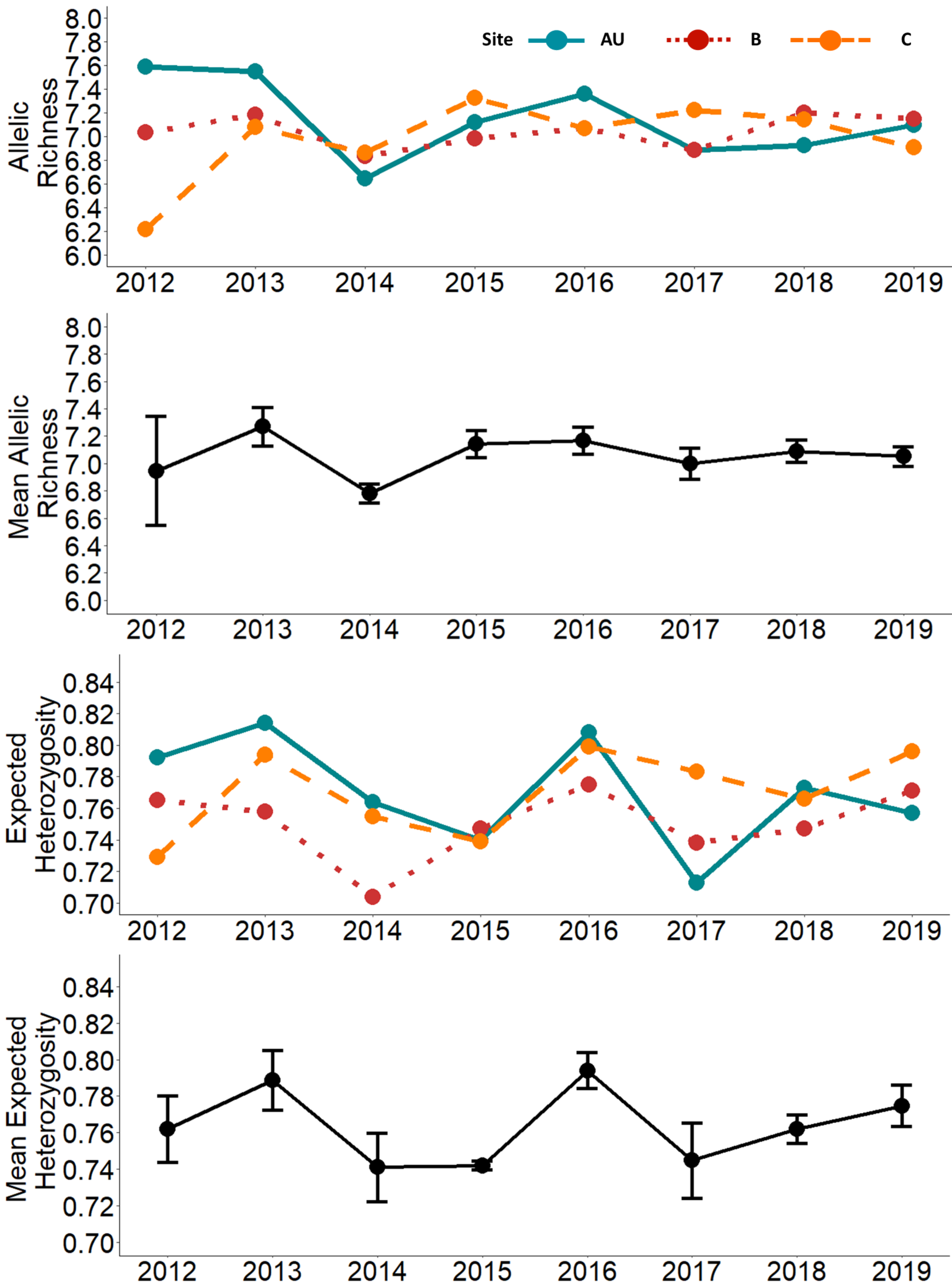


Figure 3.3 Changes in yearly (A) individual allelic richness values, (B) mean combined allelic richness values, (C) individual expected heterozygosity, and (D) mean combined expected heterozygosity for each of the three sites AU, B and C for the years 2012 – 2019.

Table 3.4 Sample sizes, genetic diversity, allelic richness, number of private alleles and effective population sizes for *M. jurtina* populations in the south of England. All values are estimated on a per population basis.

H_{exp} expected heterozygosity, A_r = allelic richness, A_p = Private alleles, $N_e(1)$ = effective population size estimated using the heterozygote excess method, $N_e(2)$ = effective population size estimated using linkage disequilibrium method.

Data set	Year	Site	Sample Size	Mean H_{exp}	(s.d.)	A_r	A_p	$N_e(1)$	$N_e(2)$
2012 Only	2012	All Sites	252	0.774	(0.200)	-	13	-	-
2012 Only	2012	Aston Rowant North (ARN)	22	0.782	(0.180)	7.624	0	∞	∞
2012 Only	2012	Aston Rowant South (ARS)	18	0.787	(0.207)	8.003	2	∞	∞
2012 Only	2012	Aston Upthorpe (AU)	18	0.792	(0.177)	8.299	2	∞	∞
2012 Only	2012	Bowdown Forest (B)	20	0.765	(0.214)	7.650	0	∞	∞
2012 Only	2012	Crabtree Plantation (C)	20	0.729	(0.252)	6.688	0	∞	∞
2012 Only	2012	Coombe Hill (CH)	20	0.763	(0.204)	7.519	1	∞	∞
2012 Only	2012	Dancersend (D)	16	0.737	(0.263)	7.646	1	∞	∞
2012 Only	2012	Howbery Park (HP)	16	0.777	(0.178)	8.099	0	∞	∞
2012 Only	2012	Lardon Chase (LC)	14	0.792	(0.187)	7.737	0	∞	∞
2012 Only	2012	Little Whittenham (LW)	20	0.766	(0.184)	7.556	1	∞	∞
2012 Only	2012	Moore Copse (MC)	20	0.774	(0.224)	8.079	1	∞	∞
2012 Only	2012	Swyncombe Down (SD)	21	0.787	(0.161)	7.937	2	∞	∞
2012 Only	2012	The Crong (TC)	16	0.760	(0.221)	7.820	2	∞	∞
2012 Only	2012	Wytham Woods (WW)	11	0.825	(0.154)	8.333	1	∞	∞
2017 Only	2017	All Sites	287	0.764	0.215	-	9	-	-
2017 Only	2017	Aston Rowant North (ARN)	21	0.760	0.176	8.278	0	∞	∞
2017 Only	2017	Aston Rowant South (ARS)	17	0.753	0.224	7.765	0	∞	∞
2017 Only	2017	Aston Upthorpe (AU)	14	0.713	0.269	8.500	0	∞	∞
2017 Only	2017	Bowdown Forest (B)	17	0.738	0.246	8.039	0	∞	∞
2017 Only	2017	Crabtree Plantation (C)	20	0.783	0.199	8.444	0	∞	∞
2017 Only	2017	Coombe Hill (CH)	20	0.805	0.145	8.689	0	∞	∞
2017 Only	2017	Dancersend (D)	15	0.776	0.223	8.806	0	∞	∞
2017 Only	2017	Howbery Park (HP)	20	0.786	0.163	8.438	0	∞	∞
2017 Only	2017	Lardon Chase (LC)	20	0.796	0.176	9.096	2	∞	∞
2017 Only	2017	Little Whittenham (LW)	20	0.784	0.229	8.682	0	∞	∞
2017 Only	2017	Moore Copse (MC)	16	0.736	0.266	8.858	1	∞	∞
2017 Only	2017	Pamber Forest (PF)	37	0.771	0.223	9.030	5	∞	∞
2017 Only	2017	Swyncombe Down (SD)	15	0.747	0.279	8.773	0	∞	∞
2017 Only	2017	The Crong (TC)	15	0.771	0.213	8.286	1	∞	∞
2017 Only	2017	Wytham Woods (WW)	20	0.745	0.192	8.097	0	∞	∞
All Years	-	All Sites	342	0.764	0.214	-	11	-	-
All Years	2012	Aston Upthorpe (AU)	18	0.792	0.177	7.591	2	∞	∞

Data set	Year	Site	Sample Size	Mean H_{exp}	(s.d.)	A_r	A_p	$N_e(1)$	$N_e(2)$
All Years	2012	Bowdown Forest (B)	20	0.765	0.214	7.034	0	∞	∞
All Years	2012	Crabtree Plantation (C)	20	0.729	0.252	6.215	1	∞	∞
All Years	2013	Aston Upthorpe (AU)	11	0.814	0.174	7.548	1	∞	∞
All Years	2013	Bowdown Forest (B)	16	0.758	0.166	7.185	1	∞	∞
All Years	2013	Crabtree Plantation (C)	16	0.794	0.155	7.079	0	∞	∞
All Years	2014	Aston Upthorpe (AU)	20	0.764	0.177	6.647	0	∞	∞
All Years	2014	Bowdown Forest (B)	9	0.704	0.354	6.833	0	∞	∞
All Years	2014	Crabtree Plantation (C)	15	0.755	0.237	6.865	0	∞	∞
All Years	2015	Aston Upthorpe (AU)	35	0.740	0.258	7.121	1	∞	∞
All Years	2015	Bowdown Forest (B)	25	0.747	0.225	6.985	0	∞	∞
All Years	2015	Crabtree Plantation (C)	10	0.739	0.288	7.324	0	∞	∞
All Years	2016	Aston Upthorpe (AU)	20	0.808	0.142	7.363	0	∞	∞
All Years	2016	Bowdown Forest (B)	20	0.775	0.189	7.069	0	∞	∞
All Years	2016	Crabtree Plantation (C)	20	0.799	0.150	7.067	1	∞	∞
All Years	2017	Aston Upthorpe (AU)	14	0.713	0.269	6.885	0	∞	∞
All Years	2017	Bowdown Forest (B)	17	0.738	0.246	6.885	0	∞	∞
All Years	2017	Crabtree Plantation (C)	20	0.783	0.199	7.226	0	∞	∞
All Years	2018	Aston Upthorpe (AU)	21	0.773	0.169	6.927	1	∞	∞
All Years	2018	Bowdown Forest (B)	20	0.747	0.246	7.203	0	∞	∞
All Years	2018	Crabtree Plantation (C)	21	0.766	0.251	7.141	1	∞	∞
All Years	2019	Aston Upthorpe (AU)	18	0.757	0.223	7.100	2	∞	∞
All Years	2019	Bowdown Forest (B)	16	0.771	0.187	7.148	0	∞	∞
All Years	2019	Crabtree Plantation (C)	10	0.796	0.183	6.908	0	∞	∞

3.4.3 Genetic structure, differentiation and gene flow

For all three datasets used within this study, using the six microsatellites described, no evidence of structure was found between these populations using the programme STRUCTURE. No population was found to be genetically distinct from any other population, i.e. K = 1 (Fig. 3.3), therefore STRUCTURE harvester was not used to determine K.

Within each dataset pairwise F_{ST} was very low and there was very little variation within the data (2012: mean = 0.003, variance = 0.00005, 2017: mean = 0.002, variance = 0.00004, Across all years: mean = 0.008, variance = 0.00008) (Tables 3.5, 3.6 and 3.7). For 2012 and 2017 no F_{ST} scores between site pair combinations were significantly greater than zero. Across all years pairwise F_{ST} was significantly greater than zero for eight site/year pairwise combinations (C-12 and AU-14, C-12 and AU- 16, C-12 and C-16, C-12 and AU-19, AU-15 and AU-18, AU-16 and AU-18, C-17 and AU-19, AU-18 and AU-19).

Additionally, there was no evidence of IBD between pairwise site combinations and pairwise F_{ST} values within any dataset (Fig. 3.4) (2012: p = 0.280 and Mantel r value of 0.079. 2017: p = 0.279 and Mantel r value of -0.078. All years: p = 0.078 and Mantel r value of 0.085), nor did habitat

similarity have a significant effect on pairwise F_{ST} (Fig. 3.5) (2012: p = 0.870, Mantel R = -0.155. 2017 p = 0.091, Mantel R = 0.174. All Years p = 0.930, Mantel R = -0.104). Yearly changes to pairwise F_{ST} can be seen in Fig. 3.6.

When only using data from the All Years dataset, the estimated number of migrants per generation according to Slatkin's private allele method ranged from 3.140 migrants in 2016 to 6.023 in 2018. However, when using all available sites for the years 2012 and 2017, in 2012 the number of estimated migrants drops to 5.085 (5.452 using only three sites) and the 2017 estimate increases to 13.141 (from 4.367) (Table 3.8).

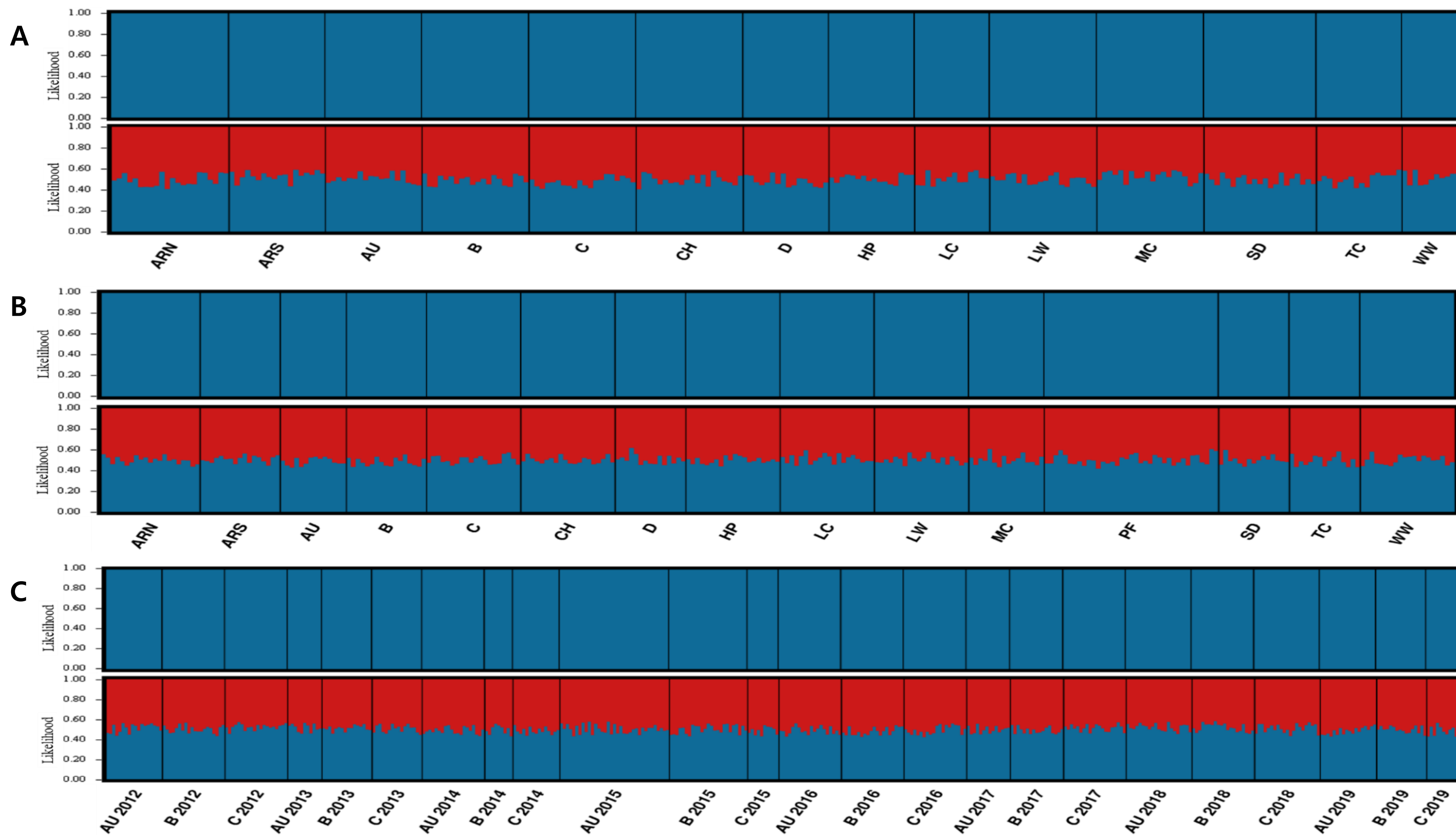


Figure 3.4 STRUCTURE individual assignment bar plots for *M. jurtina* individuals collected within the study area in A) 2012 B) 2017 C) Across eight years (2012-1017 inclusive). Individuals split by site on the x axis and likelihood of assignment of the individual into genetic clusters on the y. Colours indicate different genetic clusters. Plot show K = 1 and K = 2 for all three datasets.

Table 3.5 Pairwise F_{ST} values between fourteen pairs of sites in 2012. Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 1820 permutations, indicative adjusted nominal level (5%) for multiple comparisons is 0.000549. All values are non-significant.

	ARN	ARS	AU	B	C	CH	D	HP	LC	LW	MC	SD	TC	WW
ARN	-	0.003	0.315	0.271	0.006	0.039	0.558	0.126	0.086	0.035	0.001	0.066	0.263	0.248
ARS	0.014	-	0.039	0.003	0.001	0.048	0.027	0.051	0.423	0.037	0.391	0.009	0.014	0.057
AU	-0.002	0.005	-	0.736	0.065	0.366	0.929	0.315	0.574	0.508	0.051	0.718	0.765	0.332
B	0.004	0.011	-0.006	-	0.134	0.252	0.83	0.655	0.427	0.655	0.011	0.442	0.311	0.24
C	0.013	0.018	0.006	0.003	-	0.025	0.073	0.087	0.009	0.632	0.004	0.098	0.036	0.04
CH	0.006	0.003	-0.006	-0.005	0.005	-	0.083	0.531	0.443	0.29	0.232	0.449	0.189	0.677
D	-0.005	0.012	-0.006	-0.001	0.004	0.004	-	0.213	0.111	0.169	0.004	0.286	0.725	0.142
HP	0.015	0.010	0.006	0.002	0.005	0.003	0.014	-	0.562	0.623	0.308	0.4	0.287	0.599
LC	0.011	-0.008	-0.001	0.001	0.007	0.001	0.013	-0.003	-	0.252	0.149	0.114	0.177	0.712
LW	0.012	0.003	-0.005	-0.007	-0.005	-0.004	0.007	-0.005	0.000	-	0.495	0.702	0.087	0.148
MC	0.021	-0.002	0.008	0.010	0.015	0.003	0.020	0.005	-0.002	0.000	-	0.004	0.042	0.086
SD	0.004	0.011	-0.009	-0.004	0.004	-0.005	0.005	0.000	0.009	-0.005	0.015	-	0.246	0.419
TC	0.000	0.009	-0.006	0.000	0.008	-0.001	-0.007	0.008	0.002	0.007	0.007	-0.001	-	0.229
WW	0.004	0.006	-0.001	0.001	0.004	-0.004	0.014	-0.005	-0.016	0.001	0.007	-0.007	0.007	-

Table 3.6 Pairwise F_{ST} values between fifteen pairs of sites in 2017. Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 2100 permutations, indicative adjusted nominal level (5%) for multiple comparisons is 0.000476. All values are non-significant.

	ARN	ARS	AU	B	C	CH	D	HP	LC	LW	MC	PF	SD	TC	WW
ARN	-	0.037	0.229	0.333	0.036	0.515	0.459	0.207	0.260	0.014	0.034	0.035	0.073	0.020	0.003
ARS	0.004	-	0.290	0.073	0.060	0.142	0.450	0.387	0.140	0.103	0.336	0.062	0.414	0.053	0.019
AU	0.004	-0.003	-	0.897	0.390	0.388	0.412	0.714	0.211	0.289	0.680	0.522	0.866	0.326	0.330
B	0.013	0.006	-0.003	-	0.098	0.092	0.306	0.896	0.389	0.167	0.487	0.882	0.958	0.451	0.055
C	0.007	-0.003	0.003	0.005	-	0.370	0.275	0.730	0.545	0.766	0.263	0.090	0.788	0.146	0.018
CH	0.001	0.012	0.011	0.020	0.002	-	0.127	0.339	0.719	0.390	0.411	0.123	0.485	0.141	0.027
D	0.001	-0.011	0.002	0.001	-0.003	0.012	-	0.411	0.864	0.843	0.929	0.268	0.619	0.600	0.034
HP	0.003	-0.004	-0.003	-0.009	-0.006	0.003	-0.002	-	0.436	0.600	0.845	0.891	0.947	0.467	0.171
LC	0.006	0.003	0.013	0.004	-0.003	0.003	-0.010	-0.003	-	0.929	0.255	0.255	0.731	0.177	0.105
LW	0.012	0.000	0.008	0.012	-0.007	0.004	-0.008	0.002	-0.008	-	0.359	0.435	0.932	0.031	0.008
MC	0.009	-0.003	-0.005	-0.003	0.001	0.014	-0.012	-0.006	0.005	0.006	-	0.552	0.761	0.380	0.135
PF	0.009	0.000	-0.005	-0.004	0.000	0.011	-0.006	-0.006	0.002	-0.001	-0.006	-	0.917	0.565	0.010
SD	0.014	-0.004	-0.005	-0.007	-0.008	0.008	-0.002	-0.005	0.001	-0.009	-0.003	-0.005	-	0.428	0.149
TC	0.015	0.004	0.003	0.005	-0.004	0.003	-0.002	-0.003	0.000	0.005	0.006	-0.002	-0.002	-	0.451
WW	0.007	0.009	-0.005	0.012	0.011	0.011	0.010	0.000	0.008	0.015	0.003	0.008	0.007	0.005	-

Table 3.7 Pairwise F_{ST} values between three pairs of sites across eight years (2012-2019). Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 5520 permutations, indicative adjusted nominal level (5%) for multiple comparisons is 0.000181. F_{ST} scores significantly greater than zero are in bold

	AU-12	B-12	C-12	AU-13	B-13	C-13	AU-14	B-14	C-14	AU-15	B-15	C-15	AU-16	B-16	C-16
AU-12	-	0.747	0.057	0.022	0.803	0.849	0.042	0.156	0.078	0.011	0.084	0.283	0.007	0.078	0.259
B-12	-0.006	-	0.134	0.114	0.547	0.889	0.002	0.014	0.080	0.070	0.302	0.050	0.032	0.167	0.016
C-12	0.006	0.003	-	0.001	0.002	0.067	**0.00018**	0.002	0.001	0.001	0.0004	0.007	**0.00018**	0.0004	**0.00018**
AU-13	0.017	0.005	0.021	-	0.206	0.180	0.001	0.113	0.036	0.115	0.115	0.024	0.119	0.010	0.001
B-13	-0.004	0.004	0.025	0.016	-	0.171	0.114	0.420	0.042	0.133	0.229	0.202	0.019	0.174	0.044
C-13	-0.007	-0.009	0.005	0.004	0.016	-	0.004	0.188	0.212	0.006	0.088	0.344	0.177	0.339	0.025
AU-14	0.001	0.006	**0.024**	0.019	-0.002	0.016	-	0.118	0.377	0.002	0.149	0.664	0.030	0.195	0.031
B-14	0.007	0.008	0.011	0.025	0.006	0.011	0.005	-	0.516	0.226	0.379	0.446	0.083	0.381	0.142
C-14	0.010	0.007	0.018	0.016	0.020	0.006	-0.003	-0.003	-	0.060	0.661	0.413	0.075	0.466	0.168
AU-15	0.008	0.005	0.012	0.014	0.006	0.019	0.011	-0.003	0.007	-	0.392	0.190	0.099	0.047	0.002
B-15	0.006	0.005	0.019	0.016	0.004	0.016	0.004	-0.003	0.001	-0.005	-	0.608	0.372	0.422	0.021
C-15	0.001	0.005	0.017	0.025	0.004	0.006	-0.010	-0.011	-0.003	-0.002	-0.008	-	0.773	0.859	0.242
AU-16	0.008	0.002	**0.029**	0.003	0.009	0.004	0.006	0.013	0.009	0.003	-0.001	-0.003	-	0.328	0.058
B-16	0.005	0.000	0.018	0.018	0.010	0.002	0.002	-0.001	0.000	0.005	0.000	-0.008	-0.003	-	0.105
C-16	0.002	0.008	**0.031**	0.022	0.017	0.012	0.010	0.016	0.014	0.021	0.017	0.010	0.002	0.004	-
AU-17	0.000	-0.007	0.008	0.026	0.006	0.007	0.004	-0.005	0.011	0.001	0.005	-0.002	0.010	0.002	0.020
B-17	0.004	0.004	0.003	0.023	0.014	0.003	0.027	-0.002	0.015	-0.001	0.005	0.002	0.009	0.000	0.019
C-17	-0.005	-0.004	-0.004	0.000	0.002	-0.004	0.007	0.001	0.004	0.008	0.013	0.007	0.012	0.011	0.014
AU-18	0.002	0.004	0.013	0.023	0.013	-0.005	0.008	0.012	0.005	**0.023**	0.016	0.009	**0.022**	0.019	0.023
B-18	0.006	-0.002	0.009	0.015	0.001	0.014	0.004	-0.002	0.012	0.002	0.004	-0.004	0.010	0.007	0.020
C-18	-0.007	0.000	0.003	0.016	0.026	0.001	0.016	0.017	0.007	0.017	0.014	0.006	0.016	0.014	0.020
AU-19	0.010	0.012	**0.028**	0.019	0.020	0.015	-0.002	-0.001	-0.001	0.005	0.001	-0.014	-0.005	-0.003	0.008
B-19	0.005	0.011	0.012	0.019	0.006	0.018	0.003	-0.004	0.013	0.002	-0.001	-0.002	-0.003	0.000	0.008
C-19	0.001	0.012	0.017	-0.001	0.018	0.010	0.012	0.002	-0.004	0.007	0.010	-0.002	0.003	-0.006	0.007

Table 3.7 Continued

	AU-17	B-17	C-17	AU-18	B-18	C-18	AU-19	B-19	C-19
AU-12	0.570	0.501	0.464	0.534	0.203	0.890	0.006	0.086	0.301
B-12	0.773	0.380	0.472	0.626	0.552	0.418	0.001	0.005	0.027
C-12	0.050	0.136	0.213	0.030	0.090	0.029	**0. 00018**	0.007	0.006
AU-13	0.244	0.057	0.145	0.005	0.026	0.006	0.085	0.050	0.274
B-13	0.632	0.182	0.431	0.194	0.412	0.019	0.003	0.423	0.087
C-13	0.332	0.314	0.235	0.549	0.081	0.349	0.017	0.017	0.111
AU-14	0.067	0.001	0.001	0.003	0.072	0.001	0.381	0.539	0.067
B-14	0.508	0.178	0.198	0.096	0.314	0.003	0.584	0.589	0.243
C-14	0.072	0.037	0.011	0.069	0.012	0.017	0.544	0.258	0.607
AU-15	0.324	0.367	0.001	**0. 00018**	0.020	0.0004	0.036	0.767	0.358
B-15	0.338	0.186	0.003	0.016	0.100	0.005	0.163	0.859	0.236
C-15	0.271	0.376	0.010	0.047	0.384	0.043	0.888	0.866	0.564
AU-16	0.109	0.038	0.001	**0. 00018**	0.004	0.002	0.619	0.701	0.334
B-16	0.333	0.469	0.001	0.002	0.030	0.004	0.183	0.617	0.758
C-16	0.005	0.070	0.004	0.001	0.008	0.001	0.101	0.248	0.367
AU-17	-	0.901	0.376	0.191	0.598	0.165	0.068	0.168	0.042
B-17	-0.003	-	0.098	0.164	0.669	0.195	0.016	0.132	0.176
C-17	0.003	0.005	-	0.361	0.171	0.366	**0. 00018**	0.042	0.018
AU-18	0.015	0.015	0.003	-	0.149	0.205	**0. 00018**	0.001	0.001
B-18	-0.002	-0.001	0.006	0.011	-	0.054	0.001	0.051	0.077
C-18	0.008	0.007	-0.005	0.011	0.008	-	0.0004	0.003	0.055
AU-19	0.008	0.015	**0.018**	**0.021**	0.013	0.017	-	0.716	0.244
B-19	0.006	0.010	0.004	0.026	0.015	0.017	0.002	-	0.801
C-19	0.025	0.012	0.007	0.031	0.015	0.009	0.008	0.001	-

Table 3.8 Estimated number of migrants per population using Slatkin's private allele method.

Dataset	Year	Nm
2012	2012	5.085
2017	2017	13.141
All Years	2012	5.452
All Years	2013	3.399
All Years	2014	4.312
All Years	2015	5.440
All Years	2016	3.140
All Years	2017	4.367
All Years	2018	6.023
All Years	2019	5.779

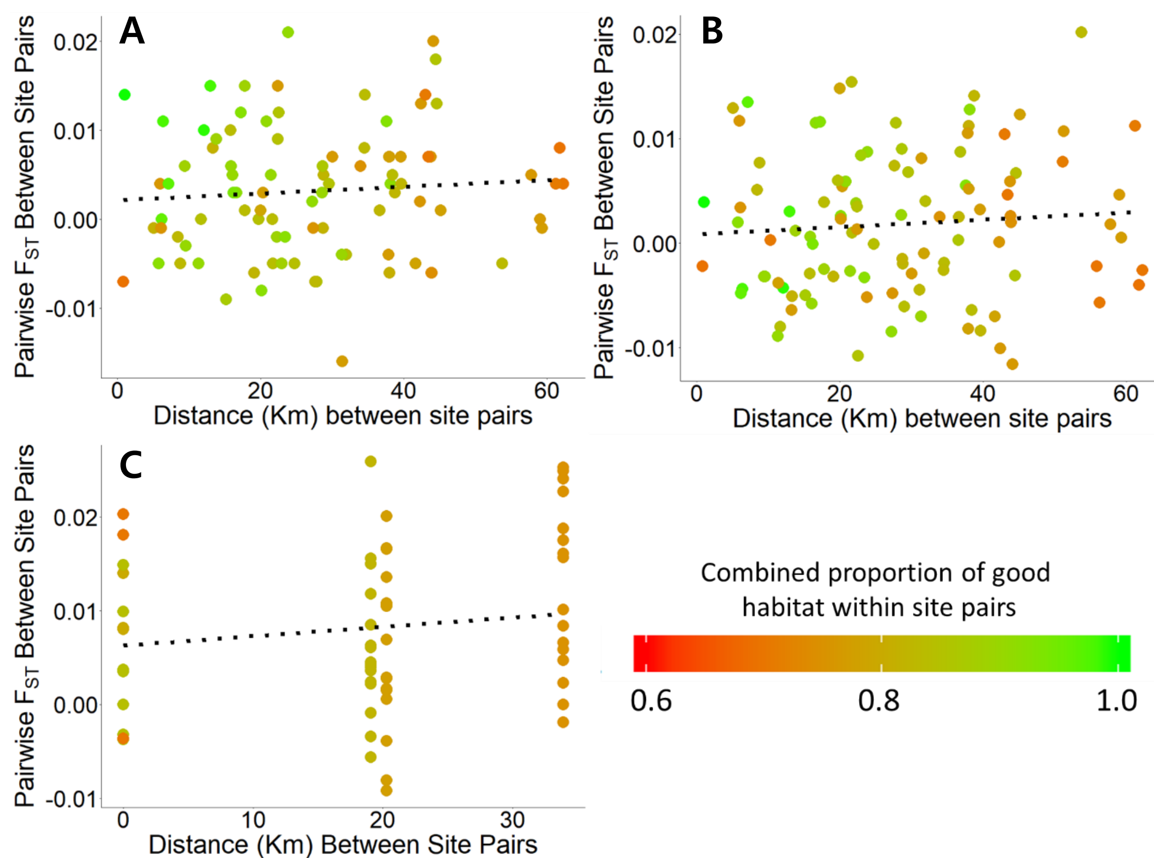


Figure 3.5 Pairwise F_{ST} values for *M. jurtina* plotted against Euclidean distances for A) fourteen sites in 2012, B) fifteen sites in 2017 and C) Three sites over eight years (2012 – 2019). All relationships are non-significant. Individual points are coloured by the combined proportions of good habitat within a 10km buffer around each site, within each site pair. Habitats are classified as good or bad based upon Villemey *et al.* (2016), whereby grassland elements were found to enhance gene flow in *M. jurtina* (classified as good habitat) and woodlands and arable lands inhibited gene flow (classified as bad habitat). Other habitat types were removed from the analysis.

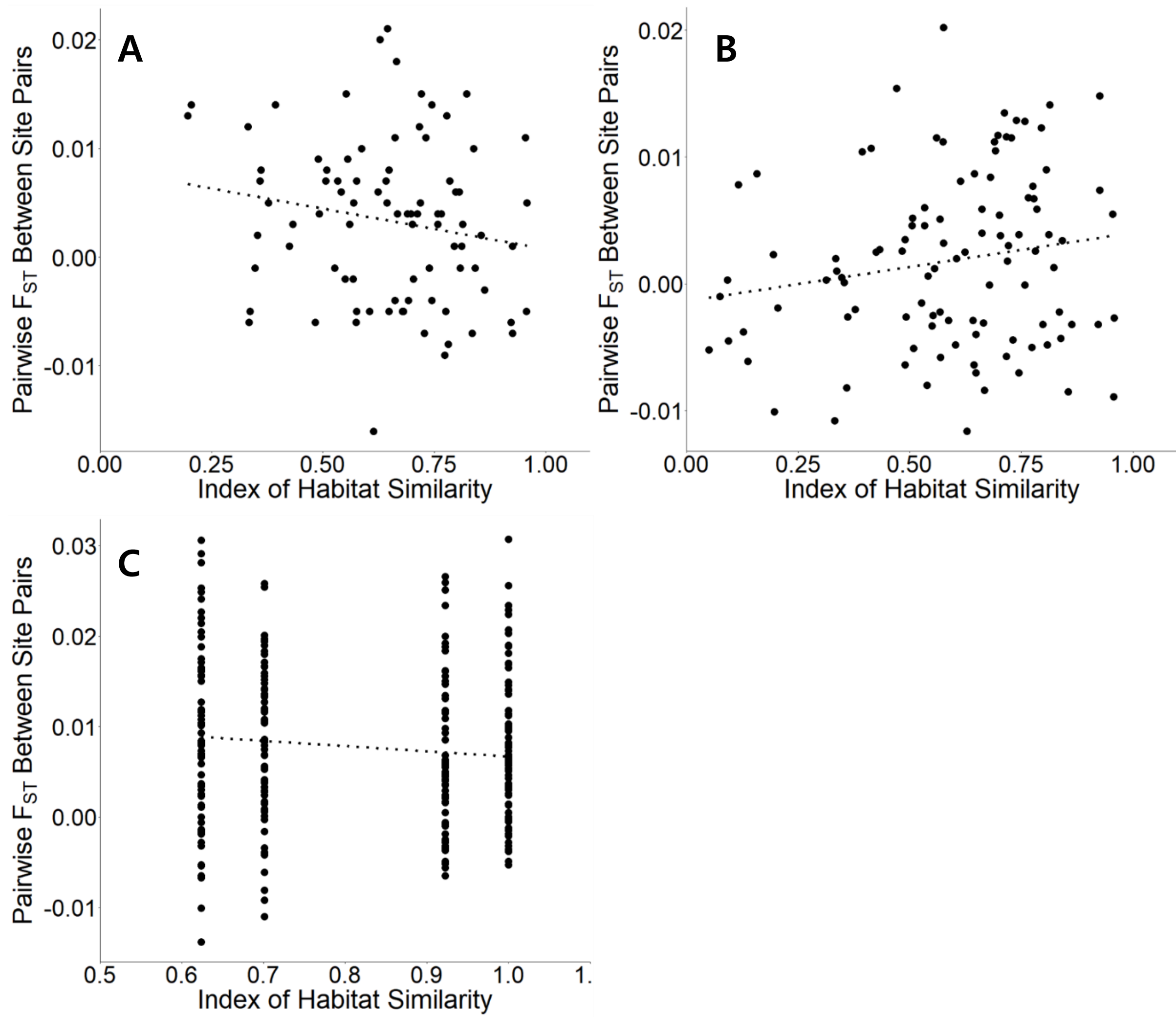


Figure 3.6 Pairwise F_{ST} values for *M. jurtina* plotted against an index of habitat similarities for A) fourteen pairwise site combinations in 2012 B) fifteen pairwise site combinations in 2017 and c) three pairwise site combinations across eight years from 2012 to 2019. All relationships are non-significant.

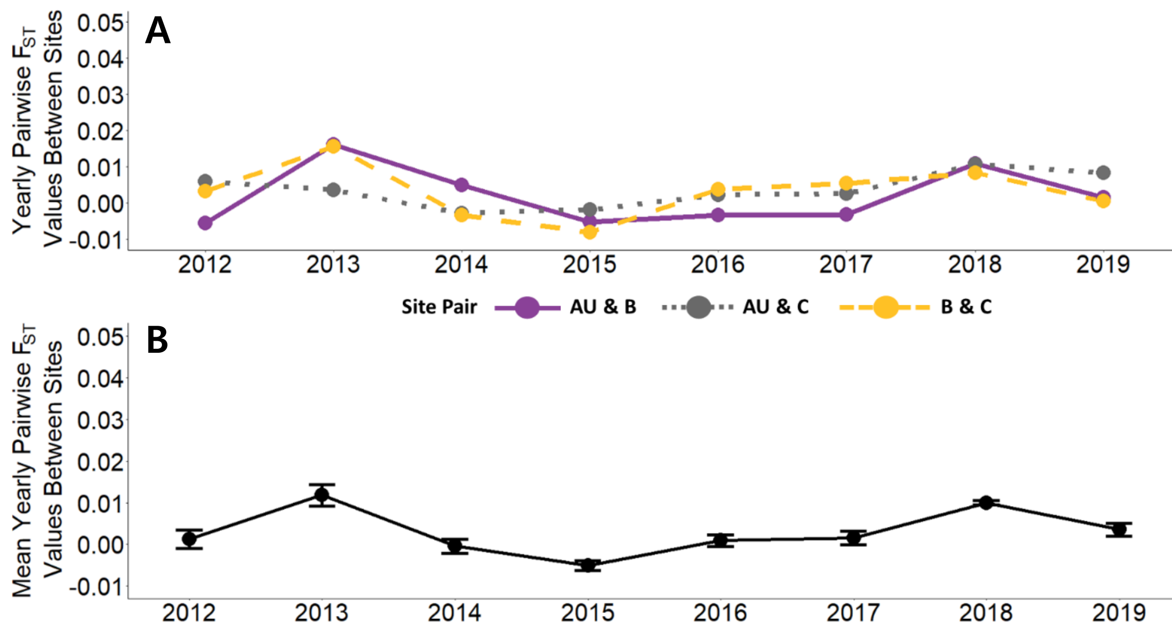


Figure 3.7 Changes in yearly mean (A) pairwise F_{ST} values for each of the three sites AU, B and C for the years 2012 - 2019 and (B) mean combined pairwise F_{ST} values for each of the three sites AU, B and C for the years 2012 – 2019.

3.4.4 Landscape genetics

For each genetic distance measure, replicate ResistanceGA runs resulted in no differences in AICc, log-likelihood values, or ranking of features between runs for the binary resistance surface. Replicate runs for the multiple grasses and multiple variables resistance surfaces showed only marginal differences between AICc or log-likelihood, however some changes occurred in the ranking of the best performing factor.

There was little evidence of either IBD or IBR in *M. jurtina* populations across the study area. Despite each of the initial runs ranking the IBR models higher than both the IBD null models based upon log-likelihood, using AICc the null model was the best fitting model across all runs. However, there was only a marginal difference in AICc (<2) between the null and binary landscape models at 250m resolution using PODA. This was not the case with F_{ST} ($\triangle$AICc = 2.056). The subsequent bootstrap analyses ranked the distance model more favourably to all other models tested (Table 3.10). The null model was not included in the bootstrapping. Further bootstrap analyses containing both resolutions of the binary surface marginally ranked the 250m resolution surface higher than the 100m resolution surface for both measures of genetic diversity (Table 3.10).

Table 3.9 Model selection results from ResistanceGA analysis (Peterman 2018) using PODA at (A) 250m and (B) 100m landscape resolution, and F_{ST} at (C) 250m and (D) 100m landscape resolution.

	Genetic Distance	Surface Resolution (m)	Optimised Surface	k	AIC	AICc	△AICc	R^2m	R^2c	LL
(A)										
	PODA	250	Null	1	-473.588	-473.28	0.000	0.000	0.293	239.794
	PODA	250	Binary Grass/Non-Grass	3	-474.472	-472.29	0.990	0.098	0.313	241.236
	PODA	250	Euclidean Distance	2	-471.606	-470.606	2.674	0.0002	0.293	239.803
	PODA	250	Multiple Grasses	6	-473.974	-463.474	9.806	0.090	0.321	240.987
	PODA	250	Multiple Variables	8	-474.091	-450.091	23.189	0.094	0.322	241.046
(B)										
	PODA	100	Null	1	-477.588	-477.28	0.000	0.000	0.293	239.794
	PODA	100	Multiple Grasses	2	-475.606	-474.606	2.674	0.0002	0.293	239.803
	PODA	100	Binary Grass/Non-Grass	3	-476.052	-473.871	3.409	0.086	0.312	241.026
(C)										
	F_{ST}	250	Null	1	-773.443	-773.135	0.000	0.000	0.280	389.721
	F_{ST}	250	Binary Grass/Non-Grass	3	-773.261	-771.079	2.056	0.060	0.286	390.630
	F_{ST}	250	Euclidean Distance	2	-771.443	-770.443	2.692	0.000	0.280	389.721
	F_{ST}	250	Multiple Grasses	6	-774.865	-764.365	8.770	0.120	0.318	391.432
	F_{ST}	250	Multiple Variables	8	-774.488	-750.488	22.647	0.109	0.314	391.244
(D)										
	F_{ST}	100	Null	1	-777.443	-777.135	0.000	0.000	0.280	389.721
	F_{ST}	100	Multiple Grasses	2	-775.443	-774.443	2.692	0.000	0.280	389.721
	F_{ST}	100	Binary Grass/Non-Grass	3	-774.241	-772.059	5.076	0.027	0.281	390.121

K = number of parameters fit in each model. AIC = original AIC value from MLPE mixed effects model. AICc = AIC value adjusted for number of populations and parameters optimized. △AICc = difference between AICc of model and the minimum AICc across all models. R^2m = marginal R^2. R^2c = conditional R^2. LL = log-likelihood of the MLPE model.

Table 3.10 Summary results of the bootstrap analysis conducted using the *resist.boot* function from *ResistanceGA* (Peterman 2018) for PODA at (A) 250m, (B) 100m and (C) combined 250m and 100m landscape resolution, and F_{ST} at (D) 250m, (E) 100m and (F) combined 250m and 100m landscape resolution.

	Genetic Distance	Surface Resolution (m)	Optimised Surface	k	Average AIC	Average AICc		Average Weight	Average R^2m	Average LL	Average Rank	Percentage Top Ranked
(A)												
	PODA	NA	Euclidean Distance	2	-246.430	-244.930	0.000	0.721	0.012	125.215	1.439	56.1
	PODA	250	Binary Grass/Non-Grass	3	-246.219	-242.790	2.140	0.279	0.115	126.109	1.561	43.9
	PODA	250	Multiple Grasses	6	-239.763	-218.763	26.167	0.000	0.100	125.881	3.000	0
	PODA	250	Multiple Variables	8	-235.838	-163.838	81.092	0.000	0.104	125.919	4.000	0
(B)												
	PODA	NA	Euclidean Distance	2	-246.664	-245.164	0.000	0.732	0.013	125.332	1.392	60.8
	PODA	100	Binary Grass/Non-Grass	3	-246.263	-242.835	2.329	0.268	0.106	126.132	1.608	39.2
(C)												
	PODA	NA	Euclidean Distance	2	-246.560	-245.060	0.000	0.585	0.013	125.280	1.804	49.4
	PODA	250	Binary Grass/Non-Grass	3	-246.306	-242.878	2.182	0.211	0.115	126.153	1.946	27.9
	PODA	100	Binary Grass/Non-Grass	3	-246.143	-242.715	2.345	0.204	0.106	126.072	2.250	22.7
	PODA	250	Multiple Grasses	6	-239.880	-218.880	26.18	0.000	0.102	125.940	4.000	0
	PODA	250	Multiple Variables	8	-235.959	-163.959	81.101	0.000	0.106	125.980	5.000	0
(D)												
	F_{ST}	NA	Euclidean Distance	2	-403.396	-401.896	0.000	0.765	0.014	203.698	1.333	70.7
	F_{ST}	250	Binary Grass/Non-Grass	3	-402.672	-399.244	2.652	0.235	0.085	204.336	1.738	26.8
	F_{ST}	250	Multiple Grasses	6	-397.535	-376.535	25.361	0.000	0.123	204.767	2.941	2.5
	F_{ST}	250	Multiple Variables	8	-393.315	-321.315	80.581	0.000	0.114	204.657	3.988	0
(E)												
	F_{ST}	NA	Euclidean Distance	2	-403.819	-402.319	0.000	0.814	0.014	203.910	1.157	84.3
	F_{ST}	250	Binary Grass/Non-Grass	3	-402.519	-399.090	3.229	0.186	0.060	204.259	1.843	15.7
(F)												
	F_{ST}	NA	Euclidean Distance	2	-403.224	-401.724	0.000	0.662	0.014	203.612	1.491	67
	F_{ST}	250	Binary Grass/Non-Grass	3	-402.518	-399.090	2.634	0.202	0.086	204.259	1.985	24.6
	F_{ST}	100	Binary Grass/Non-Grass	3	-401.813	-398.384	3.340	0.135	0.055	203.906	2.636	6
	F_{ST}	250	Multiple Grasses	6	-397.439	-376.439	25.285	0.000	0.126	204.720	3.906	2.4
	F_{ST}	250	Multiple Variables	8	-393.214	-321.214	80.510	0.000	0.117	204.607	4.982	0

K = number of parameters fit in each model. Average AIC = averaged original AIC across 1000 bootstrap iterations. AICc = averaged AIC value adjusted for number of populations and parameters optimized, across 1000 bootstrap iterations. Average Weight = averaged AICc weight across 1000 bootstrap iterations; △AICc = difference between AICc of model and the minimum AICc across all models; Average R^2m = averaged marginal R^2 across 1,000 bootstrap iterations. Average LL = average log-likelihood value across 1000 bootstrap iterations. Average Rank = Average rank position of model over 1000 bootstrap iterations. Percentage Top Ranked = Percentage of times that each model was ranked best across 100 bootstrap iterations.

3.5 Discussion

Despite a number of studies previously investigating measures of genetic diversity in *M. jurtina* (Thomson 1987; Goulson 1993a; Wood & Pullin 2002; Schmitt *et al.* 2005; Habel *et al.* 2009; Richard *et al.* 2015; Villemey *et al.* 2016; Baxter *et al.* 2017), none of these do so over more than a single time point. Many of these studies are also not directly comparable to this one, nor to each other owing to the differing analytical techniques used. However, those by Richard *et al.* (2015) and Villemey *et al.* (2016) use the same microsatellite markers as this study, allowing more reliable comparisons to be made.

By sampling *M. jurtina* individuals yearly from the same sites we have produced the first temporal dataset of genetic diversity for this species. Our results show that *M. jurtina* in the south of England appear to belong to a single, large population, with high levels of gene flow and low levels of genetic differentiation. These results are supported by the results of the above studies, which report similar findings in a range of different *M. jurtina* populations. The results also support the findings of Habel *et al.* (2013), in that the levels of genetic diversity and differentiation reported here are in line with those of other generalist Lepidoptera species.

We also show that the genetic diversity of this species in the south of England is relatively stable, with little change in measures of genetic diversity or differentiation over time. However, these results should be interpreted with caution owing to a number of limitations in the study (see below).

3.5.1 Population genetic diversity

Allelic richness was consistent across all three datasets used in this study. The mean number of alleles per locus (15.58) was considerably higher than in Schmitt *et al.* (2005) (2.65), and also higher than those of all 22 grassland Lepidoptera species reported in Habel *et al.* (2013). It should however be noted that all 22 of these studies, including Schmitt *et al.* (2005) were calculated using allozyme polymorphisms, not microsatellites and are therefore not directly comparable. It should also be noted that of the 22 species investigated by Habel *et al.* (2013) *M. jurtina* had the third highest mean number of alleles per locus (2.65), behind only *Polyommatus icarus* (common blue) (2.94) and *Coenonympha pamphilus* (small heath) (3.17).

When compared to other microsatellite analyses the results from this study are lower than those reported in Villemey *et al.* (2016), but similar to those in Richard *et al.* (2015). The differences between this study compared to Villemey *et al.* (2016) could be due to differences in sample sizes (1681 in Villemey *et al.* (2016) vs <300 in this study). Sample sizes in Richard *et al.* (2015) were lower still (96), yet the number of alleles per locus were similar to this study. A potential reason for this may be higher levels of genetic diversity within the French *M. jurtina* populations compared to those in the UK, as is often the case with island and mainland populations (Frankham 1997).

The levels of allelic richness recorded in microsatellite studies in *M. jurtina* are higher than in other satyrine species e.g. *Coenonympha hero* (Cassel 2002), *Bicyclus anynana* (Van't Hof *et al.* 2005), *Erebia palarica* (Vila *et al.* 2009b). They are also higher than the specialist chalk grassland lycaenids *Polyommatus bellargus (Adonis blue)* and *P. coridon* (chalkhill blue) (O'Connor 2014), with

which *M. jurtina* often shares habitats. Levels of allelic richness appear to be similar to those found in the meta-population forming nymphalid species *Melitaea cinxia* (Sarhan 2006). These results should be interpreted with care due to the relatively low number of samples used in the estimation of allelic richness as a result of the rarefaction process.

Excluding the locus Mj4870 all loci displayed very high levels of heterozygosity, with a high mean level of heterozygosity across all loci. This is unsurprising as previous studies using allozymes have found higher levels of heterozygosity in *M. jurtina* compared not only to other insects (Graur 1985), but to many other lepidopteran species (Schmitt *et al.* 2005). These high levels of heterozygosity and genetic diversity are found in other widespread, common, generalist butterflies (Schmitt *et al.* 2003; Habel *et al.* 2005, 2009).

Observed heterozygosity was lower than expected in the majority of cases across each dataset, however the levels of heterozygote deficit or excess did not represent significant departures from HWE at any locus. With the exception of locus Mj4870, the results from this study are again similar to those reported in Richard *et al.* (2015) and Villemey *et al.* (2016). There is some significant, but low-level inbreeding occurring within populations, indicated by significant positive F_{IS} scores, but this is unlikely to be the cause of the lower heterozygosity and greater heterozygote deficit seen in locus Mj4870. Inbreeding, as well as other biological aspects such as strong selection pressures or the Wahlund effect (whereby reductions in observed heterozygosity occur due to cryptic population structuring (Wahlund 1928; Dharmarajan *et al.* 2013)) are causes of heterozygote deficit. These generally affect all loci used, rather than a single locus (Dakin & Avise 2004; Selkoe & Toonen 2006), although this is not always the case (Dharmarajan *et al.* 2013). A more locus specific reason for heterozygote deficit would be the presence of null alleles (Selkoe & Toonen 2006). Whilst these are present for Mj4870 at a higher rate than at the other loci in this study, the mean frequency of null alleles remains below the frequency of 0.2, above which analytical problems can occur (Dakin & Avise 2004). As described in the methods the analysis was repeated with the locus Mj4870 removed, but this had no significant effect on the results. This combined with the non-significant departure from HWE meant that the locus was retained.

All estimates of effective population size were either infinite, or included confidence intervals that reached infinity, meaning "there was no evidence for variation in the genetic characteristic caused by genetic drift due to a finite number of parents" (Do *et al.* 2014), i.e. there was insufficient signal in the data to accurately predict effective population sizes. This is perhaps not surprising as methods to estimate contemporary N_e are less effective with large population sizes. In studies such as this with relatively small sample sizes and low numbers of loci it is very difficult to gain useful information regarding N_e of large populations ($n > 1000$) (Waples & Do 2010). To gain enough data for a more accurate estimation, a greater number of samples would have to be collected and analysed. However, as *M. jurtina* have large census population sizes, the increased sampling required per site would likely be impractical. N_e allows an understanding of the rate of genetic diversity loss after genetic drift has occurred (Freeland *et al.* 2011), therefore gaining a greater understanding of N_e would be highly beneficial for monitoring *M. jurtina* genetic diversity. Fluctuations in census population sizes, as recorded by population monitoring schemes, are key factors influencing N_e and

the ability to predict how N_e may be affected by severe fluctuations in census population sizes, e.g. during a drought year, would be of great interest. Whilst this would be highly beneficial, the practicality of gaining such information renders it unfeasible using the described methods.

3.5.2 Population genetic differentiation

F_{ST} scores for each locus and averaged across all loci were extremely low (<0.01), and in all cases were not significantly different from zero. Additionally R_{ST} scores (an analogous method for estimating genetic structure better suited to microsatellites (Slatkin 1995; Meirmans & Hedrick 2011) were also extremely low (means <0.002). These values indicate very little genetic differentiation between loci, suggesting high levels of gene flow between populations. Pairwise F_{ST} scores between pairs of populations resulted in mean F_{ST} values >0.01 between all sites. Again, indicating very low genetic differentiation between sites, suggesting high levels of gene flow. These results are in alignment with these previous microsatellite studies showing negligible differentiation between sites (F_{ST} < 0.05) (Richard *et al.* 2015). Additionally previous allozyme studies also report low F_{ST} scores across similar size areas (Goulson 1993a; Wood & Pullin 2002).

The lack of genetic differentiation suggested by the pairwise F_{ST} scores is supported by the results of the STRUCTURE analyses, which were unable to partition the sampled populations into anything other than single clusters. These results provide further evidence to suggest that at the scale of the study all individuals belong to a single, large population of *M. jurtina* in the south of England. Again, these results should be treated with caution owing to the low number of microsatellite loci employed in the analysis. Where too few loci are used it can be impossible for the clustering algorithms to result in anything other than a lack of evidence for population structuring. It has also been shown that even relatively large numbers of microsatellite loci can result in incorrect population structure inference (Orozco-Terwengel *et al.* 2011). However, as the loci employed were highly polymorphic the results can be interpreted with some confidence despite the low numbers used. Nevertheless, it is important to interpret the results from the STRUCTURE analysis alongside more traditional measures of population structuring such as F_{ST}.

Using Slatkin's (1985) private allele method, all estimates of Nm are greater than 2, a threshold which the author defined as indicating high levels of gene flow. The levels of migration and gene flow estimated in this study are well in excess of the 'one migrant per generation' rule of thumb that states "the appropriate level of gene flow for maintaining genetic diversity and preventing inbreeding depression in fragmented populations is one migrant individual per local population per generation" (Wang 2004). Whilst informative and lending further evidence to the idea of a single population of *M. jurtina* in the south of England, these estimates of Nm are less informative than Bayesian methods which estimate contemporary migration rates between populations, as have been used in similar studies of genetic diversity e.g. Vanhala *et al.* (2014). Such methods were not viable in this study due to a lack of resolution in the data that prevented MCMC convergence. The lack of convergence was likely due again to the low number of microsatellites available in this study.

3.5.3 Isolation by distance and the effect of landscape on gene flow

We found no evidence of IBD between pairs of sites. This is consistent with the results of Richard *et al.* (2015) who also found a negligible effect of IBD at a much larger scale. Furthermore, there was no evidence to suggest that the similarity of habitats from which populations were collected had any effect on the levels of genetic diversity.

Additionally, when surrounding habitats were taken into account there was no clustering of sites into groups based upon habitat irrespective of distance as would be expected if landscape were having a strong effect upon gene flow (Fig. 3.5). Site pairs geographically closer together with high proportions of good surrounding habitat did not exhibit higher levels of F_{ST} than distant sites with poor surrounding habitat. Evidence of this would have been site pairs with greater proportions of good habitat (green points) plotted consistently higher on the y axis of Fig. 3.5 than site pairs with bad habitat (red points), regardless of their position along the x axis. What we see instead is an even spread of green points on the y axis and a clustering of green points towards the left of the x axis. The x axis clustering is an artefact of the locations of the sites chosen, with the better sites for *M. jurtina* occurring closer together in the Chiltern Hills.

At the scale of this study distance between sites, the similarity of sites and the amount of good surrounding habitat had little effect on levels of gene flow in *M. jurtina*. Landscape does have some effect on *M. jurtina* movement and gene flow as shown by Villemey *et al.* (2016), with grassland elements enhancing gene flow and arable and woodland areas inhibiting it. However, as no site in this study had less than 60% good habitat cover within a 10km area of its centroid, it is likely that the high levels of grassland between sites allow high levels of gene flow. An interesting point of further study would be to see if the levels of genetic diversity and differentiation are similar in areas with less grassland, such as in more urbanised, agricultural or forested areas.

The results from the landscape analysis are consistent with the other findings in this study, suggesting that, at the study scale, landscape features have minimal effect on geneflow with little evidence of either IBD or IBR. The initial results were inconclusive, with minimal differences in AICc and log-likelihood (LL) values between the null, IBD and binary surface IBR models. However, bootstrapping determined that IBD was a better fit than IBR. The null model was not included in the bootstrapping, however there was no evidence of IBD found using Mantel tests. This, as well as the null model being a better fitting model than IBD according to both AICc and LL in the initial ResistanceGA analysis, is strong evidence to suggest that the landscape has little effect on geneflow at the study scale. These results provide further evidence that populations within the study area can be considered as one large population with high levels of geneflow. It should be noted that as with the other analyses in this study, the low number of microsatellites used may not provide sufficient resolution to determine landscape barriers. Additionally, sample size has a large effect on the error rate with ResistanceGA, with confidence deteriorating when fewer than 25 samples points are used (Winiarski *et al.* 2020). Although a total of 287 individuals were used in the analysis, they were collected from 15 populations, resulting in 15 data points. This may explain why Villemey et al. (2016) had differing results, with grassland areas enhancing geneflow in their study. It is likely that

at our study scale no landscape features had any significant effect on geneflow, and that if any fine scale resistance exists it cannot be determined due to the resolution of the data.

3.5.4 Changes to genetic diversity over time

The results from the two single years 2012 and 2017 are very consistent with each other indicating little change in the genetic diversity of the species between those two years. Minor differences occur, but the significance of all results is the same for both years. The results for the three sites from 2012 to 2019 inclusive are also consistent between years, with no major changes in genetic diversity of differentiation over time. A small number of site pair combinations display F_{ST} values with significant scores, however the levels of differentiation are still below the 0.05 threshold considered to indicated moderate genetic differentiation (Freeland *et al.* 2011). Aside from the small number of loci, an additional limitation needs to be taken into account when addressing the results of the time series data. As only three sites were used in the analysis it is likely that the resolution of data is poorer than for 2012 and 2017 for which 14 and 15 sites were used respectively. This is most noticeable in the estimation of migration rates. When using all 15 sites in 2017 Nm is estimated at 13.141. When using just three this drops considerably to 4.367. This is still indicative of high levels of gene flow i.e. greater than 2 (Slatkin 1985), but heavily implies that the reduced amount of data available in the time series data is affecting the analysis.

3.5.5 Limitations and further work

As mentioned throughout this discussion some limitations need to be considered when interpreting the outputs of this study. Firstly, the length of time for which the study has been running is not long enough to begin to properly analyse any changes over time to the genetic diversity of the species. Further monitoring is required over a longer period before statistically valid tests can be carried out on the time series data.

Secondly, the number of individuals sampled at each population may not have been sufficient to fully detect the levels of genetic diversity present. In microsatellite studies sampling between 20 and 30 individuals from a population is thought to be enough to accurately estimate allele frequencies, with further sampling effort yielding little further resolution (Pruett & Winker 2008; Hale *et al.* 2012). In this study attempts were made to sample 20 individuals from each site, each year. However, this was not always possible due to adverse weather conditions in some years resulting in fewer samples flying during the collection period. Sample sizes were further reduced by analytical problems such as a lack of DNA amplification. Although 20 samples would have been at the low end of sufficient population sizes, the lowest sample size in this study was reduced to just nine individuals. It is likely therefore that not all the allele frequencies estimated are completely accurate due to the low number of samples from some populations.

The major limitation of this study is the low number of microsatellite loci used. Whereas increasing sample sizes in microsatellite studies is only informative up to a point, increasing the number of microsatellites greatly increases the power and reliability of the results (Selkoe & Toonen 2006; Bruford *et al.* 2015). The problem with this is that microsatellite development can be costly

as well as resource and time intensive (Glenn & Schable 2005; Selkoe & Toonen 2006; Orozco-Terwengel *et al.* 2011).

As highlighted by the attempted novel primer development in Appendix B, microsatellite characterisation in Lepidoptera is particularly challenging for a number of reasons. Firstly the frequency of microsatellite repeat motifs can be lower in many Lepidoptera than in other species (Meglécz & Solignac 1998; Zhang 2004; Van't Hof *et al.* 2007). Secondly, microsatellite flanking regions in Lepidoptera experience high mutation rates, leading to null alleles and amplification problems (Bogdanowicz *et al.* 1997; Keyghobadi *et al.* 2002; Harper *et al.* 2003; Sinama *et al.* 2011). Flanking regions in Lepidoptera are also characteristically repetitive (Zhang 2004), with many microsatellite motifs sharing similar flanking regions (Mikheyev *et al.* 2010), or with microsatellites occurring within minisatellite repeats (Meglécz *et al.* 2004). A final reason for difficulties in Lepidoptera microsatellite isolation are their association with transposable elements (Tay *et al.* 2010), which are able to multiply and insert themselves throughout the genome (Coates *et al.* 2010) and can lead to many microsatellites having similar flanking regions.

The resolution and resulting conclusions that could be made from this study would have been greatly improved by an increased number of microsatellite loci. Only six of the 15 microsatellite loci isolated by Richard *et al.* (2015) were used in this study due to amplification issues, sex-linkage and null alleles in the other nine. This was also the case with Villemey *et al.* (2016), who were also only able to use six of the 15 loci (five of which were used in this study and one (Mj0008) that was not). Despite efforts to isolate more loci (see appendix B) only loci isolated by Richard *et al.* (2015) were used in this study. Isolation of more microsatellite loci and then analysis of the samples used in this study would be an interesting area of further investigation as it would show whether the conclusions drawn here are valid or if there is some structuring of populations that was missed due to the low resolution of the data. Isolation of microsatellite markers may be considerably easier in future studies owing to the publication of a draft genome sequence (Singh *et al.* 2020) .

Another logical area of further investigation would be to determine whether the differences in measures of genetic diversity and differentiation between this study and those of Richard *et al.* (2015) and Villemey *et al.* (2016) are due to inherent differences between the British and mainland European populations. Analysing additional samples from France and also from a number of different sites across Europe would give an indication of the genetic diversity and differentiation of the species across a much larger area.

The most important next step to be taken would be to continue to monitor the genetic diversity of this species across the south of England. Using the same sample sites and markers offers a rare opportunity to continuously monitor the genetic diversity of a wild species, as desired by the convention on biological diversity (Convention on Biological Diversity 2011a).

3.5.6 Conclusions and implications

The results of this study confirm those previously investigating measures of genetic diversity in *M. jurtina*. Using the available loci individuals in the south of England appear to belong to a single, large population, with very high levels of genetic diversity and low levels of genetic differentiation,

as is seen in other generalist Lepidoptera species with continuous populations. Whilst the distance any individual can travel is limited, gene flow over generations does not appear to be, with the population displaying properties of panmixia.

This study has also shown a potential method for monitoring the genetic diversity of a wild species over time using microsatellite markers. The standard laboratory and analytical techniques employed in this study mean that monitoring other species would be possible using the same methods. Owing to the problems of microsatellite isolation in Lepidoptera, the monitoring of the genetic diversity of other species may in fact be easier than in this study, providing researchers are able to commit time and resources to a long-term temporal data set. Such monitoring schemes are essential in order to prevent the future erosion and loss of genetic diversity in wild species.

Chapter 4. Quantifying the genetic diversity of *Maniola jurtina* across Europe

4.1 Abstract

The meadow brown butterfly, *Maniola jurtina,* is a well-studied species with much known about its life history, phenology and evolutionary history. Despite many investigations into the post glacial distribution of *M. jurtina* across Europe using molecular markers, no studies have quantified the contemporary levels of genetic diversity of the species across the continent. Using microsatellite markers, we investigate the genetic diversity of *M. jurtina* across Europe, with samples collected from 39 sites across 11 European countries, over 2500km. Overall, we find high levels of genetic diversity across the continent and low levels of genetic differentiation between populations at a lower spatial scale. We find some population structuring between populations in the United Kingdom and the European mainland, suggesting restricted geneflow, but not complete isolation, between the two. These results support previous studies into *M. jurtina* genetic diversity and allow a greater understanding of this already well studied species.

4.2 Introduction

4.2.1 Genetic diversity

Genetic diversity is important for the long-term persistence of species (Booy *et al.* 2000). High levels of genetic diversity result in greater levels of phenotypic variation, which in turn allow for populations to better adapt to environmental change (Boettcher *et al.* 2010; McGill *et al.* 2015; Bruford *et al.* 2017). In contrast, populations with low levels of genetic diversity can display reduced reproductive fitness, which can lead to subsequent population declines (Frankham *et al.* 2011). The maintenance of genetic diversity should therefore be a priority for conservation projects (Schmitt & Hewitt 2004), however this has so far been overlooked for many wild species. In order to maintain genetic diversity, it is important to first develop an understanding of how genetically diverse a species actually is. To do so typically requires the use of whole genome sequencing, molecular markers e.g. microsatellites, Amplified Fragment Length Polymorphisms (AFLPs), Restriction site associated DNA (RAD) sequencing, DNA sequences, or protein polymorphisms (Hughes *et al.* 2008) to measure quantitative characteristics such as allelic richness or heterozygosity. These can then be compared within and between populations, giving a measure of the differences between them (Booy *et al.* 2000).

4.2.2 The meadow brown butterfly, *Maniola jurtina*

The meadow brown butterfly, *Maniola jurtina* (Linnaeus 1758), is a common satyrine species. Although the most abundant butterfly in the UK (Fox *et al.* 2006) and across Europe more widely, *M. jurtina* has experienced declines across the continent over the past 20 years (Van Swaay *et al.* 2013), most severely in Finland, Malta, Luxembourg, Sweden and Russia (Asher *et al.* 2001). These declines are largely a result of increasing agricultural intensification (Delattre *et al.* 2010) and the resultant replacement of hay meadows with arable crops (Asher *et al.* 2001). As discussed in Chapter 3, the genetic diversity of *M. jurtina* is relatively high across the study area in the south of England, with all individuals recorded belonging to one large population, apparently at panmixia. The levels of genetic diversity appear to remain stable over the eight years of the study.

4.2.3 *Maniola jurtina* phylogeography

Whilst there are no previous temporal studies into the contemporary genetic diversity of *M. jurtina* (see Chapter 3), the genetic diversity of the species across Europe at a spatial scale and its historical distribution have received a great deal more attention (Tauber 1970; Thomson 1973, 1987; Goulson 1993a; Schmitt *et al.* 2005; Habel *et al.* 2009; Dapporto *et al.* 2011; Richard *et al.* 2015; Villemey *et al.* 2016; Baxter *et al.* 2017).

Initial work on the subject by Tauber (1970), using morphological measurements and paleo-ecological theory suggested that *M. jurtina* originated in Africa and spread into Europe by two dispersal routes, an eastern and a western one. Although the idea of an African origin was not widely accepted, subsequent morphological (Thomson 1973) and allozyme studies (Thomson 1987) have confirmed the presence of two distinct lineages of *M. jurtina*: a western Atlantic-Mediterranean lineage (*Maniola jurtina jurtina*) and an eastern-Mediterranean-Asian lineage (*Maniola jurtina janira*),

with a hybridisation zone between the two in central Europe (Thomson 1987). Further allozyme work by Schmitt *et al.* (2005) suggested that the two lineages diverged around 40,000 years ago, at the beginning of the last glacial maximum, with populations surviving in mainland glacial refugia in the Iberian, Italian and Balkan peninsulas.

A subsequent allozyme study by Habel *et al.* (2009) found three distinct genetic groupings across Europe: a western European grouping, a central/eastern European grouping and an Italian/Maghreb grouping. Their findings also led to the conclusion that Africa was in fact colonised by the Italian lineage: *M. jurtina junira* (Habel *et al.* 2009)*,* contrary to Tauber (1970). Using genital morphology measurements Dapporto *et al.* (2009) found contrasting evidence to Habel *et al.* (2009), supporting only the two original lineages. Differences in the genital morphologies between island and mainland populations also led to the conclusion that Mediterranean islands also acted as a glacial refuge as well as the mainland peninsulas (Dapporto *et al.* 2009).

Combining allozyme and morphometric data Dapporto *et al.* (2011) determined that recent, postglacial gene flow, as a result of range contractions and expansions had caused the discrepancies between the two previous studies. Further work by Dapporto & Bruschini (2012) suggested that the entire Mediterranean was originally occupied by the western lineage *M. jurtina jurtina* and subsequently colonised by the eastern lineage from the Balkans.

A more recent study by Kreuzinger *et al.* (2015) using both nuclear and mitochondrial genetic markers found contrasting evidence again, not supporting the findings of Dapporto *et al.* (2011) and Dapporto & Bruschini (2012). In their study the authors investigated all seven species within the *Maniola* genus and determined them all to belong to a single, highly variable "super species" *Maniola jurtina.* They also found only moderate genetic differentiation between two lineages of *M. jurtina* (when including all seven *Maniola* species) and refer back to the original "out of Africa" hypothesis suggested by Tauber (1970) almost fifty years previously (Kreuzinger *et al.,* 2015).

Clearly further research is required to fully determine the phylogeography of *M. jurtina*. However, despite no clear-cut conclusions these studies do provide valuable information regarding the contemporary genetic diversity of the species across Europe. Both Schmitt *et al.* (2005) and Habel *et al.* (2009) found very high levels of genetic diversity and allelic richness within *M. jurtina* populations, considerably higher than those of other satyrine butterflies. They attribute these high diversities to the high abundance of *M. jurtina* and its formation of near continuous populations (Schmitt *et al.* 2005).

4.2.4 *Maniola jurtina* and open water barriers

An interesting point made throughout these studies is that *M. jurtina* repeatedly colonised islands along the Mediterranean coast, with suggestions that this may have occurred more frequently during periods when sea levels were lower (Dapporto *et al.* 2011). This cannot be disputed as *M. jurtina* are found on many islands around Europe, including the Isles of Scilly off the coast of England. Interestingly, mark-release-recapture studies by Dowdeswell *et al.*, (1949) on the Isles of Scilly, backed up by recent AFLP marker analysis carried out on the same islands by Baxter *et al.*

(2017), suggest that movement between islands does not occur and that stretches of open water act as barriers to *M. jurtina* movement. This contrasts not only to the historical island colonisations of *M. jurtina,* but also to direct observations made by Dennis & Shreeve (1996) of many individuals crossing large expanses of hostile habitat including open water.

In general *M. jurtina* are relatively sedentary, with adults typically moving within a radius of 500m. However, longer movement distances up to 2.1km have been recorded via mark-release-recapture (Schneider *et al.* 2003). Whilst water may act as a barrier to contemporary gene flow, it is clear that sea crossings are possible, although they may only occur rarely. The two differing types of marker used in these studies would go some way to explaining the discrepancies in their results. Allozymes have slow mutation rates and relatively low levels of polymorphisms (Estoup *et al.* 1998). This makes them useful for reconstructing phylogenies, but less so for interpreting current population structuring (Wiens 2000). In contrast the dominant and highly polymorphic nature of AFLPs makes them more suited to studies investigating recent genetic events, but unable to resolve distant evolutionary events (Freeland *et al.* 2011). Therefore it seems likely that *M. jurtina* on the Isles of Scilly belong to the western-Atlantic-Mediterranean lineage as a result of colonisation events via sea crossings, but such sea crossings are rare and can result in genetically distinct populations through isolation, as demonstrated by the results of Baxter *et al.* (2017).

4.2.5 Chapter aims

In this study we use a third form of genetic marker: microsatellites, to investigate the genetic diversity of *M. jurtina* across Europe and also determine whether distinct genetic clusters are found across the continent. Like AFLPs, microsatellites are better suited to investigations of contemporary migration and genetic diversity than historical genetic events (Selkoe & Toonen 2006). Whilst microsatellites can be used for inferring historical events, their predictive power is greatly reduced when the number of generations increases over a few thousand (Robinson & Harris 1999). This is due to their high mutation rates, which can lead to size homoplasy i.e. where two same size alleles in different individuals appear to be the result of a shared ancestry, but are in fact the result of respective insertion and deletion mutations in two different sized alleles (Freeland *et al.* 2011). Therefore the choice of markers used mean that this study will not be a continuation of the investigations by Thomson, (1987), Schmitt *et al.* (2005), Habel *et al.* (2009), and Dapporto *et al.* (2011) into the post glacial recolonisation of *M. jurtina*, but will instead investigate the current genetic diversity of the species in populations across Europe and determine whether distinct genetic clusters have formed. This study will also shed some light onto the level of isolation that UK populations are subject to, when compared to those on mainland Europe. Should the results of Baxter *et al.* (2017) scale up to the UK as a whole it is anticipated that UK *M. jurtina* populations will display some genetic differentiation from those on the mainland as a result of reduced levels of migration across the English Channel.

4.3 Methods

4.3.1 Sample collection

A total of 810 *M. jurtina* samples were collected from 39 sites across 11 European countries in 2017 with sampling carried out by volunteers from 12 collaborating organisations (Table 4.1 & Figure 4.1). Additional samples were also collected for the years 2016 and 2018 but were not used in this analysis due to time and financial limitations. From each site 20 whole butterfly samples were collected when possible. All site locations were centred about UK or European butterfly monitoring scheme transects. At these sites volunteers record the number of butterflies seen over a number of recording dates throughout the summer, following the methodology outlined in Pollard & Yates (1993) and summarised in Van Swaay *et al.* (2008).

Table 4.1 Site codes for 39 sites (38 with coordinates) across Europe from which *M. jurtina* samples were collected in 2017.

Sample Site	Site Code	Country	Lat	Long	Number of Samples	Collaborating Organisations
Aston Rowant (north)	UK-ARN	England	51.6581	-0.94608	21	University of Reading UKCEH Forest Research
Aston Rowant (south)	UK-ARS	England	51.6582	-0.96405	17	
Aston Upthorpe Downs	UK-AU	England	51.550	1.215	14	
Bowdown	UK-B	England	51.382	-1.271	17	
Coombe Hill	UK-CH	England	51.753	-0.774	20	
Crabtree	UK-C	England	51.262	-1.051	20	
Dancersend	UK-D	England	51.777	-0.697	15	
Howbery Park	UK-HP	England	51.606	-1.113	20	
Lardon Chase	UK-LC	England	51.524	-1.155	20	
Little Wittenham	UK-LW	England	51.627	-1.190	20	
Moor Copse Nature Reserve	UK-MC	England	51.462	-1.085	16	
Pamber Forest	UK-PF	England	51.345	-1.118	37	
Swyncombe Downs	UK-SD	England	51.618	-1.026	15	
The Crong	UK-TC	England	51.771	-0.691	15	
Wytham Woods	UK-WW	England	51.787	-1.320	20	
Santa Susanna	SP	Spain	41.739	2.392	15	Granoller's Natural Sciences Museum
Issancourt et Rumel	FR-A	France	49.762	4.811	20	University of Reading UKCEH
Demange aux Eaux	FR-B	France	48.592	5.465	20	
Nogent sur Vernisson	FR-C	France	47.845	2.764	18	
Gommecourt	FR-D	France	49.067	1.602	20	
Oosterlo	B-A	Belgium	51.203	2.955	20	Research Institute for Nature and Forest (INBO)
Essenbeek	B-B	Belgium	50.718	4.257	6	
Oostende	B-C	Belgium	51.104	4.979	18	
Wageningen Campus	ND-A	Netherlands	51.984	5.666	15	De Vlinderstichting
Moerputten	ND-B	Netherlands	51.685	5.263	16	
Lamadelaine	L	Luxembourg	49.542	5.871	18	Luxembourg Institute of Science and Technology

Sample Site	Site Code	Country	Lat	Long	Number of Samples	Collaborating Organisations
Himmelried	SZ	Switzerland	47.422	7.588	18	Hintermann & Weber AG
Hestehavevej	DK-A	Denmark	56.295	10.465	19	Aarhus University
NA	DK-B	Denmark	NA	NA	20	
Halle	G-A	Germany	51.491	11.931	23	Helmholtz Centre for Environmental Research
Friedeburg	G-B	Germany	51.623	11.727	11	
Greifenhagen	G-C	Germany	51.625	11.443	20	
Siptenfelde	G-D	Germany	51.650	11.069	22	
Harsleben	G-E	Germany	51.839	11.060	20	
Rökepipan	SW	Sweden	55.672	13.361	15	Lund University
Kodijärve	E-A	Estonia	58.213	26.640	19	University of Tartu
Jõhvi	E-B	Estonia	58.511	26.774	20	
Rõhu	E-C	Estonia	58.326	26.519	22	
Lalsi	E-D	Estonia	58.442	26.085	20	

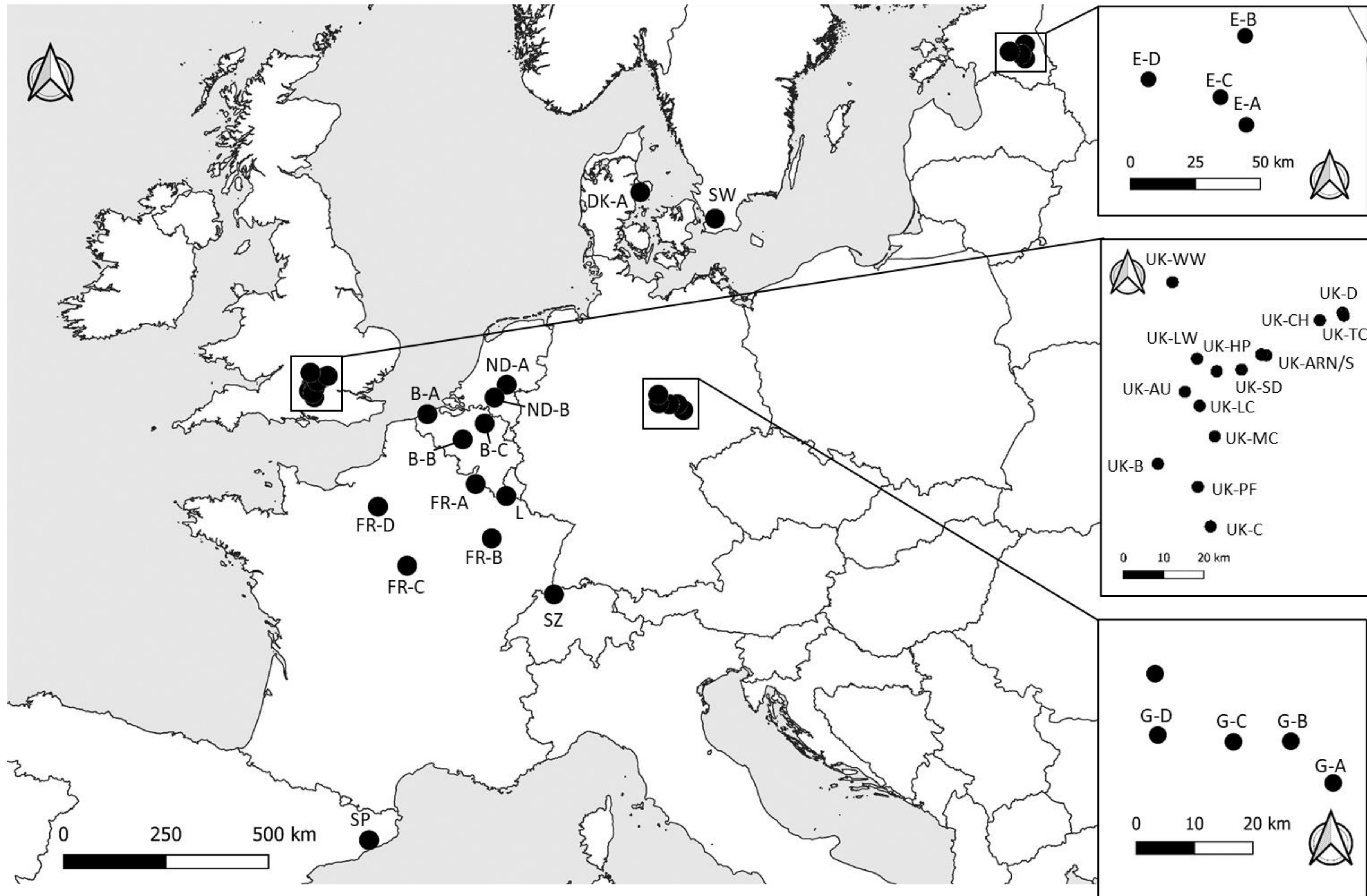


Figure 4.1 Site locations of 38 sites across Europe from which samples were collected in 2017.

4.3.2 DNA isolation and microsatellite analysis

DNA was extracted from a single leg from each individual, following the manufacturer's guidelines for insects described in the 'MicroGEM Quick-Start Guide DNA Extraction Using prepGEM Universal'. Genotyping was carried out using six microsatellite markers isolated by Richard *et al.* (2015) (see Chapter 3). For further information regarding novel primer development and evaluation, and pre-existing primer evaluation please see Appendices B and C respectively.

Polymerase Chain Reactions (PCRs) were conducted in a total volume of 11µL containing 1µL template DNA, 6.25µL QIAGEN multiplex PCR master mix (3mM $MgCl_2$), 0.625µL tagged forward primer (5' labelled with FAM, NED or PET), 0.625µL reverse primer, 1.25µL QIAGEN Q solution, 2.25µL RNase-free water. Further details, including primer concentrations, can be found in Appendix C. PCRs were carried out in an Eppendorf Mastercycler nexus eco with an initial denaturation for 15:00 at 95°C, followed by 40 cycles of denaturation for 00:30 at 94°C, annealing for 01:30 at 56°C, extension for 01:00 at 72°C, and a final extension phase for 10:00 at 72°C. All PCR reactions underwent fragment analysis using an Applied Biosystems 3730 DNA Analyser. The results were then scored using GeneMarker® version 1.5 (SoftGenetics). Any individual for which there were more than two loci with missing data were removed from the analysis (5.962% of samples). Further samples were then removed to ensure that all populations had no greater than 5% missing data per locus. This was done so that no loci were dropped in the analysis due to insufficient data. As a result, a total of 722 samples were included in the analyses.

4.3.3 Statistical analysis

All analyses were carried out first using only the mainland European sites (24 sites), then including three of the fifteen available UK sites (27 sites) and then finally using all fifteen UK sites (39 sites). This was done to determine what effect the large number of UK sites compared to other individual countries had on the overall analysis. The inclusion of all UK sites had minimal effect on the outputs of the majority of tests, with the exception of pairwise F_{ST}. As a result, all UK sites were retained in the analysis. Test results using only the European sites and including only three UK sites can be found for comparison in Appendix D.

4.3.3.1 Microsatellites

Linkage disequilibrium among pairs of loci was tested for in Genepop v4.7 (Rousset 2008), as were null allele frequencies. The observed and expected heterozygosities (H_o and H_e respectively) across the whole dataset, for individual loci and averaged across all loci, were calculated using the R package PopGenReport (Adamack & Gruber 2014; Gruber & Adamack 2015), as were deviations from the Hardy Weinberg equilibrium (HWE).

4.3.3.2 Population genetic diversity, divergence and structure

For each locus Wright's F statistics (Wright 1965) were calculated across all sampling locations using Genepop. An estimate of R_{ST} was calculated in FSTAT v. 2.9.4 (Goudet 1994). The rarefaction procedure was used to estimate allelic richness (A_r), based upon a minimum sample size of six

diploid individuals. The number of private alleles (Ap) was calculated using PopGenReport, as were the expected (H_e) and observed (H_o) heterozygosity for each locus and the percentage differences between H_e and H_o. H_e for each site and across all sites were calculated using Arlequin v 3.5.2.2 (Excoffier & Lischer 2010). To compare if there were differences in the genetic diversity of *M. jurtina* in the UK compared to those on mainland Europe mean H_e and A_r were compared using two sample t-tests.

Effective population sizes (N_e) were estimated using two methods, the bias corrected linkage disequilibrium method (Hill 1981; Waples 2006; Waples & Do 2010), and the heterozygote excess method (Pudovkin *et al.* 1996; Zhdanova & Pudovkin 2008) both implemented in NeEstimator V2 (Do *et al.* 2014).

Population structure was investigated using STRUCTURE v.2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2003, 2007; Hubisz *et al.* 2009). For all models in this study the parameter set used an admixture model and correlated allele frequencies with a 100,000 burn-in and 1,000,000 MCMC replications per chain. For each possible value of K, twenty chains were run. The most likely value of K within the sample sets was estimated using the programme STRUCTURE Harvester (Earl & VonHoldt 2012) which estimates an *ad hoc* statistic, Delta K, to determine the most likely value of K (Evanno *et al.* 2005).

Contemporary migration rates between sites were estimated using BayesAss v.3.04 (Wilson & Rannala 2003). A burn-in of 10,000 and 10,000,000 MCMC repetitions were used with a thinning interval of 800, and mixing parameters of 0.8 for migration rates, 0.4 for allele frequencies and 0.8 for inbreeding coefficients. MCMC convergence was monitored for each set of samples by plotting the profile of the likelihood and prior values over time using Tracer v1.7.1 (Rambaut *et al.* 2018). For each run the MCMC did not reach convergence. The number of MCMC repetitions was increased to 100,000,000, however convergence still did not occur and as a result BayesAss was abandoned from the study. Instead Slatkin's private allele method for estimating migrants (Slatkin 1985) was calculated in Genepop.

For measures of genetic distance individuals were pooled on a per site basis. Weir and Cockerham's pairwise F_{ST} values were calculated using FSTAT. Estimates of isolation by distance (IBD) were calculated by plotting pairwise F_{ST} values against (log) Euclidean distances between sites, with a Mantel R test using the *mantel* function from the R package Ecodist (Goslee & Urban 2007). Owing to the unevenness of the distribution of distances between sites (distance between sites: μ = 728km, min = 8km , max = 2525km), IBD estimates were also calculated with the Estonian sites removed and again with the Spanish site removed, to determine whether the large distances between these site(s) was affecting the analysis (distances between sites with Estonian sites removed: μ = 501km, min = 8km , max = 1742km, distances between sites with Spanish site removed: μ = 698km, min = 8km , max = 1979km).

4.3.3.3 Population genetic networks, barriers and bottlenecks

A standard method of detecting genetic barriers among populations is Monmonier's maximum difference algorithm (Monmonier 1973). This method uses geographical coordinates and connects

populations via Delauney triangulation with a pairwise F_{ST} matrix, resulting in the assignment of hypothetical spatial genetic boundaries between populations (Manni *et al.* 2004). However, although Monmonier's algorithm is an effective tool for investigating barriers to gene flow, it relies upon the assumptions that samples have different spatial locations and that spatial locations are equally spaced. As the data in this study violates both of these assumptions an alternative method of visualising potential barriers was required.

Instead of Monmonier's algorithm a spatial Principal Component Analysis (sPCA) was carried out to look for spatial variations within the data using the *spca* function in the R packaged adegenet *v2.1.2* (Jombart 2008). Unlike Monmonier's algorithm the sPCA is able to use allele frequency data of populations to investigate spatial patterns of genetic variability (Jombart *et al.* 2008). A spatial proximity network was built using the "K nearest neighbors" method. The principal component to interpret was determined by decomposing the genetic diversity from the spatial autocorrelation, as in Gagnaire *et al.* (2012).

A graph-based network theory analysis was used to further investigate gene flow and connectivity using the programme EDENetworks (Kivelä *et al.* 2015). All 39 populations were plotted as nodes in a weighted network connected by edges. Network analysis estimates the betweenness centrality (the fraction of all shortest paths going through a node), as an indication of gene flow, and the number of connections between nodes (Smith *et al.* 2015). Node size was proportional to the betweenness centrality and edges were weighted by F_{ST}. Firstly a Minimum Spanning Network (MSN) was plotted, representing the shortest network of edges to connect all populations, with the minimum total cost (Kivelä *et al.* 2015) i.e. the arrangement of edges to connect nodes that results in the lowest total F_{ST}. Due to a stochastic component in the algorithm used in EDENetworks the results of each visualisation can vary for the same network (Kivelä *et al.* 2015). In order to ensure that the MSN shown was representative of the results ten MSNs were calculated to test alternate network shapes. Secondly, a Threshold Network (TN) was plotted. For the TN all nodes were joined to every other node in connected network. A percolation threshold was assigned based upon both the relative size of the largest component in the network and the susceptibility score (see Kivelä *et al.* 2015). The percolation threshold indicates the point at which edge removal results in the fragmentation of the "giant component" of the network into smaller components, however, small, disconnected fragments (e.g. single nodes) can appear before this point. Edges with weights below the threshold are removed, retaining only the most important edges in the network (Kivelä *et al.* 2015). Automatic thresholding in EDENetworks was carried out in the first instance, with a threshold value set just below the estimated percolation threshold. This was then followed by manually lowering the threshold value to remove the weakest edge retained by the automatic thresholding. Edge colour and width were proportional to the genetic distances between nodes, with thicker, darker edges indicating lower genetic distances. The statistical significance of the betweenness centrality scores were tested for by bootstrapping, by re-sampling 85% of each population and randomising 1000 times.

An investigation into potential recent population bottlenecks or expansions was carried out using Bottleneck v. 1.2.02 (Piry *et al.* 1999). Bottleneck estimates heterozygosity assuming constant

population size (H_{eq}) via three models; infinite allele (IA), stepwise mutation (SM) and two-phase mutation (TPM, with SM set at 70 % and variance at 30). H_{eq} was then compared to H_o and significant excess or deficit was estimated via 1000 iterations with Sign, Standard Difference, and Wilcoxon test. H_o is classed as significantly different from H_{eq} if the majority of tests are significant. Potential recent reductions in population size were also investigated in Bottleneck using the shifted mode test as in Vanhala *et al.* (2014). For each locus alleles were grouped into frequency classes. Where no bottleneck has occurred allele distributions are expected to be L-shaped, with the rarest alleles forming the largest class. In the event of a bottleneck this distribution shifts as a result of the loss of the rarer alleles, indicating severe, recent declines in population size.

4.4 Results

4.4.1 Microsatellites, HWE and GE

There was no evidence of linkage disequilibrium between pairs of loci across the dataset (Table 4.2). With the exceptions of eight sites at Mj7232, and 19 sites at Mj4870 all site locus combinations had a low frequency (<0.2) of null alleles (Table 4.3). Across the dataset the microsatellites displayed a high level of variability (Table 4.4), with the highest values of H_O occurring at Mj5331 (0.862) and the lowest occurring at Mj4870 (0.237). All loci showed a non-significant level of heterozygote deficit across the dataset as a whole, the greatest occurring at Mj4870 with a deficit of 62.473 (Table 4.4). F_{ST} values were significantly greater than zero at Mj7232 and Mj4870. However, F_{IS} values were significantly greater than zero at four of the six loci (Table 4.4), indicating potential deviations from HWE. The high number of null alleles, combined with the low H_O and heterozygote deficit, led to the analyses being repeated with the removal of the locus Mj4870. Removal had little effect on the overall results, with the exception of Pairwise F_{ST} values which were reduced when using only five loci. Additionally, with only five loci STRUCTURE was unable to determine any clustering of populations (see below). As a result, the locus was included.

Table 4.2 Composite linkage disequilibrium test outputs for all locus pair combinations across all mainland European sites and all 15 UK sites

Locus pair		Chi^2	df	p-value
Mj7232	Mj5522	65.739	76	0.793
Mj7232	Mj0247	31.822	70	1.000
Mj5522	Mj0247	17.262	70	1.000
Mj7232	Mj4870	56.941	76	0.950
Mj5522	Mj4870	39.406	76	1.000
Mj0247	Mj4870	42.677	70	0.996
Mj7232	Mj7132	43.238	76	0.999
Mj5522	Mj7132	53.228	76	0.978
Mj0247	Mj7132	42.701	70	0.996
Mj4870	Mj7132	68.894	78	0.760
Mj7232	Mj5331	47.689	70	0.981
Mj5522	Mj5331	34.907	70	1.000
Mj0247	Mj5331	20.470	64	1.000
Mj4870	Mj5331	39.303	72	0.999
Mj7132	Mj5331	47.247	72	0.989

Table 4.3 Locus by populations estimated null allele frequencies for all European sites and 15 UK sites. Values in bold exceed 0.2 frequency of null alleles.

	UK-ARN	UK-ARS	UK-AU	UK-B	UK-C	UK-CH	UK-D	UK-HP	UK-LC	UK-LW	UK-MC	UK-PF	UK-SD	UK-TC	UK-WW	SP	FR-A	FR-B	FR-C	FR-D
Mj7232	0.028	0.145	0.000	0.137	0.079	0.096	0.022	0.032	0.086	0.000	0.171	0.050	0.048	0.013	0.032	0.041	0.181	**0.248**	0.029	0.058
Mj5522	0.000	0.055	0.048	0.000	0.000	0.000	0.030	0.055	0.006	0.013	0.029	0.001	0.076	0.074	0.000	0.129	0.000	0.040	0.017	0.000
Mj0247	0.042	0.030	0.000	0.085	0.018	0.000	0.041	0.000	0.000	0.037	0.069	0.064	0.068	0.000	0.048	0.070	0.055	0.198	0.081	0.158
Mj4870	0.075	0.187	0.000	0.000	0.000	0.064	0.086	0.003	0.159	0.151	0.000	0.080	0.150	0.189	0.128	0.167	**0.377**	**0.296**	0.173	0.076
Mj7132	0.022	0.065	0.000	0.045	0.000	0.000	0.091	0.000	0.000	0.028	0.000	0.000	0.047	0.002	0.000	0.088	0.000	0.000	0.116	0.017
Mj5331	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.030	0.028	0.028	0.027	0.000	0.000	0.000	0.000	0.084	0.129	0.000	0.000

Table 4.3 Continued

	B-A	B-B	B-C	ND-A	ND-B	L	SZ	DK-A	DK-B	G-A	G-B	G-C	G-D	G-E	SW	E-A	E-B	E-C	E-D	Mean
Mj7232	0.140	**0.242**	0.176	0.153	**0.340**	**0.305**	0.000	0.188	0.146	0.082	0.106	0.184	0.128	**0.200**	0.097	**0.261**	0.185	**0.366**	**0.426**	0.134
Mj5522	0.054	0.060	0.013	0.056	0.053	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.038	0.000	0.000	0.000	0.000	0.000	0.034	0.023
Mj0247	0.034	0.072	0.110	0.078	0.000	0.070	0.000	0.176	0.079	0.010	0.073	0.038	0.098	0.192	0.029	0.116	0.051	0.048	0.038	0.061
Mj4870	**0.227**	0.000	**0.354**	**0.315**	**0.369**	**0.284**	**0.262**	**0.287**	**0.206**	**0.312**	**0.248**	**0.371**	**0.297**	**0.372**	0.180	**0.384**	**0.221**	**0.321**	**0.358**	0.198
Mj7132	0.006	0.000	0.019	0.000	0.000	0.000	0.068	0.110	0.000	0.127	0.000	0.000	0.000	0.087	0.162	0.000	0.000	0.000	0.074	0.030
Mj5331	0.019	0.000	0.000	0.000	0.000	0.024	0.015	0.000	0.000	0.000	0.058	0.015	0.000	0.000	0.078	0.000	0.043	0.000	0.061	0.017

Table 4.4 Population-wide expected and observed heterozygosity, and percent difference ((E - O)/E*100), F_{IT}, F_{IS} F_{ST} and R_{ST} at each locus across all 24 mainland European sites and 15 UK sites. Bartlett's K-squared: 2.29, df = 1, p-value = 0.1303

Locus	No. of samples	No. of alleles	H_E	H_O	H_E vs H_O % difference	F_{IT} (p-value)	F_{ST} (p-value)	F_{IS} (p-value)	R_{ST}
Mj7232	720	16	0.837	0.738	11.892	0.120 (0.000)	0.054 (0.000)	0.070 (0.000)	0.084
Mj5522	709	13	0.845	0.794	6.046	0.062 (0.000)	0.012 (0.818)	0.051 (0.000)	0.022
Mj0247	715	33	0.940	0.779	17.104	0.171 (0.000)	0.017 (0.232)	0.156 (0.000)	0.031
Mj4870	712	8	0.632	0.237	62.473	0.630 (0.000)	0.102 (0.000)	0.588 (0.000)	0.159
Mj7132	710	11	0.771	0.737	4.430	0.046 (0.008)	0.017 (0.241)	0.030 (0.070)	0.016
Mj5331	717	23	0.895	0.862	3.629	0.038 (0.002)	0.017 (0.124)	0.021 (0.060)	-0.002
Mean	714	17.3	0.820	0.691	17.600	0.178 -	0.036 -	0.153 -	0.052

4.4.2 Population genetic diversity

The genetic diversity, relatedness and effective population sizes of *M. jurtina* across the dataset can be found in Table 4.4. H_e across the whole datasets was 0.821. Genetic diversity (H_e and A_r) was similar across all populations with H_e ranging from 0.712 (BB) to 0.831 (BC) and A_r ranging from 5.000 (B-B) to 6.19 (FR-B). A_r was significantly lower across mainland European populations (A_r = 5.699) than across UK populations (A_r 5.904) (p = 0.015). H_e was significantly higher across mainland European populations (H_e = 0.804) than UK populations (H_e = 0.764) (p > 0.001). Two private alleles (A_P) occurred at DK-A, whilst UK-LC, UK-MC, UK-PF, FR-C, ND-B, L, DK-B, G-A and E-D each had a single private allele. All other sites had no private alleles.

Using both the heterozygote excess and the linkage disequilibrium methods effective population sizes were estimated at infinity. In some case the point estimates were estimated at infinity, in others the confidence limits included infinity (Table 4.5).

Table 4.5. Sample sizes, genetic diversity, allelic richness, number of private alleles and effective population sizes for *M. jurtina* populations in the south of England. All values are estimated on a per population basis.

H_{exp} expected heterozygosity, A_r = allelic richness, A_p = Private alleles, $N_e(1)$ = effective population size estimated using the heterozygote excess method, $N_e(2)$ = effective population size estimated using linkage disequilibrium method.

Site	Sample Size	Mean H_{exp} (s.d.)		A_r	A_p	$N_e(1)$	$N_e(2)$
All sites	722	0.821	(0.108)	-	11	-	-
UK-ARN	21	0.760	(0.176)	5.776	0	∞	∞
UK-ARS	17	0.753	(0.224)	5.614	0	∞	∞
UK-AU	14	0.713	(0.269)	5.574	0	∞	∞
UK-B	17	0.738	(0.246)	5.693	0	∞	∞

Site	Sample Size	Mean H_{exp}	(s.d.)	A_r	A_p	$N_e(1)$	$N_e(2)$
UK-C	20	0.783	(0.199)	5.990	0	∞	∞
UK-CH	20	0.805	(0.145)	6.090	0	∞	∞
UK-D	15	0.776	(0.223)	6.132	0	∞	∞
UK-HP	20	0.786	(0.163)	5.935	0	∞	∞
UK-LC	20	0.796	(0.176)	6.156	1	∞	∞
UK-LW	20	0.784	(0.229)	6.100	0	∞	∞
UK-MC	16	0.736	(0.266)	5.963	1	∞	∞
UK-PF	37	0.771	(0.223)	6.121	1	∞	∞
UK-SD	15	0.747	(0.279)	5.940	0	∞	∞
UK-TC	15	0.771	(0.213)	5.781	0	∞	∞
UK-WW	20	0.745	(0.192)	5.688	0	∞	∞
SP	15	0.827	(0.099)	6.035	0	∞	∞
FR-A	20	0.819	(0.068)	5.887	0	∞	∞
FR-B	20	0.818	(0.063)	5.607	0	∞	∞
FR-C	18	0.826	(0.116)	6.186	1	∞	∞
FR-D	20	0.818	(0.112)	6.044	0	∞	∞
B-A	20	0.798	(0.124)	5.801	0	∞	∞
B-B	6	0.712	(0.273)	5.000	0	∞	∞
B-C	18	0.831	(0.066)	5.879	0	∞	∞
ND-A	15	0.811	(0.084)	5.875	0	∞	∞
ND-B	16	0.804	(0.078)	5.606	1	∞	∞
L	18	0.793	(0.110)	5.739	1	∞	∞
SZ	18	0.770	(0.119)	5.406	0	∞	∞
DK-A	19	0.810	(0.077)	5.616	2	∞	∞
DK-B	20	0.806	(0.099)	5.852	1	∞	∞
G-A	23	0.802	(0.090)	5.654	1	∞	∞
G-B	11	0.768	(0.167)	5.335	0	∞	∞
G-C	20	0.817	(0.073)	5.645	0	∞	∞
G-D	22	0.828	(0.079)	6.091	0	∞	∞
G-E	20	0.814	(0.096)	5.895	0	∞	∞
S-W	15	0.815	(0.073)	5.467	0	∞	∞
E-A	19	0.813	(0.074)	5.603	0	∞	∞
E-B	20	0.756	(0.068)	5.077	0	∞	∞
E-C	22	0.817	(0.052)	5.554	0	∞	∞
E-D	20	0.821	(0.076)	5.913	1	∞	∞

4.4.3 Population genetic structure, differentiation and gene flow

Using the six microsatellites described the initial STRUCTURE analysis found the most likely value of LnPr(X|K) was K = 4 (Fig. 4.2a). However subsequent analysis using the Evanno method (Evanno

et al. 2005) revealed the most likely number of clusters was two, then four (Fig. 4.2b). Clustering for K = 2 and K = 4 are shown in Figs. 4.3 and 4.4 respectively. Cluster 1 originated predominantly from the UK populations. Mainland European populations predominantly made-up cluster 2 and were more admixed.

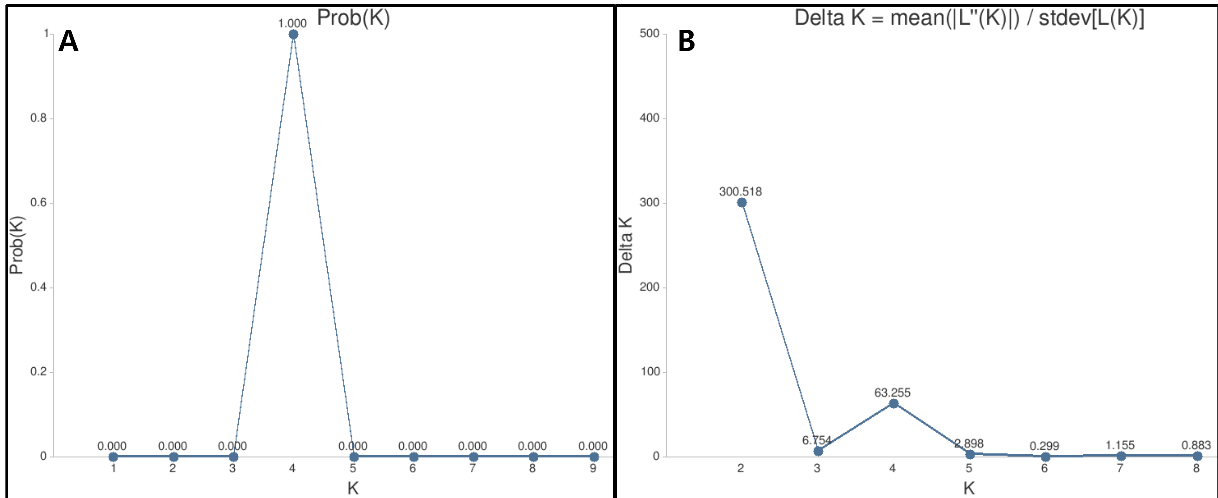


Figure 4.2 Bayesian cluster analysis in STRUCTURE. (A) Initial estimation of K for *M. jurtina* using median values of LnPr(X|K) (Pritchard *et al.* 2000), k for which Pr(K=k) is highest = 4. (B) Estimation of Delta K (Evanno *et al.* 2005) for *M. jurtina* across Europe, highest Delta K = most likely K.

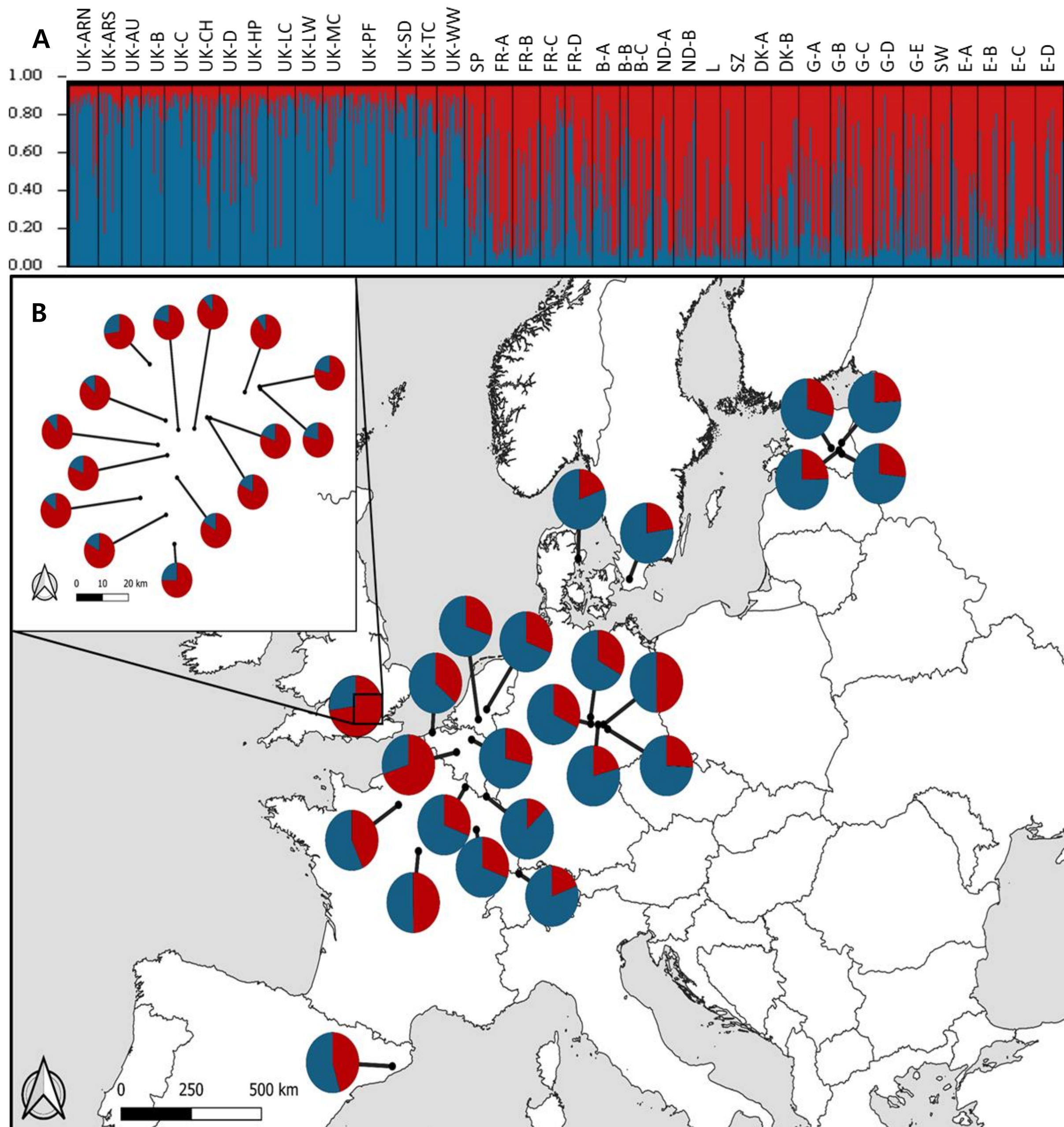


Figure 4.3 Bayesian cluster analysis in STRUCTURE (A) Individual membership coefficients for *M. jurtina* across Europe where K =2. (B) Geographic distribution of clusters, when K = 2. Colours indicate different clusters.

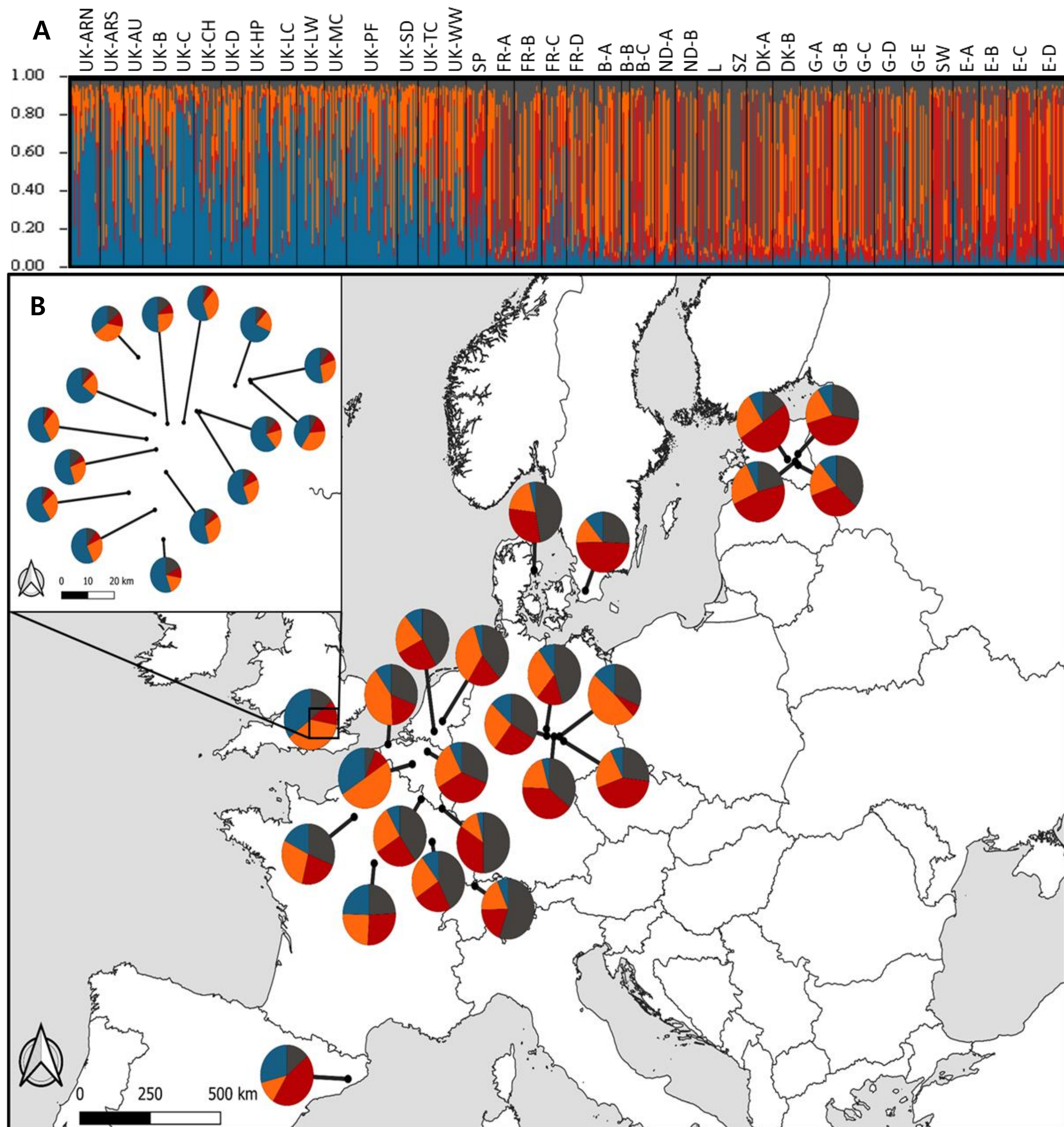


Figure 4.4 Bayesian cluster analysis in STRUCTURE (a) Individual membership coefficients for *M. jurtina* across Europe where K =4. (b) Geographic distribution of clusters, when K = 4. Colours indicate different clusters.

There was little variation within the pairwise F_{ST} values (μ= 0.028, variance = 0.0007) (Table 4.6). Pairwise F_{ST} was significantly greater than zero for 236 out of 722 site pair combinations. There was no evidence of IBD between pairwise site combinations and pairwise F_{ST} values when all sites were included in the analysis (Fig. 4.5) (p = 0.201, Mantel r value = 0.061). Removal of the Spanish site also resulted in no significant IBD, however removal of the Estonian sites resulted in a significant IBD effect (p = 0.020, Mantel r value = 0.175). The estimated number of migrants per generation according to Slatkin's private allele method was 8.004 across all sites. When UK sites were removed

from the analysis the estimated number of migrants per generation dropped to 5.980. Estimated migrants per generation for UK only sites was 13.141.

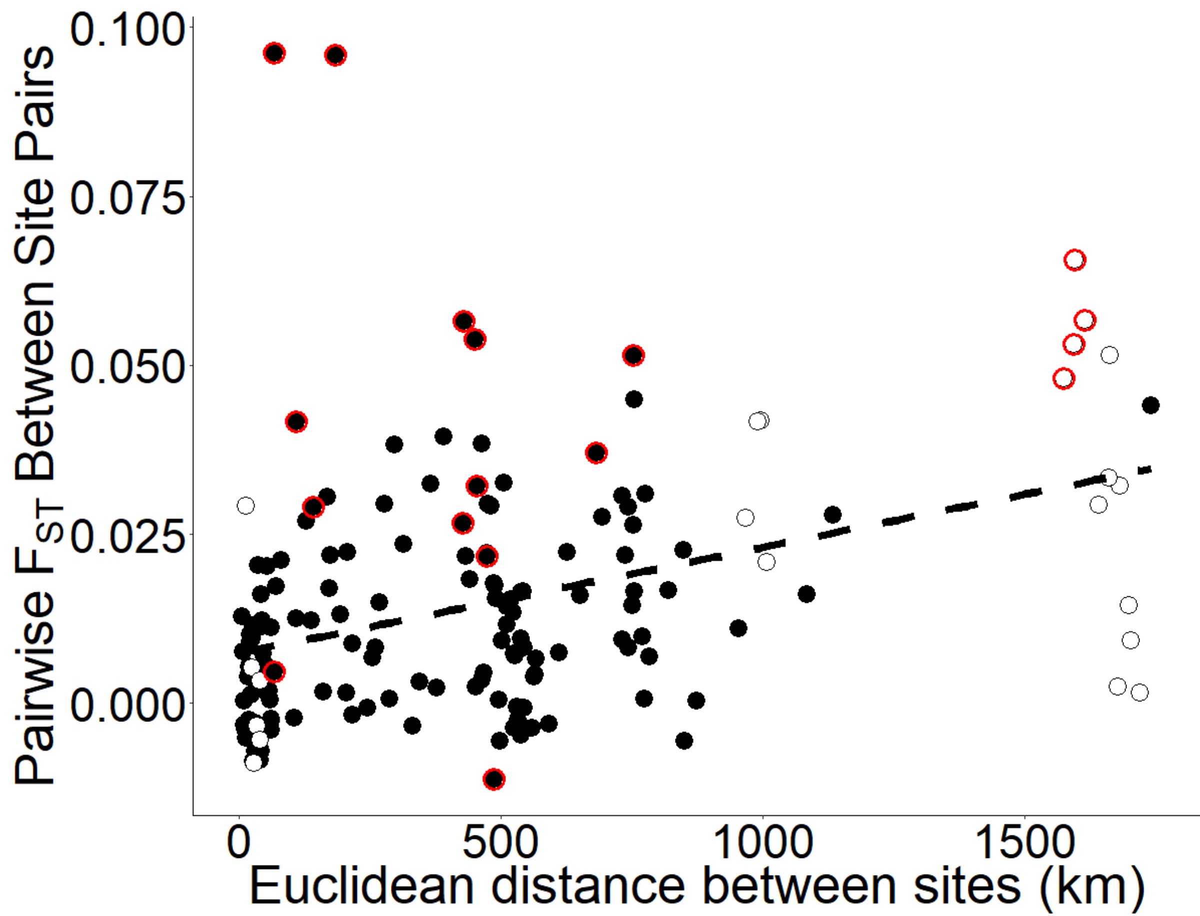


Figure 4.5 Pairwise F_{ST} values for *M. jurtina* plotted against Euclidean distances for 39 sites across Europe in 2017. Points with a red border indicate site pair combinations including site B-B, which had only six samples. White points indicate site pair combinations including at least one Estonian site, removal of these points results in significant IBD between points.

Table 4.6 Pairwise F_{ST} values between 39 pairs of sites across Europe. Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 14820 permutations.

	UK-ARN	UK-ARS	UK-AU	UK-B	UK-C	UK-CH	UK-D	UK-HP	UK-LC	UK-LW	UK-MC	UK-PF	UK-SD	UK-TC	UK-WW	SP	FR-A
UK-ARN	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*
UK-ARS	0.004	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*
UK-AU	0.004	-0.003	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
UK-B	0.013	0.006	-0.003	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*
UK-C	0.007	-0.003	0.003	0.005	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*
UK-CH	0.001	0.012	0.011	0.020	0.002	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
UK-D	0.001	-0.011	0.002	0.001	-0.003	0.012	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
UK-HP	0.003	-0.004	-0.003	-0.009	-0.006	0.003	-0.002	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS
UK-LC	0.006	0.003	0.013	0.004	-0.003	0.003	-0.010	-0.003	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS
UK-LW	0.012	0.000	0.008	0.012	-0.007	0.004	-0.008	0.002	-0.008	*0.000*	NS	NS	NS	NS	NS	*	*
UK-MC	0.009	-0.003	-0.005	-0.003	0.001	0.014	-0.012	-0.006	0.005	0.006	*0.000*	NS	NS	NS	NS	*	NS
UK-PF	0.009	0.000	-0.005	-0.004	0.000	0.011	-0.006	-0.006	0.002	-0.001	-0.006	*0.000*	NS	NS	NS	*	*
UK-SD	0.014	-0.004	-0.005	-0.007	-0.008	0.008	-0.002	-0.005	0.001	-0.009	-0.003	-0.005	*0.000*	NS	NS	NS	NS
UK-TC	0.015	0.004	0.003	0.005	-0.004	0.003	-0.002	-0.003	0.000	0.005	0.006	-0.002	-0.002	*0.000*	NS	NS	NS
UK-WW	0.007	0.009	-0.005	0.012	0.011	0.011	0.010	0.000	0.008	0.015	0.003	0.008	0.007	0.005	*0.000*	NS	NS
SP	**0.075**	**0.085**	**0.090**	**0.084**	**0.067**	0.045	**0.078**	0.054	0.059	**0.071**	**0.095**	**0.076**	0.069	0.061	0.068	*0.000*	NS
FR-A	**0.056**	**0.061**	0.070	**0.068**	**0.054**	0.041	0.055	0.039	0.042	**0.054**	0.062	**0.053**	0.065	0.045	0.044	0.042	*0.000*
FR-B	**0.054**	**0.055**	0.062	**0.061**	**0.047**	0.026	0.043	**0.035**	0.035	**0.044**	0.060	**0.050**	**0.056**	0.036	0.048	0.038	-0.008
FR-C	**0.051**	0.040	0.051	0.052	**0.034**	0.027	0.033	0.033	0.037	0.028	0.054	0.034	0.036	0.018	0.046	0.030	0.012
FR-D	**0.037**	0.027	0.033	**0.044**	**0.030**	0.019	0.023	0.019	0.018	0.025	0.036	**0.024**	0.037	0.016	0.031	0.041	0.008
B-A	**0.041**	**0.033**	0.039	0.034	**0.028**	**0.030**	0.023	0.018	0.018	**0.026**	0.037	**0.027**	0.029	0.014	0.029	**0.065**	-0.001
B-B	0.020	-0.011	-0.006	0.018	0.008	0.028	-0.006	0.012	0.002	-0.001	0.007	0.006	0.005	-0.002	0.003	0.097	0.042
B-C	**0.042**	**0.045**	0.048	**0.055**	**0.040**	0.021	**0.040**	0.029	0.030	**0.045**	**0.057**	**0.041**	0.049	0.025	0.039	0.021	0.009
ND-A	**0.051**	**0.050**	**0.064**	**0.054**	**0.043**	**0.037**	0.036	0.030	0.025	**0.041**	0.045	**0.045**	0.047	0.031	0.044	**0.066**	-0.004
ND-B	**0.050**	0.049	0.060	0.053	**0.053**	0.028	**0.052**	0.026	0.041	**0.058**	0.059	**0.053**	**0.062**	0.047	0.041	**0.052**	0.007
L	**0.093**	**0.102**	**0.119**	**0.108**	**0.094**	**0.076**	**0.091**	**0.073**	**0.081**	**0.102**	**0.104**	**0.096**	**0.119**	**0.087**	**0.091**	**0.064**	-0.002
SZ	**0.099**	**0.098**	**0.115**	**0.117**	**0.089**	**0.070**	**0.093**	**0.081**	**0.079**	**0.092**	**0.110**	**0.094**	**0.110**	**0.079**	**0.100**	**0.090**	0.021
DK-A	**0.068**	**0.061**	**0.085**	**0.079**	**0.066**	**0.050**	**0.067**	**0.041**	**0.050**	**0.065**	**0.079**	**0.069**	**0.071**	0.060	**0.060**	**0.045**	-0.003
DK-B	**0.058**	**0.044**	**0.063**	**0.056**	**0.039**	**0.039**	**0.043**	0.034	**0.031**	**0.041**	**0.059**	**0.046**	**0.037**	0.030	**0.056**	0.051	0.011
G-A	**0.045**	**0.051**	0.049	**0.059**	**0.052**	**0.032**	**0.050**	0.029	0.036	**0.054**	**0.063**	**0.048**	0.062	0.036	0.041	**0.041**	0.012
G-B	0.021	0.012	0.028	0.025	**0.023**	0.017	0.016	0.012	0.007	**0.022**	0.022	0.019	0.025	0.012	0.019	**0.081**	0.010
G-C	**0.053**	**0.058**	**0.070**	**0.069**	**0.059**	**0.041**	0.051	**0.039**	**0.045**	**0.066**	**0.064**	**0.059**	**0.072**	0.045	0.049	**0.043**	0.001
G-D	**0.049**	**0.040**	0.056	**0.055**	**0.040**	0.034	0.031	0.028	0.021	0.035	0.051	**0.040**	0.051	0.029	0.047	**0.049**	-0.001
G-E	**0.022**	**0.030**	0.044	**0.052**	**0.040**	**0.016**	0.026	0.020	0.024	**0.033**	0.036	**0.034**	0.047	0.033	0.032	**0.059**	0.017
S-W	**0.071**	**0.063**	0.071	**0.071**	**0.066**	**0.049**	0.051	**0.048**	**0.046**	**0.064**	**0.069**	**0.059**	**0.079**	0.042	**0.058**	**0.044**	0.000
E-A	**0.047**	**0.055**	**0.053**	**0.067**	**0.050**	0.019	**0.051**	**0.035**	**0.038**	**0.047**	**0.054**	**0.048**	**0.060**	0.038	**0.044**	**0.043**	**0.028**
E-B	**0.055**	**0.066**	**0.066**	**0.087**	**0.079**	**0.040**	**0.073**	**0.055**	**0.063**	**0.078**	**0.085**	**0.067**	**0.088**	**0.059**	**0.060**	**0.073**	**0.050**
E-C	**0.049**	**0.048**	**0.060**	**0.067**	**0.057**	**0.034**	**0.045**	**0.043**	**0.046**	**0.056**	**0.065**	**0.051**	**0.066**	**0.042**	**0.058**	**0.051**	**0.035**
E-D	**0.064**	**0.061**	**0.069**	**0.067**	**0.058**	**0.040**	**0.052**	**0.049**	**0.049**	**0.061**	**0.068**	**0.059**	**0.065**	0.042	**0.058**	**0.052**	**0.044**

Table 4.6 Continued

	FR-B	FR-C	FR-D	B-A	B-B	B-C	ND-A	ND-B	L	SZ	DK-A	DK-B	G-A	G-B	G-C	G-D	G-E	S-W	E-A	E-B	E-C	E-D
UK-ARN	*	*	*	*	NS	*	*	*	*	*	*	*	*	NS	*	*	*	*	*	*	*	*
UK-ARS	*	NS	NS	*	NS	*	*	NS	*	*	*	*	*	NS	*	*	*	*	*	*	*	*
UK-AU	NS	NS	NS	NS	NS	NS	*	NS	*	*	*	*	NS	NS	*	NS	NS	NS	*	*	*	*
UK-B	*	NS	*	NS	NS	*	*	NS	*	*	*	*	*	NS	*	*	*	*	*	*	*	*
UK-C	*	*	*	*	NS	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
UK-CH	NS	NS	NS	*	NS	NS	*	NS	*	*	*	*	*	NS	*	NS	*	*	NS	*	*	*
UK-D	NS	NS	NS	NS	NS	*	NS	*	*	*	*	*	*	NS	NS	NS	NS	NS	*	*	*	*
UK-HP	*	NS	NS	NS	NS	NS	NS	NS	*	*	*	NS	NS	NS	*	NS	NS	*	*	*	*	*
UK-LC	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*	NS	NS	*	NS	NS	*	*	*	*	*
UK-LW	*	NS	NS	*	NS	*	*	*	*	*	*	*	*	*	*	NS	*	*	*	*	*	*
UK-MC	NS	NS	NS	NS	NS	*	NS	NS	*	*	*	*	*	NS	*	NS	NS	*	*	*	*	*
UK-PF	*	NS	*	*	NS	*	*	*	*	*	*	*	*	NS	*	*	*	*	*	*	*	*
UK-SD	*	NS	NS	NS	NS	NS	NS	*	*	*	*	*	NS	NS	*	NS	NS	*	*	*	*	*
UK-TC	NS	NS	NS	NS	NS	NS	NS	NS	*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	NS
UK-WW	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*	NS	NS	NS	NS	NS	*	*	*	*	*
SP	NS	NS	NS	*	NS	NS	*	*	*	*	*	NS	*	*	*	*	*	*	*	*	*	*
FR-A	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
FR-B	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*
FR-C	0.008	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	*	NS	*	*	NS
FR-D	0.009	-0.001	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*
B-A	0.001	0.002	-0.001	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	NS
B-B	0.038	0.039	0.020	0.011	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
B-C	0.007	0.007	0.001	0.005	0.040	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
ND-A	-0.004	0.027	0.017	-0.002	0.029	0.018	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS
ND-B	0.002	0.033	0.022	0.017	0.042	0.022	0.018	*0.000*	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
L	0.003	0.038	0.039	0.027	0.096	0.031	0.016	0.015	*0.000*	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	*	*	NS
SZ	0.013	**0.038**	0.024	0.022	0.096	0.030	0.017	**0.044**	0.022	*0.000*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*
DK-A	0.002	0.032	**0.029**	0.018	0.054	0.016	-0.004	0.009	0.015	**0.024**	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	*	NS	*
DK-B	0.011	0.018	0.020	0.008	0.037	0.015	0.003	0.037	0.046	0.041	0.008	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	*	*
G-A	0.008	0.026	0.017	0.014	0.032	-0.003	0.014	0.012	0.023	0.036	0.014	0.034	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS
G-B	0.004	0.031	0.007	-0.001	-0.011	0.022	0.000	0.013	**0.047**	0.037	0.017	0.018	0.020	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS
G-C	-0.005	0.029	0.022	0.016	0.057	0.007	-0.001	0.004	-0.001	0.016	-0.006	0.022	0.003	0.017	*0.000*	NS	NS	NS	NS	NS	NS	NS
G-D	-0.004	0.010	0.001	-0.002	0.022	0.008	-0.008	0.015	0.014	0.020	0.007	0.012	-0.003	0.010	0.006	*0.000*	NS	NS	NS	*	NS	NS
G-E	0.007	0.031	0.009	0.018	0.027	0.004	0.008	0.015	0.037	0.031	0.015	0.031	0.010	-0.002	0.009	0.007	*0.000*	NS	NS	NS	NS	NS
S-W	0.011	**0.028**	0.016	0.017	0.051	-0.006	0.016	0.028	0.023	0.034	0.013	0.012	0.005	0.030	0.002	0.004	0.022	*0.000*	NS	NS	NS	NS
E-A	0.014	0.035	0.016	0.032	0.057	0.002	0.016	0.025	0.043	0.032	**0.021**	0.038	0.008	0.019	0.006	0.017	0.000	0.019	*0.000*	NS	NS	NS
E-B	**0.043**	**0.057**	0.026	**0.051**	0.066	0.009	**0.053**	0.039	**0.069**	0.049	**0.042**	**0.064**	0.013	0.024	0.026	**0.038**	0.009	0.030	-0.003	*0.000*	NS	NS
E-C	**0.015**	**0.029**	0.017	**0.029**	0.048	0.002	0.024	0.023	**0.041**	0.038	0.027	**0.034**	0.013	0.020	0.006	0.018	0.007	0.018	-0.006	0.003	*0.000*	NS
E-D	**0.021**	0.037	**0.032**	0.033	0.053	0.015	0.027	0.026	0.044	**0.054**	**0.042**	**0.050**	0.019	0.026	0.012	0.032	0.026	0.029	0.005	0.029	-0.009	*0.000*

4.4.4 Population genetic networks, barriers and bottlenecks

The multivariate analysis of the six loci determined that greatest amount of the variability was explained by the first principal component, as the first eigenvalue of the sPCA was highly positive, indicative of global structuring (Jombart 2008) (Fig. 4.6). As this eigenvalue could clearly be distinguished from all others only the first principal component was analysed. The structure illustrated by the individual lagged scores on the first principal component showed a distinct structure (Fig. 4.6), with the greatest difference between the UK populations and those in Europe.

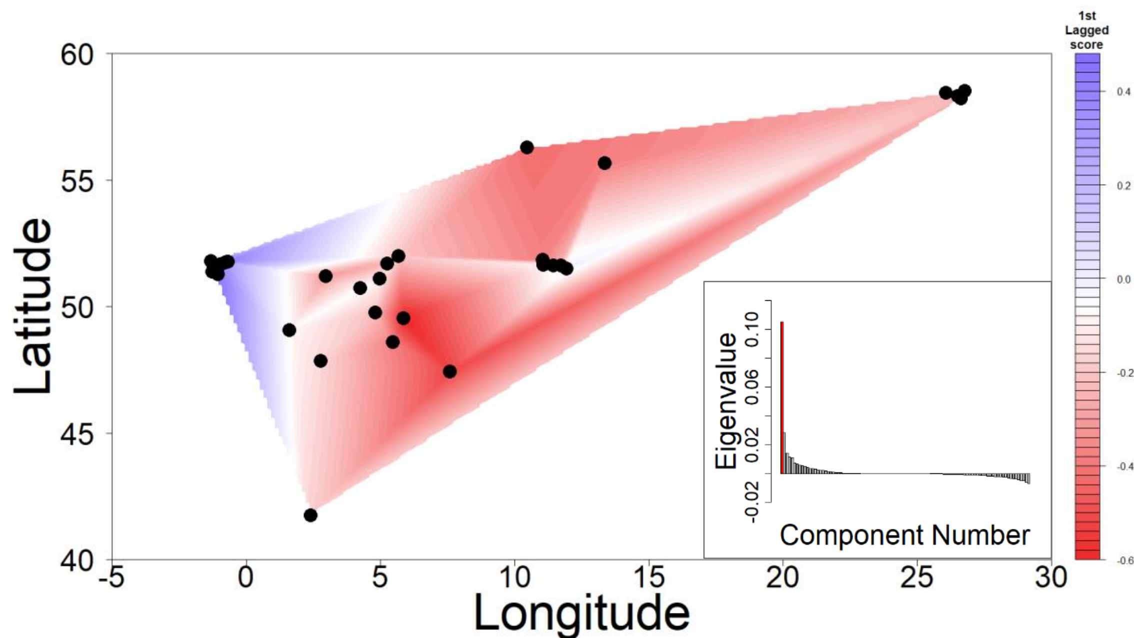


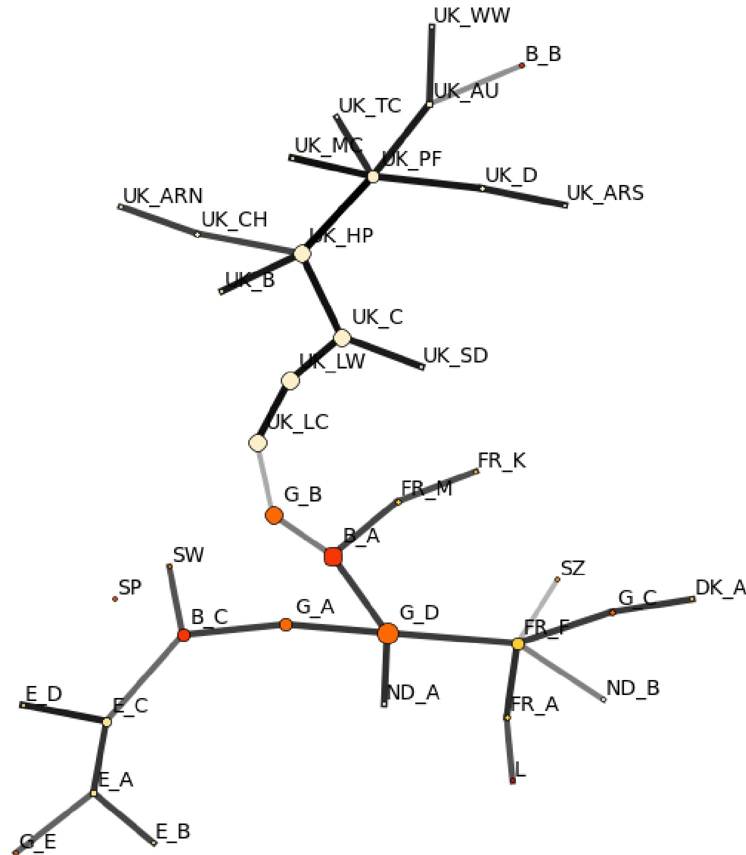
Figure 4.6 Interpolation of the lagged vectors of the first principal component. Sites are plotted by spatial coordinates. Eigenvalues are shown in the inset scree plot.

All ten MSNs calculated resulted in the same shape tree, oriented differently. A representative network can be seen in Fig. 4.7. All networks showed the node SP to be disconnected from the network. Strong edges were drawn between UK nodes, and between the majority of mainland European nodes. However, only a weak edge joined the UK nodes to the mainland European nodes via UK_LC and G_B. Weak edges were also drawn between SZ and FR_F, and between B_B and UK_AU. B_B was the only other mainland European node to connect with a UK node, with the exception of the main edge between the two described above.

The automatic TN percolation threshold value was set at 0.03. This resulted in a giant component made up of a UK network (plus node B_B) and a mainland European network, with the two connected by a single edge. The nodes SP and SZ were disconnected from all other nodes. Manually reducing the threshold to 0.0297 removed the edge between the UK and the European mainland, separating the giant component into the two constituent ones described above (Fig. 4.8). The nodes with the highest betweenness centrality scores in the TN did not consistently appear in the top five or top nodes during randomisation. This indicates that the betweenness centrality scores are affected by bootstrapping and therefore not significant.

No strong evidence of recent genetic bottlenecks was detected at the population level (Appendix E, Table E1). No populations displayed a mode shift (Appendix E, Table E2). However contrary to expectations eight mainland European populations showed some evidence of possible bottlenecks with a majority of tests being significant (Appendix E, Table E3).

A



B

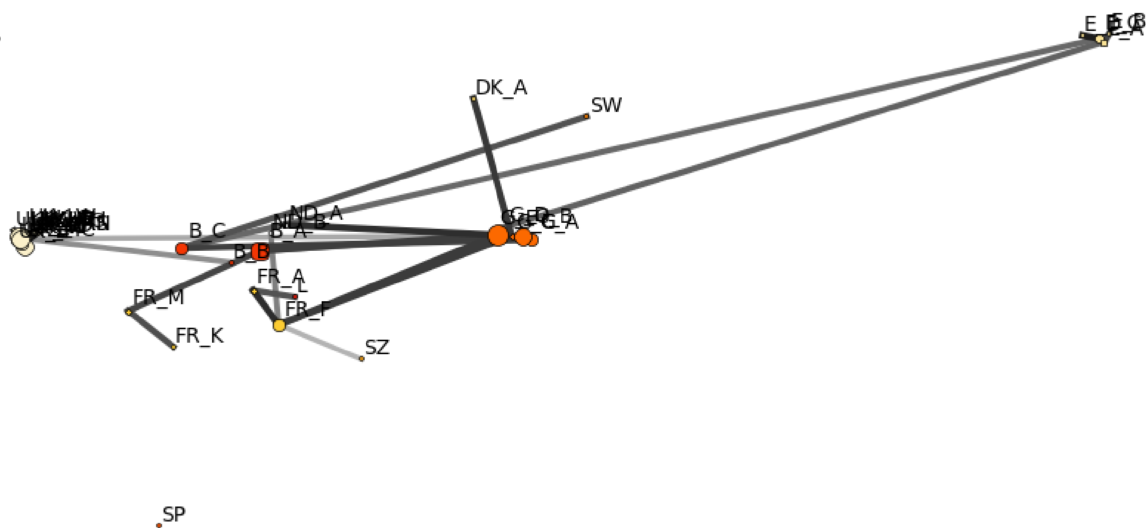


Figure 4.7 Representative Minimum Spanning Tree of pairwise F_{ST} values for *M. jurtina* populations across Europe. Line thickness and shade proportional to link strength. Node size is proportional to betweenness centrality, and nodes are coloured by country. The network is displayed twice A) without spatial coordinates and B) with sites plotted using their spatial coordinates.

Figure 4.8 Network analysis of pairwise F_{ST} values for *M. jurtina* populations across Europe. Line thickness and shade proportional to link strength. Node size is proportional to betweenness centrality, and nodes are coloured by country. The network is displayed twice A) without spatial coordinates and B) with sites plotted using their spatial coordinates.

4.5 Discussion

Our results contribute to the growing body of evidence that show *M. jurtina* to have high levels of genetic diversity across much of its European range (Goulson 1993a; Wood & Pullin 2002; Schmitt *et al.* 2005; Habel *et al.* 2009; Richard *et al.* 2015; Villemey *et al.* 2016). We found significant differences in the levels of genetic diversity of *M. jurtina* populations in the UK compared to those on the European mainland. We also found evidence of genetic differentiation across Europe, with UK populations belonging to a distinct cluster. On the mainland clustering is more complex with high levels of admixture seen within all populations. We show tentative evidence of IBD effects, although sampling design should be taken into consideration when interpreting these results.

4.5.1 Population genetic diversity

We found a lower mean number of alleles per locus than a previous study by Villemey *et al.* (2016) using five of the same markers. Our results were however similar to those in Richard *et al.* (2015). The differences between our results and those of Villemey *et al.* (2016) is likely a reflection of the differences in sampling methodologies. Investigations into the number of alleles in a population are affected by both the number of samples used, and the number of sample areas used within a region, both of which can result in a greater number of alleles being found (Kalinowski 2004). The fewer number of samples collected in this study (810 Vs 1681), combined with the collection over a greater geographic expanse (39 sites from eleven countries across 2525km vs 469 sites from three regions in France across 600km), may explain the greater number of alleles found per locus by Villemey *et al.* (2016)

We found contradicting differences in the two measures used as estimates of genetic diversity across the UK and mainland European populations, with higher levels of H_e and lower levels of A_r on the mainland. An explanation for this may be to do with a limitation of A_r within this study. Whilst both H_e and A_r are commonly used metrics of genetic diversity, H_e is the more common (Greenbaum *et al.* 2014). A key reason for this is that A_r is highly affected by sample size, meaning that rarefaction procedures need to be undertaken in order not to bias estimates (Kalinowski 2004) e.g. whereby larger sample sizes result in higher estimates due to more alleles being sampled. Rarefaction procedures use the smallest sample size across all sampled populations to estimate A_r for each population (Leberg 2002), which in this study results in a samples size of six being used. Considering that between 20 and 30 samples is often sufficient to accurately estimate allele frequencies (Pruett & Winker 2008; Hale *et al.* 2012), the low sample size used as a result of rarefaction means that our estimates of allelic richness may not be as reliable as the estimates of heterozygosity, which does not rely upon such methods. Using H_e as a measure of genetic diversity, we see the expected result of the island population (UK) displaying significantly lower genetic diversity to the mainland, as is often seen with island populations (Frankham 1997). Although significantly lower than mainland $H_{e,}$ the levels seen in UK populations are still very high. This may be due to the relatively large size and minimal isolation of the UK, meaning that inbreeding and genetic drift have less effect than they might have, were the UK smaller and more isolated (Furlan *et al.* 2012).

With the exception of Mj4870, all loci displayed high levels of H_o. All loci also displayed heterozygote deficit, although these did not constitute a significant departure from HWE. However, when UK samples were removed significant departures from HWE were observed (see Appendix D). Inbreeding, strong selection pressures and the Wahlund effect can all cause heterozygote deficit across loci (Selkoe & Toonen 2006; Dharmarajan *et al.* 2013), therefore the reported heterozygote deficits may be due significant levels of inbreeding, as indicated by F_{IS} values, across four of the six loci. Loci Mj7132 and Mj5331 displayed near significant levels of inbreeding (p = 0.07 and 0.06 respectively). As the levels of H_o displayed at Mj4870 are considerably lower than all other loci, and levels of heterozygote higher, it is likely that a more locus specific factor is having an effect. These results are most likely due to the presence of null alleles, which can cause analytical problems when occurring at a frequency higher than 0.2 (Dakin & Avise 2004). The fact that Mj4870 has a greater than 0.2 frequency of null alleles across a number of populations, all of which occur in mainland Europe explains: a) why H_o is lower than all other loci, and b) why the removal of UK sites results in a significant departure from HWE. As described in the methods, the analysis was repeated with the locus Mj4870 removed, but this had little effect on the results.

Estimates of effective population size were all infinite, or included infinity in their confidence intervals i.e. "there was no evidence for variation in the genetic characteristic caused by genetic drift due to a finite number of parents" (Do *et al.* 2014). This was the result of insufficient signal in the data to predict effective population sizes and likely an artefact of the difficulty of predicting effective population sizes of large populations (where $n > 1000$), particularly when only using small sample sizes (Waples & Do 2010)

4.5.2 Population genetic differentiation

Whilst no evidence of IBD was found when all sampling locations were included in the analysis, it should be noted that the removal of the Estonian sampling sites, resulting in a more even distribution of distances between sampling points, did result in significant IBD. Additionally, the low number of samples (six) included from site B-B appears to reduce the reliability of these results, as nine of the ten highest pairwise F_{ST} values include B-B as one of the sites in the site pair. A more even sampling regime across Europe, with more even sample numbers, could result in clearer evidence of IBD, as may be expected across large areas when individual dispersal distances are small (Wright 1943).

F_{ST} scores for the loci Mj5522, Mj0247, Mj7132 and Mj5331 were all low (<0.02) and were not significantly different from zero. F_{ST} scores for Mj7232 (0.054) and Mj4870 (0.102) were significantly different from zero and of a level considered to indicate moderate genetic differentiation ($0.05 < F_{ST} > 0.25$) (Freeland *et al.* 2011). These results indicate some differentiation, potentially as a result of obstructions to gene flow. However, the average F_{ST} score across all loci suggests that gene flow remains high. These results are higher than those reported in Richard *et al.* (2015), which may be due to the larger geographic area of the study and IBD effects, whereby larger geographic distances between samples often resulting in greater differentiation due to reduced migration between distant sites when compared to more local sites (Bradburd *et al.* 2013).

The mean pairwise F_{ST} score between pairs of populations was 0.033, again indicating low levels of genetic differentiation. However, a large number of population pairs displayed significant, moderate levels of differentiation. The majority of these pairs included one UK and one mainland European population, with a mean pairwise F_{ST} across all such combinations of 0.049. No pairwise F_{ST} scores between pairs of UK sites were significantly greater than zero. Examples of pairwise F_{ST} scores significantly greater than zero between pairs of mainland European sites are distributed throughout the data, however the majority of these occur when one population is from Estonia or Spain, i.e. the two most geographically separated sampling areas. This is further evidence to suggest that some IBD effects may be occurring, but not being picked up due to the uneven distribution of sampling distances.

The STRUCTURE analysis showed clear genetic structure between populations of *M. jurtina* in this study. There is strong evidence to suggest that UK populations are genetically distinct from those on the European mainland, with far higher levels of admixture within mainland populations. This suggests some level of isolation between the UK and mainland Europe. The low number of microsatellite loci used in this analysis means that these STRUCTURE results should be interpreted with caution (Orozco-Terwengel *et al.* 2011), however, they are in alignment with the other analyses used in this study such as the significant pairwise F_{ST} values discussed previously.

All combinations of sites resulted in estimations of migrants per generation greater than 2, a value indicative of high levels of gene flow (Slatkin 1985). Unsurprisingly the estimation using just sites from the UK is far higher than when using just those from mainland Europe, likely due to the much smaller distances between all pairs of UK sites than mainland sites. It might be expected that the estimated number of migrants would be greater when using only mainland sites, than when all sites were included, as this would be representative of the assumed barrier to gene flow between the UK and mainland Europe. However, this is not the case, with migrant estimates increasing with the inclusion of UK sites. This is probably an effect of the previously mentioned high levels of migration between UK populations, but also because of the methodology used to make the estimates. Slatkin's method is based upon the number of private alleles within populations (Slatkin 1985; Barton & Slatkin 1986). The inclusion of more sites and samples reduces the mean frequency of private alleles and therefore may contribute to the reduced values seen. These problems are less of an issue with Bayesian methods to estimate contemporary migration rates, however such methods were not possible in this study due to a lack of resolution in the data preventing MCMC convergence. An increased number of microsatellites may resolve this issue in future studies.

4.5.3 Population genetic networks, barriers and bottlenecks

The results of these analyses are complimentary with the STRUCTURE analysis and further support the idea that populations of *M. jurtina* in the UK are genetically distinct from those on the European mainland. The sPCA confirms that the greatest variation within the data occurs between the UK populations and mainland European ones.

This is also how the Network analysis groups populations. However, a number of points need to be addressed. Firstly, in both the MSN and TN the node SP is disconnected from the

network. This suggests that SP is genetically separated from both the UK sites and the European mainland ones. This could be explained by the Pyrenees acting as a barrier to gene flow, however these results are only tentatively supported by the STRUCTURE analysis when K=4. The node SZ in Switzerland is also disconnected in the TN. Again, this is not supported by the STRUCTURE analysis, and as the sample site is north of the Alps, cannot be explained by a mountain barrier, as may be a potential reason for the disconnection of the Spanish site SP. Secondly, the Belgian node B-B is connected to a UK site in the MSN and is assigned to the UK network in the TN. The most likely explanation for this is the low number of samples from that site (six). A final point to note from the network analysis is that the edges between UK nodes are generally higher than those between mainland European nodes, likely a result of the closer proximity of the UK nodes to each other.

Interestingly there is no evidence of a population bottleneck occurring with any UK population. These results suggest that whilst UK populations are genetically distinct, this does not stem from an initial colonisation and founder effect as is often the case with population bottlenecks (Mayr 1954; Nei *et al.* 1975). A possible reason for this could be that the UK was colonised by the western lineage of *M. jurtina* during the post glacial expansion described in Schmitt *et al.* (2005) and Dapporto *et al.* (2011). During this period a land-bridge (Doggerland) connected the UK with mainland Europe. The subsequent flooding of this land-bridge around 6000 years ago (Shennan *et al.* 2000) likely created a sufficient barrier to reduce geneflow, resulting in the contemporary genetic differentiation we see today. However, as this occurred after *M. jurtina* colonised the UK no bottleneck or founder event occurred.

4.5.4 Conclusions, limitations and future work

M. jurtina exhibits high levels of genetic diversity across its studied range. The complicated evolutionary history and distribution of *M. jurtina* suggests that two distinct lineages occur across Europe (Dapporto *et al.* 2014). These lineages, as evidenced by morphology and allozyme distributions, have no effect on contemporary levels of gene flow, with relatively little genetic differentiation across Europe as a whole. The exception to this is between the UK and mainland Europe, with distinct differences occurring between the two. Overall, our results suggest restricted gene flow between the UK and Europe, but not complete isolation and with no evidence of a genetic bottleneck. Therefore it is likely that stretches of open water do act as barriers to gene flow, as suggested by Dowdeswell (1961) and Baxter *et al.* (2017), but are not completely impervious, as evidenced by observations reported by Dennis & Shreeve (1996), with rare crossing events occurring. Across Europe we also find tentative evidence of IBD, although more rigorous sampling is suggested in order to clarify the situation. Our results are similar to those of other widely distributed, generalist species such as the green-veined white, (*Pieris napi*) and common blue (*Polyommatus Icarus*) (Geiger & Shapiro 1992; Schmitt *et al.* 2003).

Some limitations must be taken into account when interpreting these results (for further details see Chapter 3). Firstly, the number of samples from each population is lower than the recommended number deemed sufficient to fully detect the levels of genetic diversity present (Pruett & Winker 2008; Hale et al. 2012). This was particularly problematic for the estimations of

allelic richness discussed earlier. Secondly the number of microsatellites used was low, due to issues with microsatellite characterisation in Lepidoptera (Meglécz *et al.* 2004; Zhang 2004; Mikheyev *et al.* 2010; Tay *et al.* 2010), despite extensive attempts to characterise more (see appendix B). A greater number of microsatellites would increase the reliability of the results (Selkoe & Toonen 2006; Bruford *et al.* 2015) and may allow the use of Bayesian methods to estimate migration rates, shedding more light onto barriers to gene flow across the continent.

Chapter 5. The influence of chalk grasslands on the phenology and ecology of *Maniola jurtina* in the UK

Submitted to *Insect Science* as Greenwell MP, Botham MS, Bruford MW, Day JC, Evans LC, Gibbs M, Roy DB, Watts K, Oliver TH. The influence of chalk grasslands on butterfly phenology and ecology

5.1 Abstract

The meadow brown butterfly, *Maniola jurtina*, is one of Europe's most abundant butterfly species. Found throughout the western Palearctic, it is the most common species recorded across European monitoring schemes. The phenology of the species is unusual for a univoltine grassland butterfly, with a long flight season throughout its range, spanning several months. Protracted flight periods have previously been reported in populations on chalk grassland sites in the south of England, although no attempt has yet been made to quantify this at a national level. Using data from 540 sites across the UK these differences in phenology are quantified and *M. jurtina* phenology is found to be strongly associated with both site geology and topography, independent of levels of abundance. Further investigation into aspects of M. jurtina ecology at a subset of sites finds no genetic structuring or drought tolerance associated with these same site conditions.

5.2 Introduction

The meadow brown, *Maniola jurtina* (L. 1758), is one of the most common and widely distributed butterfly species in Europe. Despite recent declines, it remains a species of 'Least Concern' under the red list of British butterfly species (Fox *et al.* 2011) and is considered stable across much of Europe (Van Swaay *et al.* 2019). The larvae feed on a range of grasses and adults feed on a variety of flower species, with a preference for knapweeds (*Centaurea sp.*) and thistles (*Cirsium sp.*) (Lebeau *et al.* 2018). The wide distribution of these resources, though not the only factor limiting species' distributions (Quinn *et al.* 1998), may go some way to explaining the wide distribution and high abundance of *M. jurtina* across a broad climatic envelope.

The phenology of *M. jurtina* is unusually long for a univoltine, grassland species in the UK, with adults typically on the wing from mid-June through to September (Thomas & Lewington 2010). The flight period varies depending on the environmental conditions and habitat, for example in the south of the UK on some lowland calcareous grassland habitats (herein chalk grasslands) it continues later into the year (Thomas & Lewington 2010). These sites are typically warmer and dryer, favouring thermophilic species (Mortimer *et al.* 1998), and a second peak in emergence has been occasionally observed (Goulson 1993b; Thomas & Lewington 2010). Additionally, chalk grasslands are often associated with areas of variable topography and steep slope angles, which may also play an important role in the phenology of *M. jurtina*, through increased exposure to solar radiation (Bennie *et al.* 2008).

Whilst the protracted flight period of *M. jurtina* on chalk grasslands in the UK has received previous investigation (Shreeve 1989; Goulson 1993b), no effort has yet been made to quantify these differences in phenology at the national scale. Compared to landscapes such as farmland or woodland, chalk grasslands are warmer, dryer, and more topographically heterogeneous, containing a wider range of microclimates (Mortimer *et al.* 1998; Diacon-Bolli *et al.* 2012). This results from differences in vegetation structure and topography that combine to produce substantial variation in ground temperature (Maclean *et al.* 2019). Microclimates may allow individuals to persist in specific locations when surrounding areas of habitat are climatically unsuitable (Bennie *et al.* 2008; Suggitt *et al.* 2011), potentially broadening the flight period. Similarly, extreme warm temperatures in some microclimates may result in local drought conditions which are likely to affect larval development, for example, larvae of speckled wood (*Pararge aegeria*) reared on drought-stressed plants show longer development times (Talloen *et al.* 2004; Gibbs *et al.* 2012, 2018). Thus, longer development times for some individuals coupled with climatically suitable patches, may both contribute to the longer flight periods. Protracted flight periods are also observed in some *M. jurtina* populations in southern Europe (Haeler *et al.* 2014), however, this results from adult females entering a period of aestivation (Brakefield 1984), which has hitherto not been observed in anecdotal observations of UK populations.

Although it seems likely that the variation in UK *M. jurtina* phenology results from differences in conditions that occur within chalk grasslands, the mechanisms that cause these responses are unclear. A parsimonious explanation of the protracted flight period is that these sites contain more favourable habitat and therefore higher abundances, with the broad flight periods simply a result

of the mathematical relationship between mean and variance (Taylor 1961). If, however, the broad flight period of *M. jurtina* on chalk grasslands is not purely the result of high abundances, differences in the ecology of populations at these sites may be affecting phenology.

The broader flight periods on chalk grasslands may be the result of genetic differences between populations, with some anecdotal suggestions of locally adapted races. Whilst we do not explicitly look at local adaptation here, we do investigate the potential for genetic structuring between populations, based upon the type of site that individuals are found. Clear genetic clustering of individuals into chalk and non-chalk populations would suggest a high level of genetic differentiation which may support the idea of locally adapted races as an explanation for the differences in flight periods.

Variation between *M. jurtina* populations has been recorded at large spatial scales (Thomson 1987; Schmitt *et al.* 2005; Habel *et al.* 2009), with two distinct lineages recognised across Europe; the western *M. jurtina jurtina* and the eastern *M. jurtina janira* (Dapporto *et al.* 2014). Whilst large scale phenotypic variation exists in *M. jurtina*, previous studies have also found high levels of genetic diversity at a range of geographic scales (Thomson 1987; Goulson 1993a; Schmitt *et al.* 2005; Habel *et al.* 2009; Richard *et al.* 2015; Villemey *et al.* 2016; Baxter *et al.* 2017). In theory, this could afford a greater ability to adapt to the local environment (McGill *et al.* 2015), leading to protracted flight periods where environmental conditions allow.

It is also possible that populations on chalk grasslands exhibit different levels of drought tolerance, with a possible association between the broader flight periods on chalk grasslands and populations subject to different environmental conditions showing developmental plasticity (variation in physiological development as a result of environment; Breuker *et al.*, 2010). These differences may be a result of either phenotypic plasticity, or more general plasticity across the population in responses to the environment, i.e. mediated by high allelic diversity leading to wider variation in phenotypes. Whilst local adaptation can lead to genetic differentiation, plasticity itself does not necessarily imply genetic differentiation between populations.

To explore these possibilities, we examine the flight periods of *M. jurtina* in the UK at 540 sites differing in geology and topography and quantify the variability in phenology. We confirm that flight periods are protracted on chalk grasslands as previously reported (Goulson 1993b; Thomas & Lewington 2010). After controlling for abundance in our models, we then investigate levels of genetic diversity and differentiation, and drought tolerance at a subset of sites to determine whether differences in phenology are associated with genetic structuring of populations and whether there is evidence of increased drought tolerance from chalk sites that may influence the flight period length. Overall, we test the following:

1) To what extent are *M. jurtina* population flight periods protracted on chalk grasslands in the UK.?
2) Are populations of *M. jurtina* clustered into genetically structured populations based upon the same habitat conditions?
3) Are populations of *M. jurtina* on chalk grasslands more drought tolerant than populations in other habitats?

5.3 Materials and methods

5.3.1 Long-term butterfly monitoring sites and landscape context

Abundance data from 540 long-term monitoring sites of the United Kingdom Butterfly Monitoring Scheme (UKBMS) were used to investigate *M. jurtina* phenology and determine the effects of abundance on phenology. The UKBMS sites were selected if they had both relevant Natural England priority habitat map and digital elevation data (see below). UKBMS data is collected by volunteers using the 'Pollard walk' method, in which volunteers conduct a line transect of 0.5 to 3km, recording all butterflies that occur within a moving 5m x 5m x 5m box in front of the recorder (Pollard & Yates 1993). The UKBMS uses a two-step method, using these data to fit Generalised Additive Models which produce fitted weekly counts and an overall collated annual index of abundance at each site (Dennis *et al.* 2013).

To quantify local site characteristics and capture the focal habitat within survey areas, we analysed a 500m radius buffer around the centroid of each of the 540 UKBMS sites, using data from the Natural England priority habitat maps (Natural England 2019). These maps capture a range of habitat characteristics, including lowland calcareous grassland (chalk grassland). Using a 50m resolution digital elevation map (Morris & Flavin 1990), topographic slope angles were estimated for 539 UKBMS sites, using a systematic sampling of points at 250m intervals within the 500m radii of the site centroids, as described in Oliver *et al.* (2010). It should be noted that site steepness is positively correlated with increased variation in slope angles i.e. areas with steeper slopes are also more topographically variable (see Appendix F, Tables F1 – F3).

For the population genetics analyses distinct categories of sites were required. Sites were defined as either chalk or non-chalk sites based upon the presence of lowland calcareous grassland. The lowest percentage cover was 4.7% at Dancersend. Whilst this represents a small percentage of the total site it is worth noting that few sites, across all UKBMS sites where lowland calcareous grassland is present are dominated (>50% cover) by lowland calcareous grassland and that 25% of these sites (n = 70) have less than 4.3% cover. All of the chalk sites used in the analysis fall within the interquartile range of chalk cover across all UKBMS sites.

5.3.2 Drought tolerance experiment

All drought experimentation was carried out by Melanie Gibbs from the UK Centre for Ecology & Hydrology (UKCEH) following the methodology described in Gibbs *et al.* (2012). A summary of the methods are provided here. Potted host plants (*Poa trivialis*) were grown under standard conditions, with each plant watered via individual trays. Once mature, plants were randomly assigned into the treatment groups drought-stressed or control. Control plants were watered daily from 20 days prior to larval hatching and then throughout the experiment. Plants were never oversaturated with watered, but enough to prevent soil drying and wilting. Drought-stressed plants received no water from 20 days prior to larval hatching and were then only watered every six days throughout the experiment. This treatment ensured that, at all stages of the experiment, green leaves were available for plant, but ensured moderate drought stress occurred. At the end

of the experiment green leaves were still present on all plants. This ensured that food availability was not a factor limiting larval growth and survival. Rainwater was used in both treatments.

A total of 324 newly hatched *M. jurtina* larvae were collected from populations originating from nine of the 15 sites used in the molecular analysis (Fig. 5.1). Adults from these source populations were live captured and mated with individuals from the same population and eggs were collected. In a common garden experiment, 12 newly hatched larvae from each source population were raised on three non-drought stressed (control) host plants (four larvae, originating from the same source population, per plant) and 24 larvae were raised on six drought stressed host plants (four larvae, originating from the same source population, per plant) under controlled conditions until eclosion, using the methods described in Gibbs *et al.* (2012). A higher number of larvae were raised on drought stressed plants due to an expected higher mortality rate (see Talloen *et al.* 2004), totalling 108 and 216 larvae on control and drought stressed host plants respectively. *M. jurtina* overwinter as small larvae, during which little growth occurs (Brakefield 1984). As such larvae were monitored at three time points: 49 days after the first larval hatch date (pre-overwintering), 162 days after hatching (post overwintering during larval growth) and 309 days after hatching (late larval growth and pupation phase). Larval masses (g) were recorded for all individuals that survived up to the second monitoring point and the number of larvae that survived until the third monitoring point were recorded. Individuals were monitored until they reached pre-pupa stage, at which point they were removed.

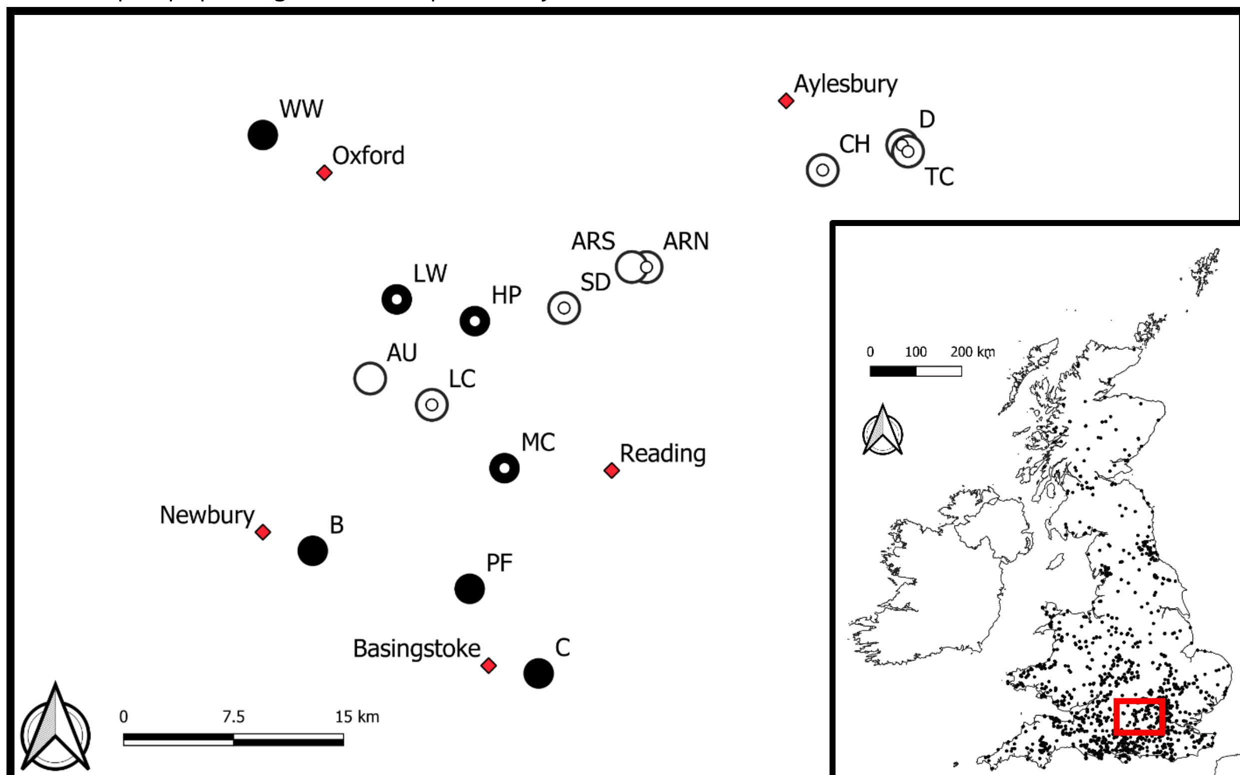


Figure 5.1 Fifteen sites around the Chiltern Hills from which Maniola jurtina samples were collected for genetic analysis. Large black circles signify where no chalk grassland was present within the 500m radius of each site (n = 7), large white circles signify sites where chalk grassland occurred within a 500m radius of the site centroid (n = 8; percentage cover 4.7-21%). Circles with smaller

dots at the centre were sites from which individuals were also collected for the drought experiment (n = 9). Main towns are marked with red diamonds. Inset map shows the locations of 517 UKBMS transect sites used in the phenology analysis,

5.3.3 Molecular analysis

We conducted a molecular analysis of 287 *M. jurtina* individuals sampled from 15 of the 540 UKBMS sites around the Chiltern Hills in the south of England in 2017 (Fig. 5.1). To assess how landscape factors affect gene flow, distances between sites ranged from 0.8km to 62km, and intervening landscape encompassed urban areas, arable farmland, woodland and semi-natural habitats. DNA extraction and analysis methods are described in detail in Chapter 3 and summarised here. For further information regarding novel primer development and evaluation, and pre-existing primer evaluation please see Appendices B and C respectively. DNA was extracted from a leg of each individual using prepGEM Universal DNA extraction kits (Zygem), following the recommended protocol for insects. Six microsatellite markers, isolated in Richard *et al.* (Richard *et al.* 2015), were used to genotype the samples: Mj4870, Mj7232, Mj7132, Mj5522, Mj5331 and Mj0247. DNA was amplified in two multiplex sets using the following reaction mixture: 1μL template DNA, 6.25μL QIAGEN multiplex PCR master mix (3 mM $MgCl_2$), 0.625μL tagged forward primer, 0.625μL reverse primer, 1.25μL QIAGEN Q solution, 2.25μL RNase-free water. Multiplex set 1 contained Mj7232, Mj5522 and Mj0247, all at 3μM. Multiplex set two contained Mj4870 at 1.5μM, Mj7132 at 5μM, and Mj5331 at 4.5μM. PCRs were carried out in an Eppendorf Mastercycler nexus eco with an initial denaturation for 15:00 at 95°C, followed by 40 cycles of 00:30 at 94°C, 01:30 at 56°C & 01:00 at 72°C, and a final extension 10:00 at 72°C. All PCR products were diluted by 100x and run on an Applied Biosystems 3730 DNA Analyser. Allele peaks were then scored by using GeneMarker® version 1.5 by SoftGenetics, using the microsatellite calibration settings.

5.3.4 Statistical analysis

5.3.4.1 Phenology

To calculate butterfly flight periods, all weekly fitted count values for *M. jurtina* abundance were summed per UKBMS site in each year and the day number of the recording period at which 10% of the total occurred was recorded as the flight period start date. The day at which 90% of the total occurred was recorded as the flight period end date. We used 10th and 90th percentiles to avoid the effect of outliers (Van Strien *et al.* 2008). Length of flight period was calculated as the number of days between these two values. The mean flight dates for each site per year were also recorded. A second dataset using the 1st and 99th percentiles as start and end dates was generated to determine what effect the trimming of the data might have on the analysis.

We fitted statistical models to understand whether the inferred geology (herein geology) and topography of the site predicted *M. jurtina* phenology (Equation 1). The four measures of timing for *M. jurtina* flight periods (start, mean and end dates of the flight period and length of flight period) were each fitted as response variables into separate linear mixed effects models, against the percentage cover of chalk grassland and mean slope angle of each site. The additional factors

abundance, northing (km north on Ordnance Survey grid), easting (km east on Ordinance Survey grid), mean site altitude and mean site aspect (cos((aspect x **π**)/180), such that 1 = due North, -1 = due South)) were included as fixed effects and site and year as random effects. We included annual abundance as a covariate in these models because larger populations are likely to have a greater flight period range due to mathematical mean-variance relationship (Taylor 1961). Northing was included in the model to account for the temperature gradient across the UK, with cooler average temperatures occurring at higher latitudes. This was necessary, firstly, because previous studies have shown that *M. jurtina* flight periods are shorter and begin later at northern latitudes (Brakefield 1987) and secondly because temperature has been shown to affect *M. jurtina* phenology, with a predicted 4.7 and 5.4 day advance to first appearance and peak flight dates respectively per 1°C increase (Roy & Sparks 2000). Easting was included to account for longitudinal differences in site conditions e.g. differing levels of rainfall which can affect butterfly phenology (Roy *et al.* 2001). Site altitude and aspect were included to account for the effects these two factors might have on local temperatures. To reduce the range of magnitudes across the data, northing was scaled by subtracting the mean from each value followed by dividing by the standard deviation. Site and year were included as random effects to account for repeated measures at each site and variation in phenology between years, often associated with weather (Roy & Sparks 2000).

All mixed effects models were carried out using the *lmer* function from the *lme4* package in R (Bates *et al.* 2015). Model assumptions were checked using diagnostic plots for all mixed effects models. Diagnostics from the initial model fits demonstrated that phenology at sites with very low abundances was much more variable, violating homoscedasticity. This is likely because at sites with very low abundances there is increased detectability-related sampling error, increasing the uncertainty of the phenology estimate (McCarthy *et al.* 2013). To overcome this problem, all sites with an abundance index value of less than 20 were removed from the analysis.

$$P = C + S + A + N + E + H + F + i + y + \varepsilon \quad [1]$$

Where P is the phenology metric of interest (either flight period start, mean, end day or range), C is the percentage cover of chalk grassland per site, S is the mean slope angle per site, , A is the site total abundance, N is the site northing, E is the site easting, H is the mean altitude per site, F is the mean aspect of each site, i is a random intercept for site, y is a random intercept for year and ε indicates error term with zero mean and normal distribution.

All models were tested for spatial autocorrelation via a Moran's I test. Residuals were extracted from each model and run against an inverse matrix of distance between sampling points using the *Moran.I* function from the *ape* package in R.

5.3.4.2 Drought tolerance

A generalised linear mixed effects model was used to determine whether larval survival rates varied between sites in association with site characteristics. The model was fitted with a binomial error structure and with host plant drought treatment and percentage chalk cover (geology) as fixed effects with an interaction term, and population as a random (Equation. 2). Slope angle was not included due to a 0.8 correlation with chalk cover.

$$S = T + G + T.G + p + \varepsilon \quad [2]$$

Where S is the larval survival rate, T is the treatment (drought/control), G is the geology of the origin site (percentage cover chalk grassland, p is a random intercept for the origin population of the larvae and ε indicates error term with zero mean and normal distribution. A series of model simplifications were carried out (removal of the interaction term, removal of geology variable and removal of treatment variable) and all versions of the model were compared using the *model.sel* function from the R package MuMIn.

5.3.4.3 Population genetics

Population genetics analyses are described in detail in Chapter 3 and summarised here, with specific tables and figures available in Appendix H. Measures of genetic diversity and differentiation (based on 287 individuals from 15 sites; Fig. 5.1), including Wright's F statistics, heterozygosity, allelic richness and effective population sizes were carried out using GenePop v4.7.0 (Rousset 2008), FSTAT v2.9.4 (Goudet 1994), Arlequin v3.5 (Excoffier & Lischer 2010), NeEstimator v2 (Do *et al.* 2014) and PopGenReport (Adamack & Gruber 2014), and can be found in Appendix H, Tables H1-H4.

Population structure was estimated using STRUCTURE v.2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2007), using an admixture model and correlated allele frequencies with a 100,000 burn-in and 1,000,000 MCMC replications per chain. The potential number of genetic clusters (K) was tested from one to six, with 20 chains run per K. The likeliest K within the sample sets was estimated using the programme STRUCTURE Harvester (Earl & VonHoldt 2012) and visualised using CLUMPAK (Kopelman *et al.* 2015). Four separate STRUCTURE runs were conducted i) all individuals allocated by the population from which they were sampled (15 populations, n = 287), ii) all chalk site and all non-chalk sites grouped into two populations (n = 137 and 150 respectively), iii) only the individuals from the eight chalk sites (n = 137), and iv) only the individuals from the seven non-chalk sites (n =150).

Individuals were then pooled by site to generate allele frequencies for genetic distance analysis. Weir and Cockerham pairwise F_{ST} values were calculated using Genepop. Mean allelic richness across all loci for each site was calculated using FSTAT. A Mann-Whitney U test was carried out to compare the allelic richness of individuals on chalk compared to non-chalk sites. Pairwise F_{ST} values were calculated for each site pair combination, with each combination assigned into one of three categories based upon the individual geologies of the two sites: a) both chalk, b) both non-chalk, c) one chalk and the other non-chalk. Pairwise F_{ST} values were fitted into a linear regression with geology and Euclidean distance between sites as a fixed effect (Equation 3). Slope angles were not included in the equation due to the high corelation (0.75) between chalk cover and slope angle.

$$F = G + D + \varepsilon \quad [3]$$

Where F is the pairwise F_{ST} score between each pair of sites, G is the site geology (chalk/non-chalk), D is the Euclidian distance between sites and ε indicates error term with zero mean and normal distribution.

As pairwise F_{ST} values between sites are not independent, Mantel randomisation tests with 999 permutations were conducted to assess whether the predictor variable (geology) was significant following the methodology described in Powney *et al.* (2012). The number of significantly different groupings within site type pairs was determined via a Tukey HSD test.

5.4 Results

5.4.1 Phenology

All phenology measures were positively significantly associated with differences in chalk cover (start 0.07, $p = 0.009$; mean 0.14, $p < 0.001$; end 0.19, $p < 0.001$; range 0.13, $p < 0.001$ Fig. 5.2, Appendix G, Table G5), and mean slope angle (start 0.36, $p < 0.001$; mean 0.62, $p < 0.001$; end 0.81, $p < 0.001$; range 0.43, $p < 0.001$ Fig. 5.3, Appendix G, Table G5)) i.e. average flight period dates were later on sites with greater levels of chalk cover or steeper slope angles and average flight periods were longer on sites with greater levels of chalk cover or steeper slope angles. Northing and abundance were also significantly associated with all four measures of phenology, with two exceptions: i) northing was not associated with flight period mean date and ii) mean local abundance was not associated with flight period end date (Appendix G, Table G1). Estimated model values for Equation 1 regarding abundance and northing can be found in Figs. G3 & G4. Easting, aspect and altitude were not significantly associated with any measure of phenology, however aspect and altitude were both marginally significantly associated with flight period range ($p = 0.08$ and $p = 0.09$ respectively).

The results when using the 1st and 99th percentiles as start and end flight dates were remarkably consistent with those using the main dataset. The only variable that was significantly affected was flight period start dates, which were no longer significantly associated with percentage chalk cover (Appendix G, Table G2).

The residuals from each model showed no evidence of spatial autocorrelation using a Moran's I test (start day model I: Observed (O) = 0.001, Expected (E) = - 0.0001, s.d. = 0.001, $p = 0.259$; mean day model I: O = 0.0007, E = -0.0002, s.d. = 0.001, $p = 0.476$; end day model I: O = 0.0008, E = -0.0002, s.d. = 0.001, $p = 0.421$; range model I: O = 0.001, E = -0.0002, s.d. = 0.001, $p = 0.277$).

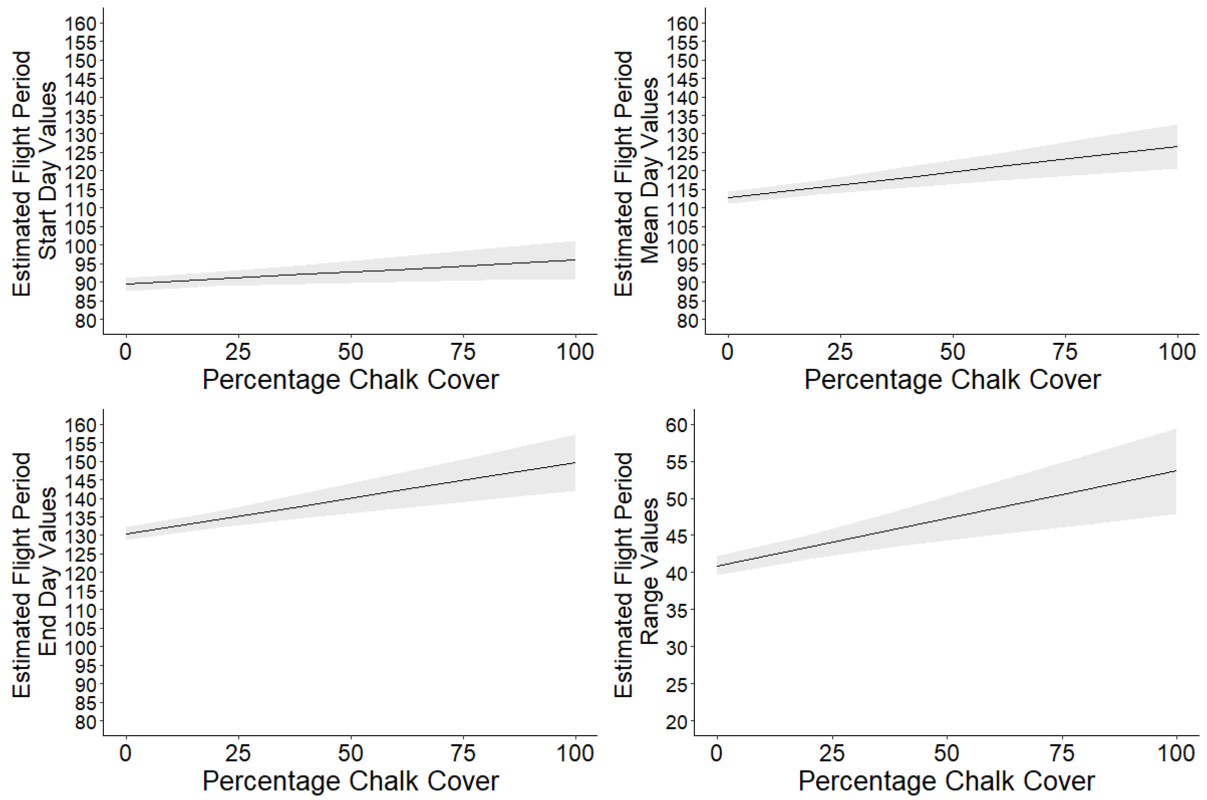


Figure 5.2 Estimated model values from equation 1 for four measures of phenology for *M. jurtina* in relation to percentage cover of chalk.

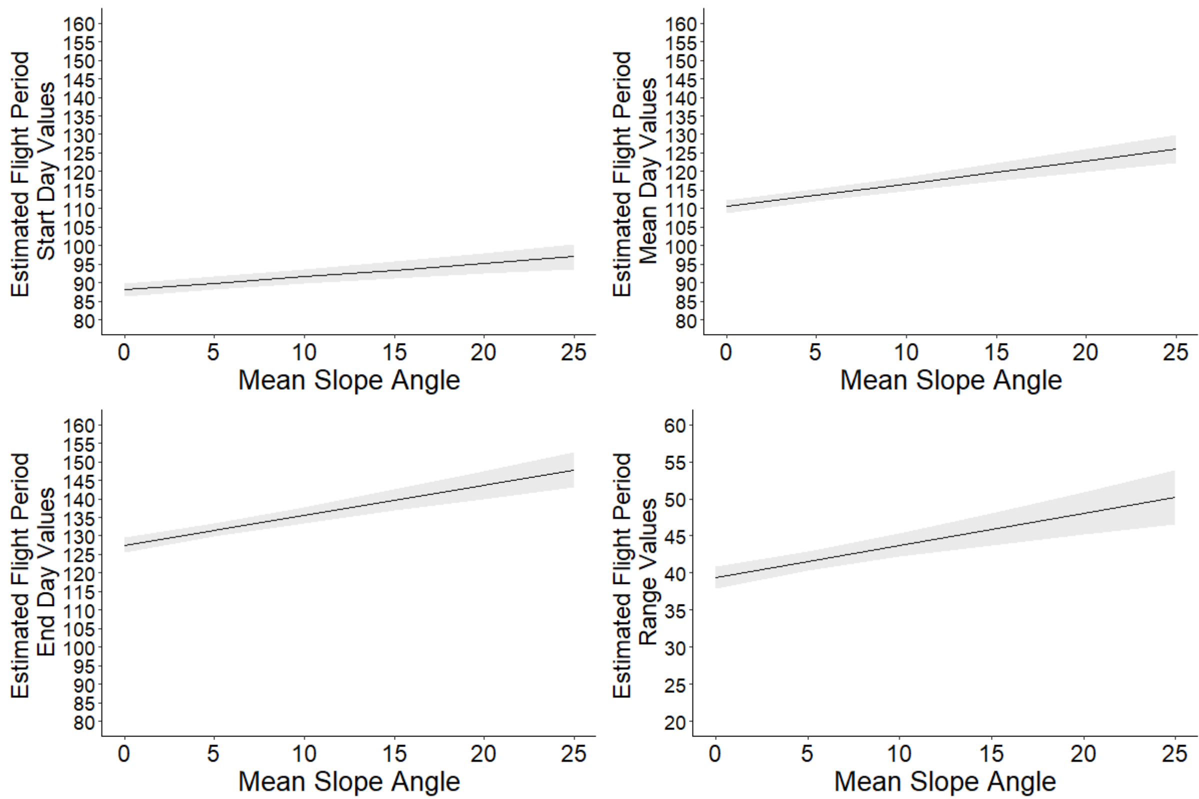


Figure 5.3 Estimated model values from equation 1 for four measures of phenology for *M. jurtina* in relation to site mean slope angle.

5.4.3 Drought tolerance

Model simplification determined that the best fitting model did not include chalk cover as a fixed effect (AICc 377.9 vs 339.5 (treatment and geology additive), 341.4 (treatment and geology interaction) and 346.1 (geology only)). I.e. larval survival rates were significantly affected by host plant drought treatment (Intercept: estimate = 0.78, se = 0.33, z-value = 2.36, p = 0.018; Drought: estimate = -0.84, se = 0.28, z-value = 2.97, p= 0.003) (Fig. 5.4), but chalk cover had no effect on larval survival rates (Fig 5.4).

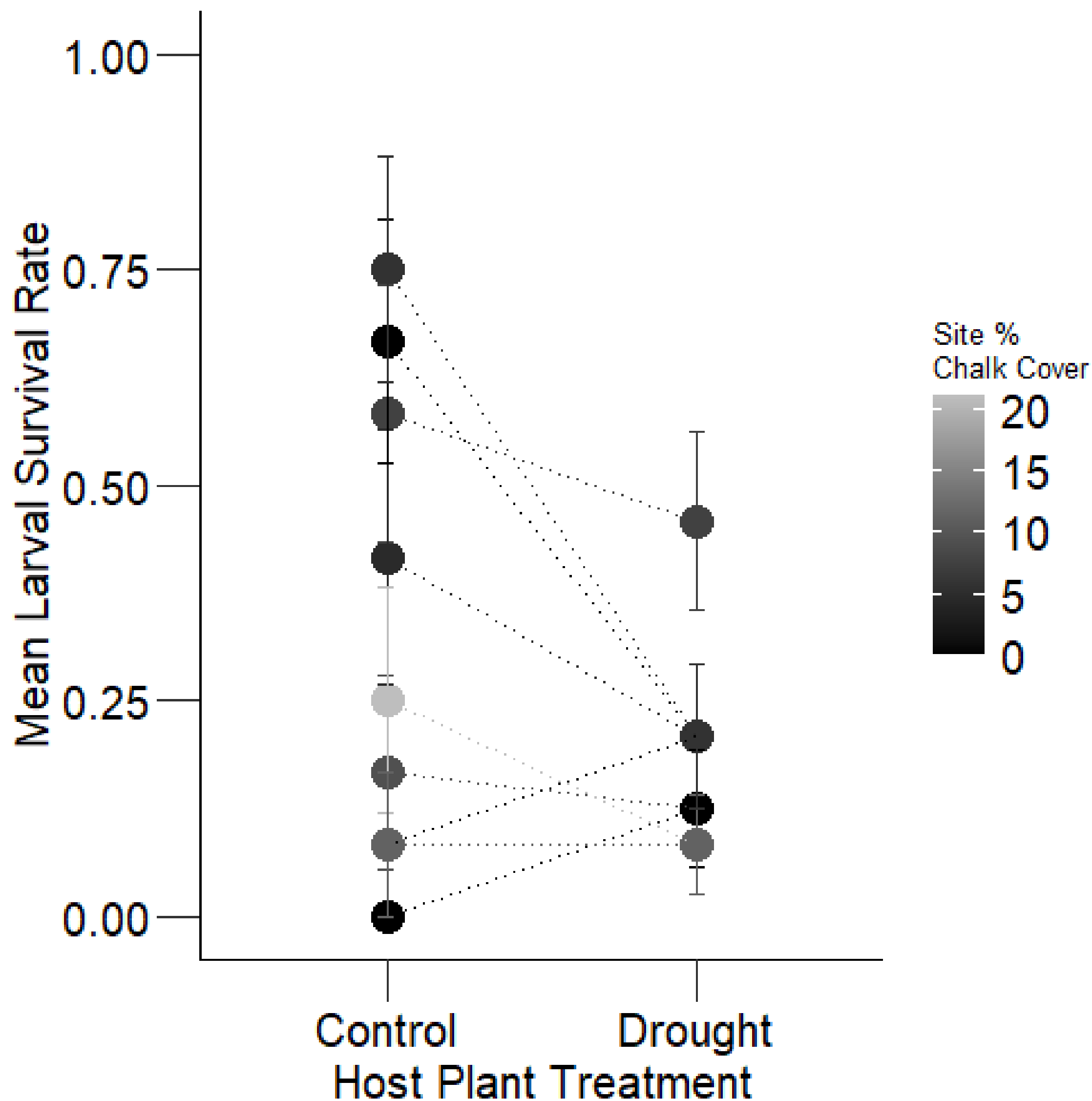


Figure 5.4 Mean survival rates of *M. jurtina* larvae when reared on control and drought-stressed host plants. Populations are coloured by percentage chalk cover at each site, however chalk cover had no significant effect on larval survival rates.

5.4.3 Population genetics

All populations within the 15 sites in southern England displayed high levels of genetic diversity and low levels of genetic differentiation (Appendix H, Tables H1 – H5, Fig. H1). In summary, no linkage disequilibrium occurred between any pair of loci (Table H1). Null allele frequencies were <0.2 for all site loci combinations except for Mj4870 at ARS (Table H2). The microsatellites used displayed a high level of variability (H_O = 0.279 to 0.902) and no locus displayed significant heterozygote excess or deficit (Table H3). No F_{ST} values per locus were significantly different from zero, however, F_{IS} values were significant at four of the six loci (Table H4). All populations displayed a high level of heterozygosity, with high levels of allelic richness and infinite estimated effective population sizes (Table H10). Allelic richness was not significantly affected by site geology (p =

0.867), with a mean allelic richness of 8.2 for sites with chalk grassland present and 8.3 for non-chalk sites (Fig. 5.5).

Pairwise F_{ST} scores between pairs of sites were extremely low (mean = 0.002, variance = 0.00004) and none were significantly greater than zero (Table H5). However, when site pairs were grouped by geology (i.e. chalk & chalk, non-chalk and non-chalk, chalk and non-chalk), combinations within site pairs had a significant effect on pairwise F_{ST} (Table 5.1, Fig. 5.5), indicating evidence of weak population differentiation. Distance between sites had no effect on pairwise F_{ST} (Table 5.1). No evidence of population structure was found between these 15 populations. No population was found to be strongly genetically distinct from any other population, regardless of the number or allocation of sites included in the analysis (Appendix H, Fig. H1).

Table 5.1 Effects of site pair geology and distance between sites on pairwise F_{ST} (Equation 3)

Model	Response	Factor	Degrees of Freedom	Sum of Squares	Mean Square	F value	p-value
Equation 3	F_{ST}	Geology	2	0.0003	0.0005	3.8105	**0.025**
Equation 3	F_{ST}	Distance	1	0.000009	0.000009	0.243	0.623
Equation 3	F_{ST}	Residuals	101	0.004	0.00004	NA	NA

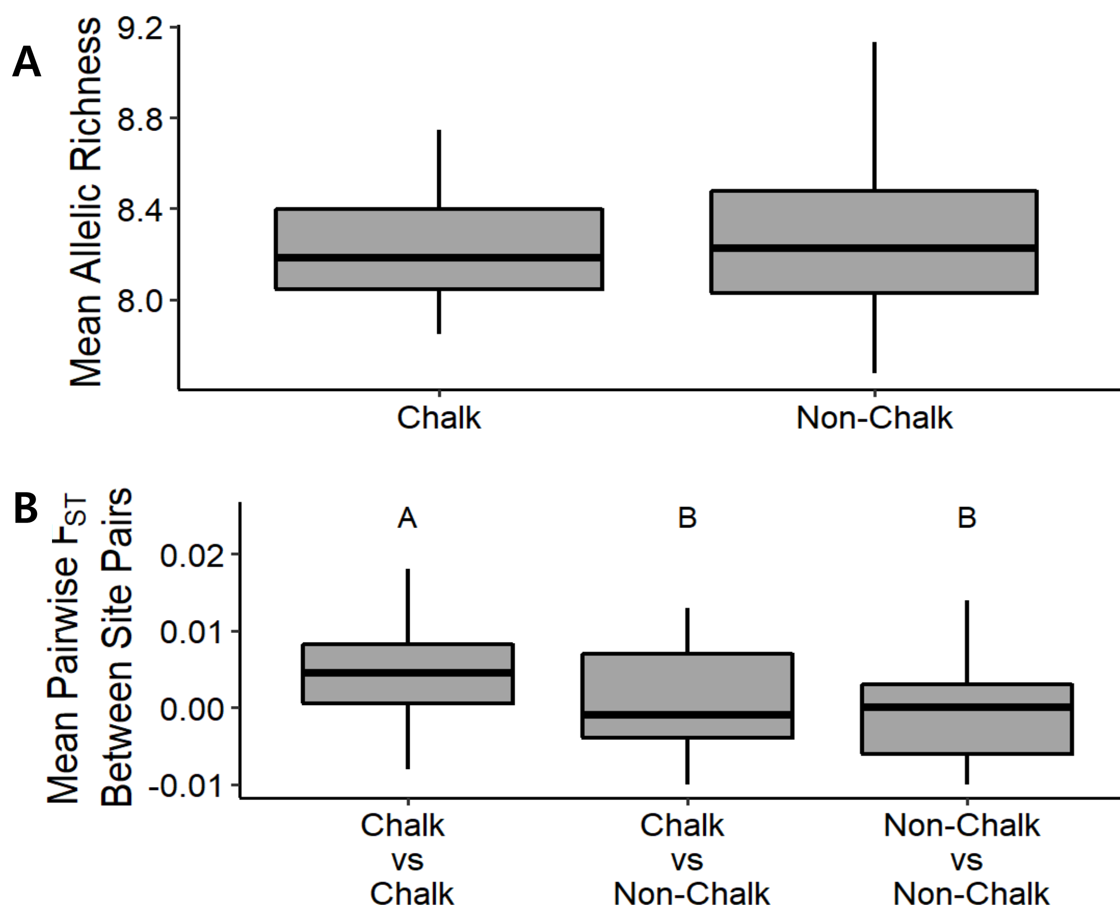


Figure 5.5 The effects of site geology vs two measures of genetic diversity: (a) allelic richness (b) pairwise F_{ST}, letters indicate significance groupings.

5.5 Discussion

In this study, we quantified characteristics of *M. jurtina* flight periods with respect to geology and topography. We also determined whether differences in other aspects of ecology (population genetics and drought tolerance) were also associated with the same landscape attributes. We found significant, positive, associations between the phenology of *M. jurtina* and geology (chalk grassland) and topography (steepness of sites being a general proxy for topographical heterogeneity), i.e. key flight dates are delayed with increasing chalk cover and slope angle. These associations remained after accounting for abundance, i.e. aspects of geology and topography are associated with phenology independent of mean local abundance. We found no strong evidence of genetic structuring of *M. jurtina* populations linked to geology, and only very weak evidence of genetic differentiation among populations. Finally, we found no effect of geology on larval survival (drought response).

Microclimatic heterogeneity may explain the longer flight periods on steeper (more topographically diverse) chalk grasslands. Habitat and topographic diversity can allow species to persist in areas of suitable microclimate when the surrounding climate is no longer favourable (Bennie *et al.* 2008) and habitat heterogeneity has been promoted as a method of improving species resilience under climate change (Crick *et al.* 2020). For example, south-facing chalk grassland hillsides were found to harbour populations of the warmth loving species the silver-spotted skipper (*Hesperia comma*), absent from other habitat types (Davies *et al.* 2006). However increasing ambient temperatures at sites due to climate change has seen a change in the local distributions of this species more recently (Wilson *et al.* 2010; Lawson *et al.* 2014). If microclimate heterogeneity alone causes the longer flight periods, we might expect to see a two-tailed expansion to the flight period on steep chalk sites, with suitable habitat patches available earlier as well as later in the year.

In contrast, we found that all measures of phenology were positively associated with chalk cover, including start date. This means that sites with more chalk have later start dates, creating a long, single-tailed extension to the flight period later into the season. Similar results were found regarding topography. These results indicate that phenology differences are likely not a simple effect of either warmer conditions or the heterogeneous nature of chalk grassland sites and the range of microclimates available (Diacon-Bolli *et al.* 2012). However, it should be noted that when the 1st percentile was used as flight start dates no significant association with chalk was found. This may support the microclimate heterogeneity hypothesis but should be interpreted with caution.

Flight start dates are typically a result of the effects of spring temperatures on larval development (Roy & Sparks 2000), therefore drought conditions on steep chalk grassland sites may additionally impact larval development and hence, adult phenology. Drought conditions have been shown to lead to lengthened larval development times and later emergence dates, in species such as the speckled wood, as a result of physiological stress (Gibbs *et al.* 2012). In habitats with heterogeneous microclimates, such as hilly chalk grasslands, certain microhabitats (e.g. with thinner soils on south-facing slopes) may lead to host plants becoming particularly drought stressed. This would result in a certain proportion of individuals at a site with delayed emergences, and a more

protracted flight period overall, but one that is single-tailed. One point that should be addressed is that the fixed effects in the models account for relatively little variation within the data (7-15%) and the majority of variation (46-77%) is explained by the random effects for site and year. This is unsurprising as year captures weather effects, which are known to have a large effect on butterfly phenology (Roy *et al.* 2001; Mills *et al.* 2017).

A limitation of this study is that UKBMS data do not fully encompass the flight period of *M. jurtina.* Protracted flight periods into October have been reported on these southern chalk grassland sites (Thomas & Lewington 2010), whilst UKBMS recording runs from the start of April until the end of September. Therefore, it is feasible that flight period end dates are later than those used in this analysis. Additionally, an assumption of this work is that sites with some percentage cover of chalk have similar microclimates. This is not unreasonable to some extent, however plant assemblages at different sites could well vary, leading to different microclimates and temperature extremes (Suggitt *et al.* 2011). Such differences in local vegetation characteristics could be having effects that are not accounted for in this analysis, as local microclimates due to vegetation structure have been shown to affect butterfly ecology (Suggitt *et al.* 2012).

Our molecular analysis results support those of Richard *et al.* (2015) and Villemey *et al.* (2016) in finding high levels of genetic diversity within *M.jurtina* populations and low levels of genetic divergence between populations using microsatellite markers. These results are consistent with other studies though not directly comparable due to the use of differing techniques (allozymes and AFLPs) (Thomson 1987; Goulson 1993a; Schmitt *et al.* 2005; Habel *et al.* 2009; Baxter *et al.* 2017). Despite being statistically significant, the differences in genetic differentiation between site types (as indicated by pairwise F_{ST} scores) are extremely low, being below the 0.05% threshold typically viewed as indicative of genetic differentiation (Freeland *et al.* 2011). This suggests that populations on chalk sites are marginally more genetically distinct from populations on other chalk sites than they are from populations in the surrounding environment. Additionally, no population structuring was found via any combination of sites. Therefore, it appears that all populations included in the study belong to a single, large population with properties similar to one at panmixia with random mating. Very low levels of differentiation are present, although insufficient to have any great effect on population structuring. The suggestion that populations of *M. jurtina* on chalk grasslands form a distinct genetic race is not supported; in fact, the opposite is found, with populations on chalk sites being more distinct from each other, although these levels of differentiation are very low. Therefore, we conclude that the differential phenology associated with geology and topography found in this study is unlikely to be explained by differentially adapted host races. However, it should be noted that due to the high correlation found between chalk percentage cover and site steepness we cannot determine the effect of topography with this experimental set-up. Therefore, caution in interpretation is required as our other analyses have shown that site topography can have an effect on aspects of *M. jurtina* ecology.

Contrary to our expectations we found no association between the percentage chalk cover from source sites and larval survival when exposed to drought conditions. However, these results should be interpreted with caution owing to the relatively small sample size and spatial scale of the

analysis, and the fact that slope could not be included in the drought models despite affecting phenology. Additionally, in wild situations larvae would be able to move from plant to plant, ensuring that a sufficient quantity of food could be consumed. In the experimental setup larvae were constrained to single pots containing food plants and therefore unable to move to fresh sites. However, sufficient green plant material was available throughout the experiment and remained at the end to ensure that food quantity was not a limiting factor in larval growth. Our results suggest that whilst drought conditions reduce larval survival rates, the effects are not mitigated by local adaptation specific to chalk sites.

In conclusion, we find butterfly phenology varied at the national scale with geology and topography. We find no evidence of genetic structuring of populations based upon these site conditions, nor any differences in drought tolerance. Future research may benefit from a detailed analysis of factors influencing phenology and the potential for local adaptation; for example, slope aspect, microclimate, vegetation cover, and habitat management (Brakefield 1987; Bennie *et al.* 2006; van Noordwijk *et al.* 2012). Such studies will become increasingly important for understanding and predicting species responses to a rapidly changing climate.

Chapter 6. General discussion

Biodiversity loss has a detrimental effect on humanity (Díaz *et al.* 2006; Cardinale *et al.* 2012). To understand the magnitude of these effects requires, amongst other things, an understanding of the levels of loss. The need for biodiversity monitoring is clear; the more we know about ecosystems the more we can do to protect them. Fortunately, this is something that is gaining increasing recognition, with an increasing number of monitoring schemes being set up over time (Pocock *et al.* 2015, 2017b; Hayhow *et al.* 2016). Despite this recognition there remain some major understudied areas. One such overlooked subject is genetic diversity.

All aspects of biodiversity are underpinned by genetic diversity (Bruford *et al.* 2017), yet genetic diversity monitoring schemes are few and far between outside of domestic livestock or socioeconomically important wild species (Laikre 2010; Laikre *et al.* 2020). Consequently, the monitoring of genetic diversity should be seen as a conservation priority and should be coupled with the more traditional diversity monitoring that already occurs.

This thesis highlights three ways in which monitoring data can be applied to answer ecological questions and aid conservation. Firstly, through the novel application of existing monitoring datasets, secondly through the creation of a novel genetic monitoring dataset, and finally by demonstrating how monitoring data can be used in conjunction with other available data to investigate species' ecology to a greater degree than would be possible using each dataset independently. Throughout this thesis all studies have focused on butterflies. This is a combined result of data availability and the value of butterflies as indicators of insect biodiversity (Thomas 2005).

In this final chapter I review each of the previous chapters and discuss their implications in the context of both biodiversity monitoring, and conservation more generally. I further discuss the importance of genetic diversity and the need for monitoring. This thesis both highlights the need for an increase in genetic diversity monitoring schemes, but also demonstrates the relative ease at which this can be accomplished and the further applications to which the data can be used.

6.1 Thesis overview and implications

The importance of biodiversity monitoring for answering ecological questions and informing conservation actions was introduced in **Chapter 1**. A number of research gaps were also addressed, the most pressing of which being a general lack of genetic diversity monitoring (Laikre 2010; Laikre *et al.* 2020). The application of monitoring data to novel methods and the combination of monitoring data with alternative data types e.g. experimental datasets, were also introduced. In this chapter I explained the importance of genetic diversity for conservation and species persistence, and why the general lack of monitoring for wild species is problematic and short sighted.

In Chapter 1, the model species for the majority of this thesis (**Chapters 3-5**), the meadow brown, *Maniola jurtina*, was also introduced. As one of the best studied and most abundant butterflies in Europe, this species provided an ideal subject for the studies reported here. Despite the large number of studies on *M. jurtina* over the last century, including investigations into genetic

diversity, none have monitored genetic diversity, nor combined genetic data with both population and experimental data. Therefore, the work presented in this thesis adds valuable content to the already considerable literature on this species.

In **Chapter 2** I demonstrated the application of long-term population monitoring data to a novel method for predicting the resilience of ecosystem functions. This chapter represents a significant milestone in this area of research, as it offers a new approach to answering a question that had reached an impasse for the past 20 years (Díaz & Cabido 1997; Funk *et al.* 2017), i.e. the ability to predict how ecological communities respond to environmental change and what effect this has on ecosystem functioning.

By correlating the long-term population dynamics of all possible pairs of butterfly species within the UKBMS, I created a matrix of similarity in response to environmental change. This matrix was then used to build a dendrogram allowing the visualisation of similarities in species' population dynamics. Species could then be allocated into response guilds, with all species within a guild responding more similarly to the environment than to species outside of the guild. The distribution of specific ecosystem functions could then be mapped onto the dendrogram to determine whether or not they fell into aggregated clusters or were evenly spread across the butterfly community. These assertions were tested empirically using Mantel tests. Ecological theory suggests that the more species within an ecosystem, the more stable an ecosystem function should be as a result of redundancy, i.e. species performing similar functions (Oliver *et al.* 2015). This is because a greater range of species have a greater range of population dynamics, therefore even if some species decline, some functioning will still occur through portfolio effects (Tilman 1999). However, where all species have similar population dynamics, e.g. those within the same response guild, these species effectively respond as one. Therefore, if an ecosystem function is wholly provided by species within a single response guild it is far less stable than one which is provided by species from multiple guilds. Although all proxies for ecosystem functions used in this chapter were clustered evenly, the methodology shown could be applicable to any ecosystem service with quantifiable values. Furthermore, the methodology demonstrated should be applicable to any of the increasing number of taxonomical group or species assemblages for which long-term monitoring data are available.

In **Chapter 3** I described the creation of a new genetic diversity monitoring scheme for the meadow brown butterfly *M. jurtina.* Existing microsatellite markers, characterised by Richard *et al.* (2015), were used to measure levels of genetic diversity across fifteen populations of *M. jurtina* in southern England over a period of eight years. Although additional microsatellite isolation was attempted, no loci were successfully characterised, highlighting the complications of microsatellite isolation in Lepidoptera (Zhang 2004; Sinama *et al.* 2011). The results of the study showed high levels of genetic diversity across the study landscape, coupled with low genetic differentiation between populations, indicating a single large population with panmictic properties. These results were likely due, partly to the low number of microsatellites used, but also to the high abundance of *M. jurtina* across the study landscape and the high availability of its host plants, resulting in minimal barriers to gene flow. I also showed that across all fifteen study sites there was minimal change in the levels of genetic diversity or divergence after a period of five years. Additionally, for

a subsample of three sites, yearly changes in genetic diversity were extremely low over eight consecutive years. These results suggest that the genetic diversity of *M. jurtina* within the study landscape is stable, however continued monitoring is required to develop a greater understanding.

In **Chapter 4** I built upon the data collected for the previous chapter and investigated the genetic diversity of *M. jurtina* across Europe. By collaborating with a number of institutions across Europe an analysis was carried out using samples from eleven countries, spanning a total range of 2525 km. Using the same microsatellite markers as used in the previous chapter genetic diversity was shown to be high across the continent, however there was a greater level of differentiation across mainland Europe than found in the UK. The results of this chapter suggest that contemporary gene flow between the UK and Europe is restricted, resulting in multiple clusters of *M. jurtina* across the continent. No evidence of isolation by distance was found, however this is likely to be the result of an uneven sampling design, as demonstrated by the removal of the most distant sites. Additionally, no evidence was found to support the idea that the differentiation of the UK populations was due to a recent population bottleneck. Despite only being for a single year, the methodology employed could easily be repeated, allowing the creation of a temporal dataset across Europe, as seen in Chapter 3, but over a far larger scale. The inclusion of a greater number of sites, resulting in more evenly distributed sample areas would greatly improve the results of future studies.

Finally, in **Chapter 5** I return to the use of long-term monitoring data. I combine this with *M. jurtina* genetic data, resulting from the genetic diversity monitoring scheme, and independent experimental data. This chapter demonstrates the range of applications that different forms of monitoring data can be applied to and how experimental data can easily be used in conjunction with monitoring data to answer ecological questions. Using the long-term population monitoring data, I show that *M. jurtina* phenology is affected by both geology and topography. Genetic and experimental data are then used to determine whether these differences in phenology are associated with genetic structuring of populations of differential tolerance to drought conditions. While there is some evidence for local adaptation, this is independent of geology and topography, indicating that other local environmental factors may be mediating *M. jurtina* selection. I hypothesise that differences in phenology are likely the result of differences in habitats, with chalk grasslands typically being warmer, more heterogeneous environments (Mortimer *et al.* 1998; Diacon-Bolli *et al.* 2012). This heterogeneity may result in the overall extension of flight period seen on chalk sites, through different patches becoming climatically suitable at different times.

6.2 Limitations

Although all four chapters vary in terms of analysis, they all suffer from the same basic limitation: data availability. The limitations of the method proposed in **Chapter 2** are that there are relatively few datasets to which it can currently be applied (e.g. birds, butterflies, moths (Morecroft *et al.* 2009)) and that there is often insufficient data regarding which species contribute to the delivery of specific ecosystem functions. There is no easy fix to either of these problems; however the first should lessen over time, as an increasing number of monitoring schemes are set up and start to provide usable data (Pocock *et al.* 2015; Hayhow *et al.* 2016). Furthermore, increasing participation

in citizen science schemes (Pocock *et al.* 2017b) and the ability to use species records as a proxy for time series data (Mason *et al.* 2018) should mean that this limitation is only temporary. The second limitation is harder to address as it requires detailed study of specific species and detailed measurements of ecosystem functioning. However, with ecosystem services valued at over $125 trillion each year (Costanza *et al.* 2014) and increasing recognition being given to the importance of such services (Millennium Ecosystem Assessment 2005; Geijzendorffer & Roche 2013), it is likely that such research will increase in the future, particularly for economically important functions such as pollination (Hanley *et al.* 2015).

Further limitations should also be noted for **Chapter 2**. Firstly, whilst no longer reliant upon a series of correlations between effects and response traits, some analyses using this method will still be reliant upon the collection of trait data. For example, the proxy for food provisioning to higher trophic levels was based upon the effect trait larval mass. In such instances this method does not provide a less data intensive approach, but does provide a simpler, more easily applicable one. However, there are some instances where large amounts of trait data will not be required. The best example of this are the pollination proxies used. For example, detailed knowledge of which bee species visit certain crop plants is already known (Hutchinson *et al.* In prep). Therefore, without specific trait data such as body size, or tongue length, we can still determine the stability of this function based upon which species visit specific crops and how their population dynamics are correlated.

Another limitation in **Chapter 2** are the proxies used, which are extremely basic and do not to stand up to much ecological scrutiny. For example, the proxy used as food provisioning for higher trophic levels is purely based upon larval mass and abundance. Whilst these factors do influence provisioning to higher trophic levels other factors such as population densities, behaviours and environmental interactions (including crypsis and toxicity) also play a major role (Holling 1961). Whilst this needs to be addressed it does not detract from the methods described. Simplistic proxies were chosen to highlight the method rather than answer the questions themselves and the results found in the chapter should not be interpreted as anything other than a demonstration of the method. There is an argument that instead of unrealistic proxies random numbers should have been used to avoid confusion. Whilst this would have been valid and demonstrated the method equally well, it is hoped that using proxies provides clearer understanding of how the method might be applied with real world examples.

Chapters 3, **4** and **5** all share a specific limitation; the low number of microsatellites employed in the study, limiting the power and resolution of the analysis (Selkoe & Toonen 2006). As has been discussed at length in **Chapters 1, 3** and **4** microsatellite isolation in Lepidoptera is challenging (Meglécz *et al.* 2004; Zhang 2004; Tay *et al.* 2010; Sinama *et al.* 2011). Of the fifteen loci reported by Richard *et al.* (2015) only six were found to be suitably robust for our population studies. Although attempts to characterise additional loci were unsuccessful the six loci provide, with minor exceptions, remarkably consistent results. Therefore, though the resolution may be low, the results are likely to be indicative of the true population structure and genetic diversity of the

studied populations. As a result, I have a good deal of confidence that more loci would probably add further detail, but not radically change any of the results.

A further limitation of **Chapters 3 and 4** are the uneven distribution of sampling sites. This is particularly apparent in **Chapter 4**. The range of distances between sites (from kilometres to hundreds of kilometres) influences certain statistical tests and meant that others could not be applied. For example, when the most distant sites were removed, reducing the variability between site spacing, a significant isolation by distance effect occurred. In theory this problem should be easily rectifiable in future studies, with the sampling of more sites, filling in the gaps between current sites. An ideal sample design would have seen sites regularly spaced across a grid. This would have increased the confidence in tests such as isolation by distance, which assume equal distances between sites. As UKBMS monitoring sites are selected by volunteers, the distribution of sites across the UK is not regular. The same issue occurs across Europe, with the added problem that not all countries have monitoring schemes set up, leading to large gaps in the data. A more evenly distributed sampling scheme would have been most beneficial in the resistance analysis carried out in Chapter 3. In this analysis the levels of resistance to geneflow between sites was estimated, in order to determine how landscape affects *M. jurtina* genetic diversity. As well as a more even distribution of sampling sites, a gradient of habitat fragmentation would also have benefited the study, as it would have helped to determine whether resistance values were the result of the habitats present in an environment, or the distances between patches of habitat. The resistance analysis found no evidence of isolation by resistance in the study landscape. Carrying out the same analysis along a gradient of fragmentation may result in changes in resistance, but the ability to carry out such an analysis is currently unfeasible due to the restrictions on site selection described above.

Chapter 5 also suffers from uneven sampling. A total of 540 sites were used for the phenological analysis, compared to fifteen sites for the genetic analysis and only eight sites for the experimental drought conditions. This unbalance becomes more apparent when sites are defined by their geology, with 123, eight and five chalk sites and 417, seven and three non-chalk sites for each respective analysis. Further sampling to increase both the genetic and experimental data used would be beneficial and allow more powerful statistical tests to be used, however sufficient sampling to match the number of sites used in the phenological analysis would be unfeasible.

An additional limitation of this chapter is that no concrete conclusion can be drawn regarding whether differential responses are due to phenotypic plasticity, or high allelic diversity resulting in adaptation. Phenotypic plasticity (the expression of multiple phenotypes by a single genotype (Bradshaw 1965) could allow a population to persist under different environmental conditions. Alternatively, in populations with high levels of allelic diversity, some individuals by chance may be better adapted to alternative environments (McGill *et al.* 2015). In theory it should be possible to tease these two explanations apart, as environmental persistence as a result of genetic diversity should lead to local adaptation and genetic differentiation. Therefore, as there was no evidence of local adaptation as a result of site geology it would be possible to conclude that phenotypic plasticity was the main driving force. However, this study specifically used a

heterogeneous environment, and as such selection pressures were not consistent throughout the population. High genetic diversity could still be the cause of differences in measures of phenology, as different genotypes may be suited to different microclimates.

6.3 Implications and applications

6.3.1 Predicting ecosystem function stability

The ability to understand how changes in species assemblages affect ecosystem functions is a priority of functional ecology (Díaz *et al.* 2013; Oliver *et al.* 2015; De Palma *et al.* 2017). In particular much research has gone into attempts to predict ecosystem functioning from species traits (Lavorel & Garnier 2002; Suding & Goldstein 2008; Funk *et al.* 2017). However, despite substantial research in the subject area the ability to predict changes to ecosystem functioning in real world scenarios is limited. This is largely a result of previous methods requiring detailed data on species response and effects traits. The method and analysis proposed in Chapter 2 offers a novel approach to this problem, one that is not dependent on such fine scale, detailed data.

Where sufficient long-term data are available this method should be easily applicable. Although not previously used to estimate ecosystem function resilience, the correlation of long-term trends is not a new concept (Siriwardena *et al.* 1998). As such, the relative ease at which these correlations can be calculated means that this method should be applicable to any long-term dataset. The challenge, therefore, will be in the calculation and quantification of ecosystem function delivery. One such group with relatively detailed information regarding an ecosystem function (crop pollination) is bees, with some species more important for crop pollination that others (Kleijn *et al.* 2015). Increasing amounts of data regarding bee distributions and abundances are becoming available e.g. the Bees, Wasps and Ants Recording Society (BWARS 2020). These population data, combined with a detailed knowledge of ecosystem function delivery, makes this an obvious area of future research, especially when the economic value of pollination is considered (Garratt *et al.* 2014).

6.3.2 Monitoring genetic diversity

The development of the pilot genetic diversity monitoring scheme for *M. jurtina* represents one of the very few examples of such a scheme for a wild species (Hutchinson *et al.* 2003; Hoffman & Blouin 2004; Nussey *et al.* 2005; Poulsen *et al.* 2006). As such it makes an important contribution to the field of genetic diversity monitoring, and makes some progress towards the CBD's Aichi Target 13 (Convention on Biological Diversity 2011a). Clearly a great deal of work with many more species is required before these targets are met, but the work in this thesis is a valuable step forward. As such, the most obvious area of future study in this thesis would be the continuation of the genetic monitoring scheme. At the very least, continued tissue samples should be taken and stored, ready to be analysed should time and funding allow. Currently the number of years in the study is insufficient for anything other than a general comparison of genetic diversity between years. Continued sampling and analysis would allow the implementation of robust statistical tests that are used with other long-term monitoring schemes e.g. Dennis *et al.* (2013). This would allow empirical determination of whether there are any significant changes in the levels of *M. jurtina* genetic

diversity over time. Additional microsatellite characterisation should also be considered a priority of any further study, as this would add statistical power to the results. However, as previously discussed, the characterisation of microsatellite loci in Lepidoptera is challenging (Zhang 2004). The implementation of alternative, next generation sequencing technologies such as SNPs or RADSeq could greatly increase the power and resolution of these studies. As all samples used in this study were retained, re-extraction of DNA for such purposes would be entirely possible although potentially costly and requiring specialist bioinformatics training (Davey & Blaxter 2010) .

The tissue archive, made up of over 2000 whole frozen *M. jurtina* specimens, holds a potential wealth of future research opportunities. Firstly, it provides an element of future proofing as DNA can be re-extracted for re-analysis using more powerful and modern molecular markers. Secondly, it provides opportunity for further morphometric studies, adding to the previous literature on genital morphology (Thomson 1973) and wing spotting (Brakefield 1979). The use of sophisticated imaging software e.g. Breuker *et al.* (2010), has already increased the efficiency of such studies (Baxter *et al.* 2017). Thirdly, no organism lives in isolation. During sample collection and preparation, a proportion of *M. jurtina* individuals were found carrying mites of the *Trombidium* genus, which both use them for transportation and feed on host body fluids (Conradt *et al.* 2002). These mites remained with their hosts after capture and are also present within the tissue archive. Investigations into whether certain aspects of *M. jurtina* ecology and habitat utilisation makes them more or less susceptible to parasitism or phoresy would be a fascinating area of study.

This project demonstrates that genetic monitoring of species is an achievable aim and one that could be applied to a range of species across multiple taxa. An obvious next step would be to apply the techniques used here, or alternative methods where costs allow, to other species of Lepidoptera. Whereas *M. jurtina* occurs in large, continuous populations (Thomas & Lewington 2010), the results would likely be very different for species that form more discrete populations such as adonis blue *Polyommatus bellargus* (Harper *et al.* 2003; O'Connor 2014). It should be noted that the methods used here would only be possible for species with large, stable populations. While the collection of whole samples and the creation of tissue archives provides further research opportunities and an element of future proofing, whole sample collection is not always possible or practical when species numbers are low (Hamm *et al.* 2010). In fact, the creation of a tissue archive, of similar size to ours, for the recently reintroduced chequered skipper butterfly in England, would be sufficient to eradicate the species approximately 60 times over. Fortunately non-lethal sampling techniques such as wing clipping or leg removal have been shown to yield sufficient DNA of high enough quality for population genetics studies, using a range of molecular markers (Keyghobadi *et al.* 2009; Vila *et al.* 2009a; Hamm *et al.* 2010). These techniques mimic wing wear and leg loss that is often seen in wild individuals and therefore are generally not detrimental to an individual's overall success (Koscinski *et al.* 2011). As such there is no reason that similar projects and monitoring schemes could not be set up for rarer lepidopteran species.

Whilst there have been many previous studies into the genetics of *M. jurtina* at the continental scale, the work presented in Chapter 4 represents the largest scale study of genetic diversity of *M. jurtina* using microsatellite markers. Whilst previous studies have relied upon

morphometrics and allozymes to answer specific questions regarding genetic lineages and post glacial expansions, this study focuses on contemporary gene flow. These approaches are complementary and the work presented here allows further understanding of this already well studied species.

A potential avenue of research from Chapter 4 would be to turn this single year study into a temporal dataset. There are currently samples from the same locations for the years 2016 and 2018 in the tissue store, however time and financial limitations did not facilitate their analysis. The yearly sampling of *M. jurtina* individuals in the UK can be done with relative ease, in a short period of time and at low cost. In contrast the sampling of individuals across Europe requires a great deal more organisation, time and financial backing as all samples need to be shipped frozen to the UK for analysis. This is due to the subjective nature of microsatellite allele scoring, as different labs and computer software can result in incomparable results (Bruford *et al.* 2015). This makes conducting such an endeavour yearly a challenge and raises an ethical question. Scientific research can result in large carbon footprints, particularly through the use of flights (Achten *et al.* 2013). The setup of a pan-European genetic monitoring scheme for *M. jurtina* would necessitate regular air shipments of samples, which would result in a large carbon footprint. Whether the conservation value of such a scheme would outweigh the environmental impact is something that should be considered, especially when ecologists and environmental scientists should be leading by example in terms of minimising carbon emissions (Burke 2010). Different analytical techniques such as SNPs may reduce this problem, as they would allow the analysis of samples in their country or origin and are more comparable between labs (Corlett 2017). However, whilst this would reduce the carbon emissions associated with such a scheme, it would greatly increase the level of commitment required from participating institutions.

Perhaps a more feasible area of further research would be to increase the number of sample locations and in doing so create a more evenly spatially distributed dataset. This would solve some of the limitations mentioned previously regarding uneven sampling. However, if further sampling were to occur it would be highly beneficial to focus on expanding the study area further into southern Europe. Currently the majority of samples occur within central and northern Europe, north of the Pyrenees and Alps. The only exception to this is the single Spanish site. Samples taken from southern Europe, especially from populations which undergo summer diapause could reveal greater levels of population structure across Europe. One question that could be addressed would be what effect do large mountain ranges such as the Pyrenees or the Alps have on *M. jurtina* gene flow. Further sampling of individuals from island populations would also be of great interest. As there appears to be restricted contemporary gene flow between *M. jurtina* in the UK and those in mainland Europe, it may be that this is also the case for populations on Mediterranean islands or those in the Canaries. The effects of the differing sizes of these islands, their distances from the mainland and the periods of time that they have been separated from their respective mainlands could yield interesting results.

6.3.3 Phenology

The final experimental chapter represents a significant, empirical contribution to the study of *M. jurtina.* The extended flight period of some *M. jurtina* populations on chalk sites in the UK is well known (Thomson 1971; Brakefield 1982b; Thomas & Lewington 2010), with empirical studies confirming this for specific sites (Goulson 1993b). Where this study differs is the use of long-term monitoring data to quantify these differences at the national level, as opposed to specific sites.

Understanding how environmental factors affect the phenology of populations may become increasingly important in conservation biology. The phenology of many species around the world is changing as a result of climate change (Miller-Rushing *et al.* 2010), resulting in species emergences no longer aligned with host plants or food sources (Visser *et al.* 2006; Hindle *et al.* 2015). By understanding what drives shifts in phenology we may be able to reduce the impact of phenology mismatch between trophic levels. For example, increasing levels of topographic diversity has already been suggested as a method of reducing mismatch between a UK butterfly species and its favoured nectar source (Hindle *et al.* 2015).

As highlighted previously, Chapter 5 would benefit from a more rigorous sampling scheme for both the genetic and experimental data. Whilst this would be a valid area of future research, a more interesting question would be to investigate why the flight period of *M. jurtina* is so broad compared to other similar species. If the analyses conducted here could be repeated for similar, univoltine butterflies such as gatekeeper (*Pyronia tithonus*), marbled white (*Melanargia galathea*) and ringlet (*Aphantopus hyperantus*) we would greatly increase our understanding of all of these species. If it were found that *M. jurtina* has much higher levels of genetic diversity than these other species, it could be that a greater ability to adapt to environmental change is the reason for this species success. In contrast if all these species were similarly genetically diverse then alternative explanations would need to be found.

6.4 Conclusions

The overall aim of this thesis was to demonstrate the use of monitoring data to answer ecological questions and to highlight the importance of genetic diversity. This aim has been achieved through three very different approaches, using pre-existing population monitoring data, novel genetic monitoring data and by combining the two with experimental data. Another aim was to highlight the importance of genetic diversity and to develop and maintain a monitoring scheme for *M. jurtina*. Whilst these aims have been achieved it is important that this is not seen as an end point. A great deal of work is still required in terms of monitoring genetic diversity. This thesis describes only a single, highly abundant species. The genetic diversity of countless other wild species remains unknown and whilst this thesis starts to address this issue it is just a small drop in a large ocean. It is my great hope that over time as molecular analysis costs continue to plummet that genetic diversity becomes part of the standard toolbox for monitoring biodiversity, alongside more traditional population monitoring and observations. Monitoring brings greater understanding and such an understanding will be required if we are to prevent species declines and losses in the Anthropocene.

Appendices

Appendix A Chapter 3 – Sample capture and preparation

From each site 20 whole butterfly samples were collected yearly from 2012 onwards. Sample collection from 2012 to 2016 was carried out by Tom Oliver (TO), Marc Botham and volunteers from the UK Centre for Ecology and Hydrology (UKCEH). From 2017 all samples were collected by TO, Matt Greenwell (MG) and volunteers from the University of Reading and Forest Research.

Adult butterflies were captured in nets and euthanized by pinching between the head and thorax. Samples were inspected for mites and when present the number of mites was recorded. Each sample was then placed in an individual envelope and stored in an electric cool box below 0°C until they could be transferred to permanent storage in a -20°C freezer at the University of Reading.

Leg dissections for all samples from 2012 and 2013 were carried out prior to the project starting by Melanie Gibbs (MGi) from UKCEH. All samples from 2014 onwards were dissected by MG. A leg was removed from each sample using a sterilised pair of forceps. Sterilisation was carried out by placing each pair of forceps into a bead sterilizer at 250°C for one minute after each use. Each leg was divided into three, separating the femur, tibia and tarsus. All three leg sections from each individual were placed into an individual tube of a PCR eight tube strip, which was sealed and returned to the -20°C freezer along with the original sample. Butterfly legs were transferred to the -20°C walk in freezer at UKCEH. The remainder of the lab work took place at UKCEH. This preparation occurred for all samples from the fifteen sites in 2017 and for all samples from three sites (Aston Upthorpe downs (AU), Bowdown forest (B) and Crabtree plantation (C)) for each year from 2014-2019.

Appendix B Chapter 3 – Novel primer development and evaluation

B.I Primer types used

5′ fluorescently labelled oligonucleotide primers from Thermofisher Scientific and Eurofins Genomics were used throughout much of the isolation and amplification processes (Fig. B.1). Oligonucleotide primers are usually sequences of 16 to 30 nucleotide bases, specific to target sequences of DNA. Primer pairs are designed in order to anneal to complementary sequences of template DNA and result in the amplification of segments of DNA when used in a PCR (McPhearson & Moller 2006).

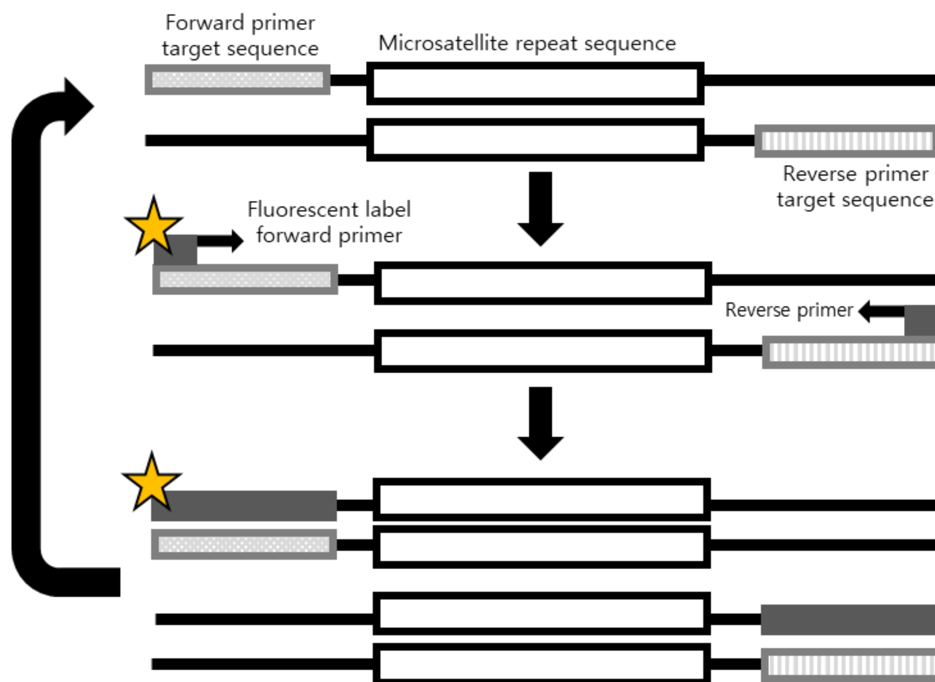


Figure B.1 Oligonucleotide primer microsatellite amplification method. Samples undergo repeated cycles of denaturation, annealing and extension with fluorescently labelled oligonucleotide forward primers and non-tagged reverse primers.

Both fluorescently labelled and non-labelled M13 primers were also used during the isolation process. M13-tailed primer analysis can be a highly efficient, cost effective method of generating labelled amplified products (Boutin-Ganache *et al.* 2001). Methods for M13 primer analysis are detailed in Culley *et al.* (2013) and summarised here (Fig. B.2). Primer pairs are developed for the microsatellite fragments being trailed. The primer pairs are then tested by amplifying the microsatellite fragment with three different primers in a single PCR reaction with multiple reaction cycles. Firstly, forward tailed primers tagged with an M13 DNA sequence at the 5′ end, and reverse primers are used to amplify the DNA fragment during the first amplification. This incorporates the M13 sequence into the first strand at the end of the targeted region. A fluorescently labelled tailing primer made up of the same M13 tail sequence and an added fluorescent label at the 5′ end, binds to the strand which includes the newly formed M13 complementary site. Amplification results in

the labelled tail annealing to the tail sequence, incorporating the fluorescent label to the amplified allele. A lower concentration of forward unlabelled primers is required to ensure they depleted in the first set of PCR cycles. Additionally, a lower annealing temperature is used for the second round of PCR cycles to increase the likelihood of annealing the larger labelled product (Culley *et al.* 2013)

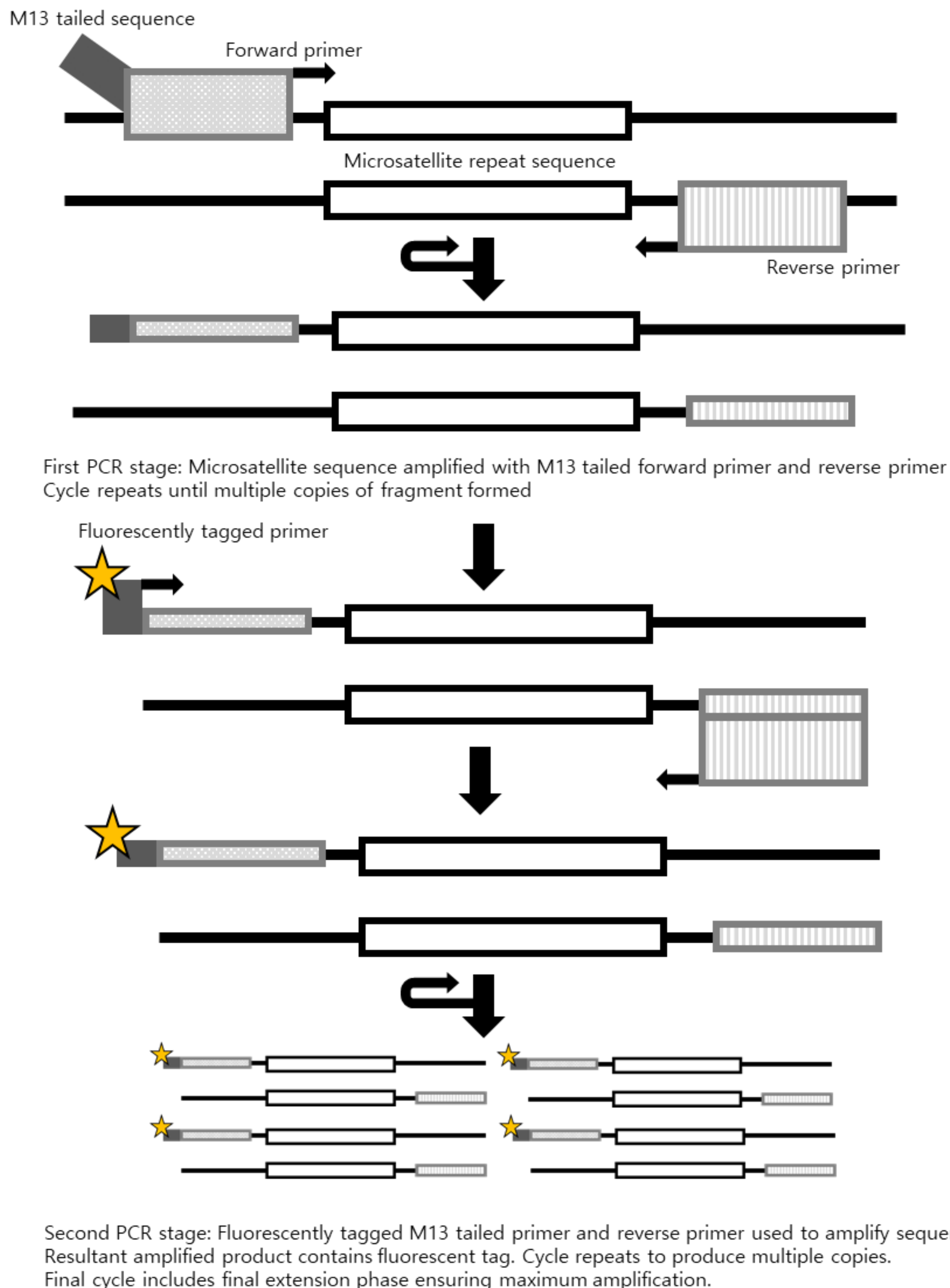


Figure B.2 M13 dual primer labelling method. Adapted from Culley *et al.* (2013): Figure 1.

B.II Individual PCR of non-tagged primer sequences

Nineteen non-labelled M13 tagged primer sequences (Table B.1), characterised prior to the study by John Day (JD) underwent individual PCR using *Taq* polymerase under the following conditions:

initial denaturation for 02:00 at 95°C, followed by forty cycles of 00:30 at 94°C, 0:30 at 56°C, 01:00 at 72°C and a final extension phase of 10:00 at 72°C. The loci Mj01, Mj04, Mj18 and Mj19 showed single banding patterns after gel electrophoresis and were selected for further testing (Fig. B.3). All other loci were discontinued from further analysis.

Table B.1 Nineteen unlabelled microsatellite loci forward and reverse primer sequences isolated using the program Primer3 (Untergasser et al., 2012). Loci taken on for further testing are in bold.

Locus	Forward and Reverse Primer Sequences	Repeat Unit
Mj01	**F: TTGCGTTGGCTCTGAAACG** **R: GGAAACAGCTATGACCATGATGCTAGGTCAACGGGAC**	**(CCGT)^{9}**
Mj02	F: CGAAGACTGTCGGAAACGC R: GGAAACAGCTATGACCA TGAACGATACTATAACCGAATCAC	(ACAG)^{7}
Mj03	F: AGGCTTGTGATTTATAGTGGGC R: GGAAACAGCTATGACCATTTCACGGCCCAACCGC	(GAT)^{7}
Mj04	**F: GGAAACAGCTATGACCA TGGAGCGATGCAGAAATGC** **R: CGTGAATTACGTGCTTACATGAG**	**(AGGTT)^{9}**
Mj05	F: GGAAACAGCTATGACCA TGCCCAATCCGTCCAGTAG R: TTCGCCAAATAAGCCGAGC	(AACCT)^{23}
Mj06	F: GGAAACAGCTATGACCA TTCGGCATTCAGTCCGCTC R: GGTTAGGTTCACCCGAAAGC	(TCA)^{7}
Mj07	F: AGAAACCTTTATAGGATCACTGTTG R: GGAAACAGCTATGACCATACGCCCTGGTCAATCCTC	(GTCT)^{11}
Mj08	F: TGCGGTGGCGCTCTTTAG R: GGAAACAGCTATGACCA TTTCCGATCGAAGCCTGGG	(GAT)^{7}
Mj09	F: GGAAACAGCTATGACC TCTCCACTTAGGAGTACGGC R: AGGCCCTGAAGGCTACTTG	(GAT)^{7}
Mj10	F: CGAGTCGCAGGAAGCCG R: GGAAACAGCTATGACCA TGTCGATGGCGGGATGAG	(GAT)^{8}
Mj11	F: GGTTATTTGAACGATACCATAACTG R: GGAAACAGCTATGACCA TGGAACTACGATTGTATGAAGTACG	(GTC)^{7}
Mj12	F: GGAAACAGCTATGACCA TTGTTTGCTTCGGCGTTCC R: CATGATTCTTGGTCAACGGG	(GTC)^{7}
Mj13	F: GGAAACAGCTATGACCA TCTGCAGGAGAACCAAGGTC R: AGATCCCAGTCAAGGGCAG	(TGA)^{8}
Mj15	F: AACGTGATGTTGCGACCTG R: GGAAACAGCTATGACCA TGTTTGAAGGAGGAGATCCG	(AGGTT)^{28}
Mj17	F: GGAAACAGCTATGACCAT GGCGCTCTATAGGGAAGG* R: AGATACTTGTGTAAGTCAAAGCC	(GAT)7
Mj18	**F: GGAAACAGCTATGACCAT** **R: CAATTCTCTACCTATTACGGTTCG**	**(ACGG)^{11}**
Mj19	**F: AGCAACTTACTCATTTCGAATAGC** **R: GGAAACAGCTATGACCAT CTATAAGAGACCTATGTCCTACCC**	**(ATC)^{13}**
Mj20	F: GTCATGTCTATGGATTCACGC R: GGAAACAGCTATGACCAT CGGCCCAGTAGTTTAGGC	(TATG)^{8}
Mj21	F: ACGAAATGGAAGAAACCTTTGC R: GGAAACAGCTATGACCAT TGAGCCAACATGAACTAGG	(ACAG)^{8}

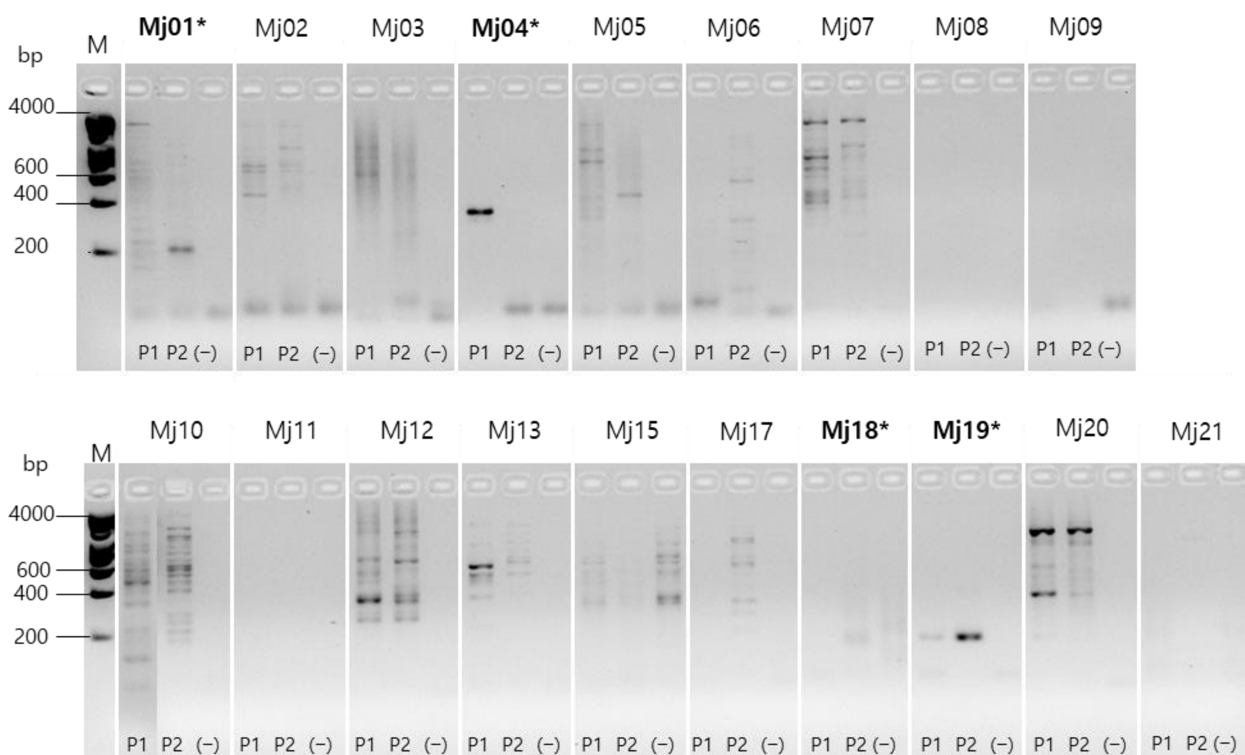


Figure B.3 Evaluation of PCR amplified products using 3μM primer concentrations of unlabelled primers. Gel electrophoresis carried out on a 3% agarose gel stained with ethidium bromide. Electrophoresis ran for 50 minutes at 95V and 400mA. DNA ladder (M) is shown to the left of each row. For each locus three PCR products are shown, two with DNA (P1 and P2) and a third with a negative control (−). Loci taken on for further analysis due to efficient amplification in correct regions are highlighted in bold and asterisked.

Four M13 tagged dyes (6-FAM, YAK, ATTO550, ATTO565) were individually added to the 5′ end of each of the four loci selected: Mj01, Mj04, Mj18 and Mj19. In total 16 sets of 10μL reactions were set up with 1μL DNA, 6.6μL dH_2O, 1μL 10x *Taq* polymerase buffer, 0.8μL dNTP's, 0.1μL M13-tailed forward primer (20μM), 0.1μL reverse primer (20μM), 0.05μL M13-labeled oligo (20μM) and 0.05μL *Taq* polymerase. The loci then underwent M13-tailed primer PCR under the following conditions: initial denaturation for 05:00 at 95°C, followed by ten cycles of 00:30 at 94°C, 01:00 at 57°C, 00:30 at 72°C and 27 cycles of 00:30 at 94°C, 01:00 at 55°C, 00:30 at 72°C followed by a final extension phase of 10:00 at 72°C.

All PCR reactions failed to produce any amplicon. As a result, the loci were then tested using standard PCR, resulting in effective amplification of correct size fragments. The following PCR conditions were used: initial denaturation for 02:00 at 95°C, followed by 40 cycles of 00:30 at 94°C, 01:30 at 56°C, 01:00 at 72°C and a final extension of 10:00 at 72°C.

After PCR, all samples underwent gel electrophoresis for 50 minutes on a 3% agarose gel at 95V and 400mA (unless otherwise specified all other gel electrophoresis were carried out under the same conditions). All primers showed strong banding, indicating successful amplification. As a result, all sixteen M13 variations of the four loci were taken on for further analysis.

B.III Individual PCR of tagged primer sequences

Eight pairs of primers (Table B.2) characterised prior to this study by John Day were evaluated. Each forward primer was labelled with one of four fluorescent primer dyes ATTO550, ATTO565, 6-FAM and YAKIMA Yellow (YAK), and a series of individual PCRs with *Taq* polymerase were carried out under the following conditions: initial denaturation for 02:00 at 94°C followed by 40 cycles of 00:30 at 94°C, 00:30 at 56°C - 60°C, 01:00 at 72°C and a final extension for 05:00 at 72°C. The 6-FAM tagged locus Mj4870 was used as a positive control. Gel electrophoresis showed that many loci did not amplify the expected size product. However, electrophoresis of the loci ATTO550 Mj43, ATTO565 Mj44 and YAK Mj51 produced amplicon of predicted size (Fig. B.3). As a result, these loci were taken on for further evaluation. The remaining five loci were discarded from the study.

Table B.2 Eight *M. jurtina* microsatellite loci forward (F) and reverse (R) primer sequences isolated using the programme Primer3 (Untergasser et al., 2012). Loci taken on for further evaluation are in bold.

Loci	Forward and reverse primer sequences	Repeat unit	Forward primer dye
Mj13	F: TCTGCAGGAGAACCAAGGTC R: AGATCCCAGTCAAGGGCAG	$(TGA)^{8}$	6-FAM
Mj14	F: TACCTGGTTCGTGAACGGC R: TGGCGCTCTTTAGGGAAGG	$(CAT)^{11}$	ATTO550
Mj34	F: TCGGCCCAGTAGTTTAGGC R: AACCGACTGAGGCAAAGC	$(CATA)^{8}$	ATTO565
Mj40	F: TATGGTCTAAACCCTTCGCATTC R: GTTCTCTGCCGCGATGTAAG	$(GAT)^{11}$	YAK
Mj42	F: CCCAAAGTTGAAAGAGGCGG R: GCTCGCGTTCATAGTCCAC	$(GAT)^{15}$	6-FAM
Mj43	**F: TGCTCAAATCGCTGCTAAGAC** **R: GGTCCCGTAGCTGACTCC**	$\mathbf{(GAT)^{9}}$	**ATTO550**
Mj44	**F: TAAGCCTTTGGATAACCTTGGAAAC** **R: GGCGCTCTTTAGGGAAGGC**	$\mathbf{(ATC)^{10}}$	**ATTO565**
Mj51	**F: CACTCATGGTGGTGCCCTC** **R: CGATTCGCAGACGTTGCTC**	$\mathbf{(ACTCT)^{10}}$	**YAK**

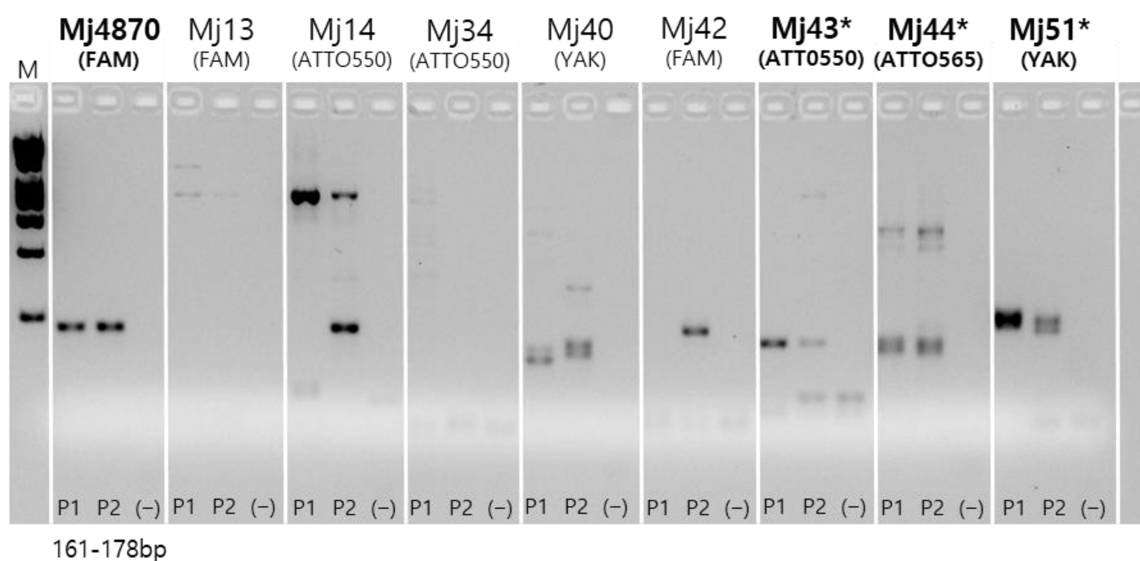


Figure B.4 Evaluation of amplification of PCR products using 3µM primer concentrations of labelled loci. Gel electrophoresis carried out on a 3% agarose gel stained with ethidium bromide. Electrophoresis ran for 50 minutes at 95V and 400mA. DNA ladder (M) is shown to the left. For each locus three PCR products are shown, two with DNA (P1 and P2) and a negative control (−). Loci taken on for further analysis due to efficient amplification in correct regions are highlighted in bold and asterisked. Locus Mj4870 is shown in bold as an example of a previous working locus for comparison, with expected fragment size shown.

B.IV Fragment analysis

To determine whether the sixteen M13 tagged loci and the three loci ATTO550 Mj43, ATTO565 Mj44 and YAK Mj51 could be used to genotype individuals from UK populations, representative amplicons for each locus underwent ABI fragment analysis using the following protocol.

For each ABI sequence run, amplicon was diluted by 100-fold, as undiluted products resulted in peak saturation on the Sequencer chromatogram. For each individual, 1µL of diluted PCR product was used. A mixture of 9µL Hi-Di™ formamide and 1µL of ABI LIZ500 size standard was vortexed, incubated at 95°C for 3 minutes, chilled on ice for 2 minutes and then added to each 1µL diluted PCR sample. All samples were then mixed and run on an Applied Biosystems 3730 DNA Analyser.

B.V Allele scoring

GeneMarker® version 1.5 (SoftGenetics) was used to score alleles and genotype individuals for each locus, using the microsatellite calibration settings. Scoring was carried out by eye and alleles which were unable to be easily scored due to stuttered, multiple, or saturated peaks were removed from the analysis.

All PCR products from the sixteen M13 tagged loci produced ambiguous electropherograms that could not be accurately scored. As a result, all loci were dropped from further testing. The PCR products from the loci ATTO565 Mj44 and YAK Mj51 also produced unclear electropherograms, showing multiple peaks for each locus, rather than distinct peaks that could be scored as

heterozygous or homozygous. This indicated that the loci may have been amplifying multiple areas of the genome and that the primer sequences were not unique to the microsatellite flanking regions, a known factor in Lepidoptera microsatellite development (Meglécz *et al.* 2004). The electropherograms for the locus ATTO550 Mj43 showed distinct peaks which could be scored, allowing the genotyping of individuals as either heterozygous or homozygous for the locus. As a result, the locus ATTO550 Mj43 was retained for further development and application. The other two loci were dropped from the analysis.

B.VI Primer redesign

During further testing of the locus ATTO550 Mj43 it was observed that no product was amplified during PCR for the DNA sample ARN-12-06, under a range of primer concentrations (1µM-3µM) and annealing temperatures (50°C - 56°C). Previous results with alternate loci had proved that the DNA sample could be successfully amplified, therefore the issue was with the locus and not the DNA. The DNA sample came from a male, therefore the allele was not sex-linked (Lepidoptera have a ZZ/ZW sex chromosome, with males having ZZ (Richard *et al.* 2015)). The lack of amplification may have been a result of a null allele, a mutation in the flanking region preventing primer binding and therefore product amplification (Chapuis & Estoup 2007).

Primers for the locus ATTO550 Mj43 were redesigned by JD by selecting different sections of the microsatellite flanking regions using the programme Primer3 (Untergasser *et al.* 2012). Two new reverse primer sites and one new forward site were selected. Using the new primers in combination with the originals, five sets of PCR reactions were set up. Each set containing one of the two forward primers and one of the three reverse primers, creating all permutations of primer pairs. All loci were amplified in duplicate, at different annealing temperatures (52°C and 56°C) using three DNA samples, two of which had amplified previously (ARN-12-05, ARN-12-07) and ARN-12-06. No combination of primers resulted in successful amplification of the ARN-12-06 DNA, despite all positive control samples amplifying successfully (Fig. B.5). This suggested the problem was not a point mutation in the original flanking, to have a large succession of point mutations affecting all newly designed primers was unlikely. A further possibility is that the result may be due to a large indel, an insertion or deletion of nucleotide bases (Albers *et al.* 2011) at the primer sites. This would result in primers not being able to bind, even after being redesigned. The locus was dropped from further analysis.

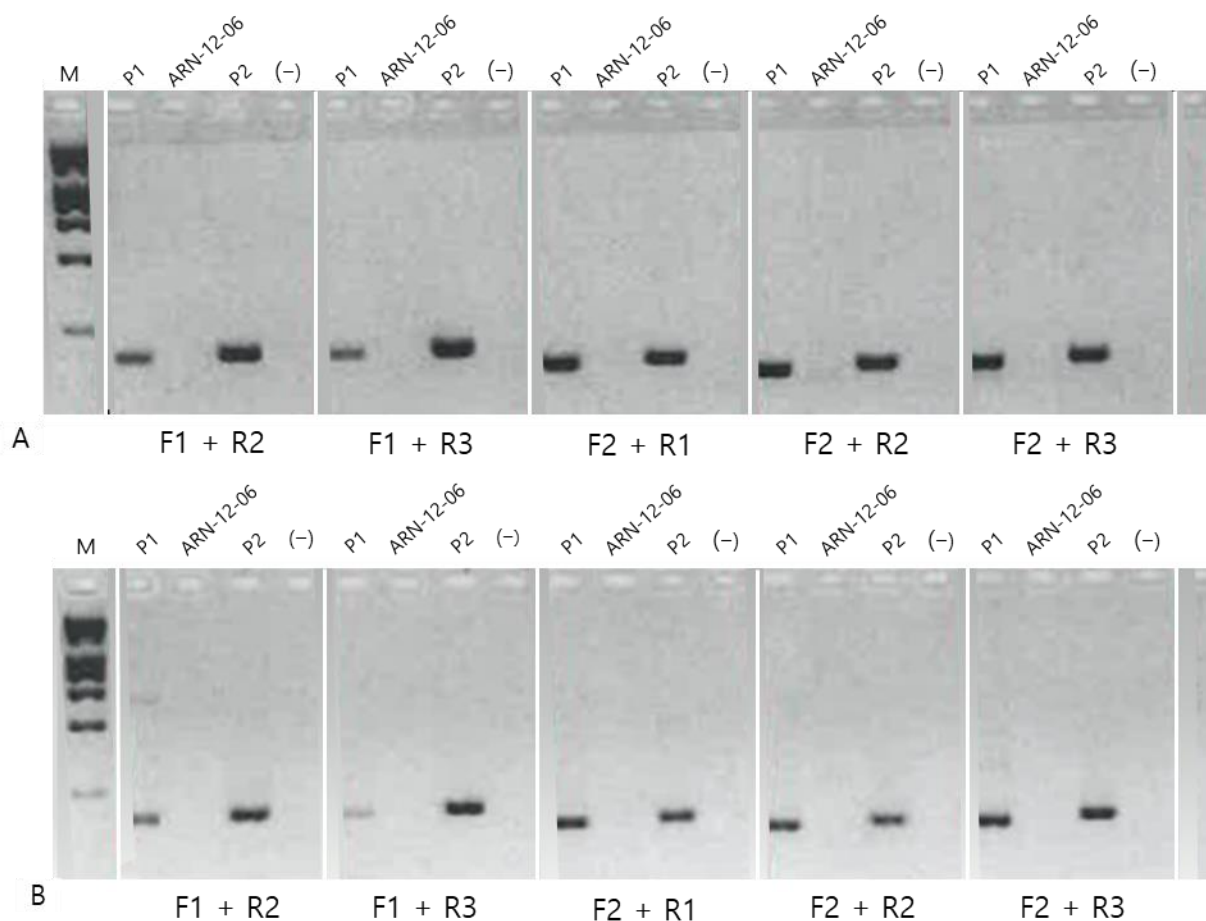


Figure B.5 Evaluation of amplification of PCR products using 3μM concentrations of ATTO550 labelled locus Mj43. Gel electrophoresis carried out on a 3% agarose gel stained with ethidium bromide. Electrophoresis ran for 50 minutes at 95V and 400mA. DNA ladder (M) is shown to the left. Each set of four PCR samples shows a different combination of forward primers (F1 and F2) and reverse primers (R1, R2 and R3). Each set of four samples contains two successfully amplified DNA samples (P1 and P2), the attempted amplification of DNA sample ARN-12-06, and a negative control (−). Reactions shown on the top row (A) were carried out with an annealing temperature of 56°C. Reactions shown on the bottom row (B) were carried out with an annealing temperature of 52°C.

Appendix C Chapter 3 – Pre-existing primer testing

Fifteen polymorphic microsatellite loci were isolated for *M. jurtina* (Table C.1) by Richard *et al.* (2015). Using 96 samples from six populations across five French regions, the authors found a mean overall expected heterozygosity (He) of 0.74 across all fifteen loci. Five loci departed significantly from Hardy-Weinberg equilibrium and one exhibited high levels of null-alleles and was discarded from further investigations (Mj0272). A further locus was found to be sex-linked (Mj2410). Seven loci showed moderate levels (~15%) of null alleles (Richard *et al.* 2015).

In this study, 12 of the 15 microsatellites isolated by Richard et al. (2015) were used in an analysis to investigate the genetic diversity of populations of *M. jurtina* in the UK. The loci Mj0272 and Mj2410 were not selected due to a high proportion of null–alleles and sex-linkage respectively as described in Richard et al. (2015). The locus Mj5647 was also not selected, as it would have been a surplus PET dyed locus and could not have been included in multiplex reactions, due to the predicted size of fragments overlapping with other PET dyed loci during ABI fragment analysis. The remaining loci from Richard et al. (2015) were used to save significant time and costs compared to isolating novel loci for this study.

Table C.1 Fifteen microsatellite loci isolated from *M. jurtina.* Reproduced from (Richard *et al.* 2015). The six viable loci used throughout the analysis are highlighted in bold.
* sex-linked locus, # high null allele frequency.

Locus	Primer sequence (5'-3')	Repeat motif	Multiplex group in Richard *et al.* (2015)
Mj0008	F: PET-CGTGTCGCCTAAACCACATC R: TGGCAACCCTAAACCCTACG	(ACAT)^{7}	1
Mj3956	F: PET-CAACATCGGGAGTCGAAACG R: CTCAGCCAGGATACCCACTC	(GATA)^{7}	2
Mj5331	**F: PET-TTAGACCGTGATCCCACTGC** **R: ATTTCGATAGGCAACGAGGC**	**(TATC)^{10}**	**3**
Mj5287	F: 6FAM-GCTAGCTCGTGGGTACTCTG R: CTCCAAGCAATAAGACCGCC	(GATA)^{11}	1
Mj7232	**F: 6FAM-AAGTTACAAGAGCGTTGGCG** **R: GCGGGAACTCTTGGGTTTTC**	**(CTGT)^{7}**	**2**
Mj4870	**F: 6FAM-ATGATCCATAGCTGCGTTGC** **R: CTCCTTAGCGCTTACACGTC**	**(ATGT)^{7}**	**3**
Mj7132	**F: NED-ATCTGCGGATTTGCAGTTGG** **R: CACTATTGAGCACGTGTGTCC**	**(TATG)^{13}**	**1**
Mj5522	**F: NED-TGATCTTTGCCAGCAGGAAC** **R: AGTGTAAGCTGGCCCTAAAC**	**(GATA)^{8}**	**2**
Mj3637	F: NED-CTTCCGCAAAATAACGTCTGC R: AGATACTCCATTGACCCGGC	(TCTA)^{7}	3
Mj5647	F: PET-GCGTTCTGATTACCACCCTG R: GCGACAGTCCCCTAAGATCG	(TATG)^{13}	1
	F: PET-ATTCCACAAACGAGCCAACG		

Locus	Primer sequence (5'-3')	Repeat motif	Multiplex group in Richard *et al.* (2015)
Mj0247	**R: ACTCCGATGGTAAGAGGTGC**	**$(GATG)^8$**	**2**
Mj2410*	F: PET-TAATTAGAGTTTGCGCGGGG R: CGCACACCGCAGTATAAGTG	$(TGTA)^7$	3
Mj5563	F: VIC-CGGTTTTGCCGATAGCGTAG R: CGCAAGGCAATAGACCACTC	$(ATCT)^7$	1
Mj0272#	F: VIC-GTTGCATTGGCACACTCCTC R: CAGCTGCACACTACGACAAG	$(AGAT)^7$	2
Mj0283	F: VIC-CCCTTAGAATAAGAACTCGGCTC R: TGTTCGCACATGCTTAGTCC	$(AGAT)^9$	3

C.I Individual primer PCR testing

To determine whether the selected loci could be used to amplify correct product size amplicon from *M. jurtina* individuals in UK populations, all 12 loci selected from Richard et al. (2015) were tested individually under a series of PCRs, with *Taq* polymerase, under the following conditions: initial denaturation for 02:00 at 94°C followed by 35 cycles of 00:30 at 94°C, 00:30 at 56-60°C, 00:30 at 72°C and a final extension phase for 05:00 at 72°C. A range of primer concentrations between 0.2μM and 20μM, and annealing temperatures between 50˚C and 60˚C were used. All PCR products were visualised by gel electrophoresis, confirming all primers effectively amplified the correct size product from eight samples across the study area (Fig. C.1).

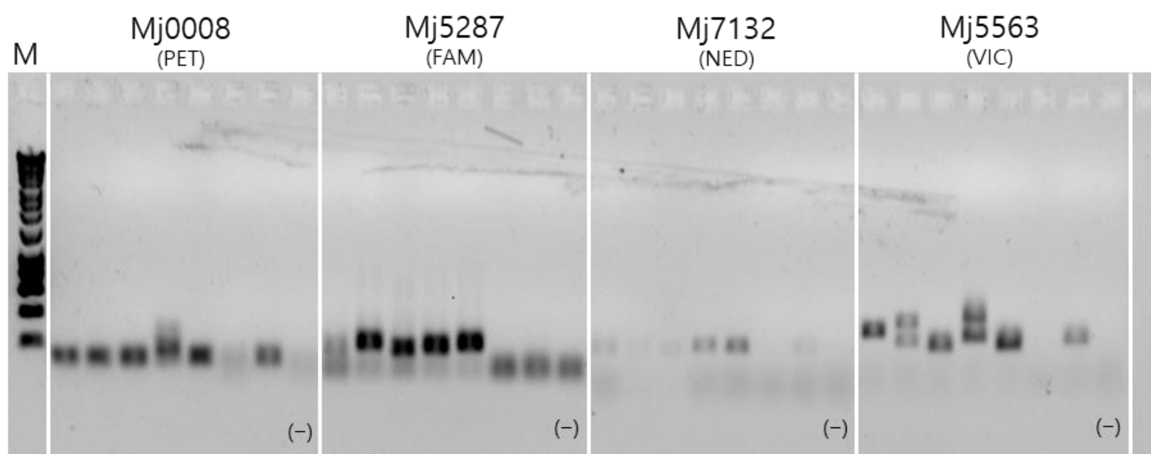


Figure C.1 Evaluation of amplification of PCR products using 5μM concentrations of labelled loci. Gel electrophoresis carried out on a 2% agarose gel stained with ethidium bromide. Electrophoresis ran for 50 minutes at 95V and 400mA. DNA ladder (M) is shown to the left. Each set of eight samples contains seven amplified DNA samples and a negative control (−) as the eighth sample.

C.II Multiplex PCR optimisation

All multiplex PCRs were carried out using Qiagen Multiplex PCR kits with the following reagent mixes: 1μL template DNA, 6.25μL QIAGEN multiplex PCR master mix (3mM $MgCl_2$), 0.625μL tagged

forward primer, 0.625μL reverse primer, 1.25μL QIAGEN Q solution, 2.25μL RNase-free water. Initial multiplex testing began using the four loci used in multiplex one in Richard *et al.* (2015) (Table C.1), in a number of quadruplex PCRs with primer concentrations from 0.2μM to 3μM, under a number of different reaction conditions all starting with an initial denaturation for 15:00 at 95°C, followed by 35-40 cycles of 00:30 at 94°C, 00:30 at 56°C - 60°C, 00:30 - 01:30 at 72°C and a final extension phase of 05:00 at 72°C.

Successful fragment amplification occurred using higher primer concentrations, longer extension times and an increased number of reaction cycles. From this point onwards the primer concentrations used by Richard *et al.* (2015) were abandoned and all PCRs were carried out with higher primer concentrations (between 1μM-5μM) and the following conditions: initial denaturation for 15:00 at 95°C followed by 40 cycles 00:30 at 94°C, 01:30 at 56°C, 01:00 at 72°C and a final extension phase for 10:00 at 72°C.

The 12 loci were grouped into three sets of four, following the multiplex groupings in Table C.1. Each set of four loci underwent a quadruplex PCR. To determine whether all loci were being amplified equally in the quadruplex PCRs, each set of four loci also underwent PCRs as pairs of loci in duplex reactions, and as individual loci in four individual PCRs. Comparisons of the resultant gel images showed that some loci were lost in the quadruplex reactions as a result of preferential amplification of other loci. An example of four loci amplified individually, in duplexes and as a quadruplex PCR can be seen in Fig. C.2.

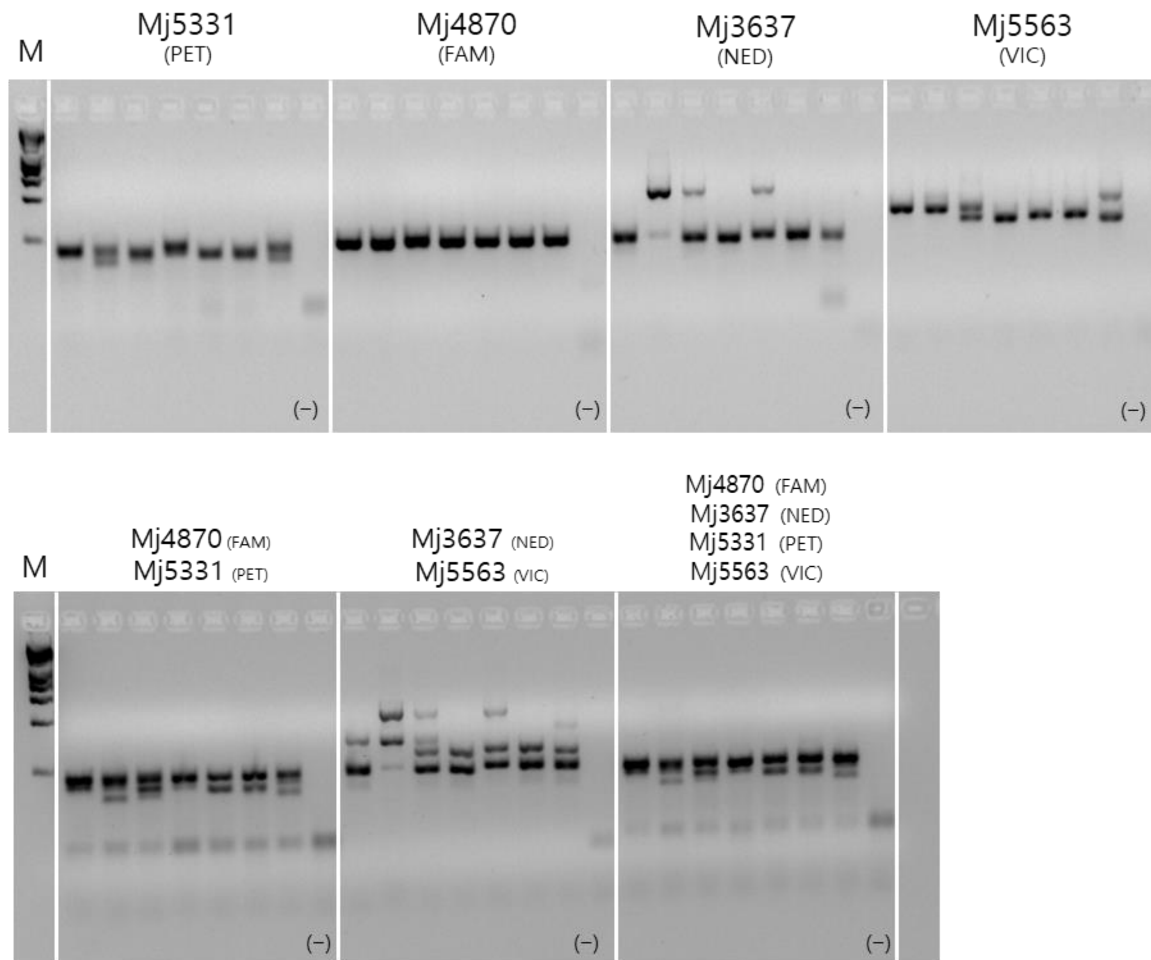


Figure C.2 Evaluation of amplification of PCR products using different combinations of 3μM concentrations of labelled loci. Gel electrophoresis carried out on a 3% agarose gel stained with ethidium bromide. Electrophoresis ran for 50 minutes at 95V and 400mA. DNA ladder (M) is shown to the left. Each set of eight samples contains seven amplified DNA samples and a negative control (−) as the eighth sample.

To reduce the levels of preferential amplification, the primer concentrations of loci within each quadruplex PCR were altered. Loci that did not amplify in the quadruplex PCRs (e.g. Mj5563, Mj3637) were increased to 4.5μM primer concentration, or loci that had been preferentially amplified (e.g. Mj4870, Mj5331) were reduced to 1.5μM, 1μM or 0.5μM primer concentrations. The alterations in primer concentrations reduced the preferential amplification of loci over others and resulted in a more even amplification of loci during the quadruplex PCR reactions (e.g. as in Fig. C.3).

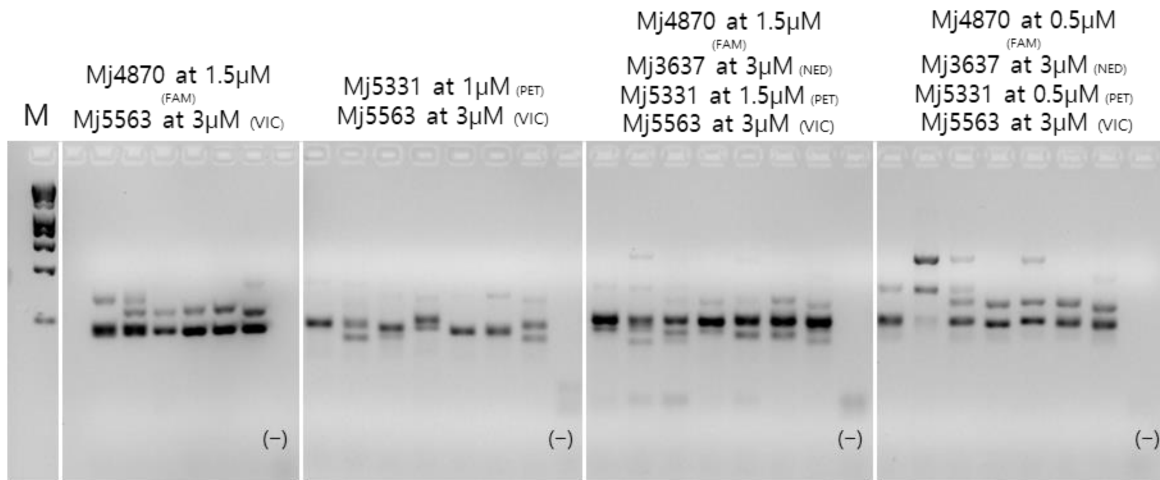


Figure C.3 Evaluation of amplification of PCR products using different concentrations of labelled loci in a series of duplex and quadraplex reactions. Gel electrophoresis carried out on a 3% agarose gel stained with ethidium bromide. Electrophoresis ran for 50 minutes at 95V and 400mA. DNA ladder (M) is shown to the left. Each set of eight samples contains seven amplified DNA samples and a negative control (−) as the eighth sample. Images show return of bands seen in individual reactions but lost in multiplex reactions in Fig. C.2.

C.III Fragment analysis and genotyping

To determine whether the loci could be used to genotype *M. jurtina* individuals from a range of UK populations, samples underwent ABI fragment analysis, were genotyped and scored following the same protocols as described above (B.IV & B.V). For each ABI 3730 run, final PCR products were diluted by 100x as undiluted products resulted in saturation of peaks on the sequencer chromatogram. An additional reason for diluting samples was that in multiplex reactions high concentrations resulted in pull-up or bleed through, where the software was unable to distinguish between dye types and produced artefact peaks.

Dilutions of the final quadruplex PCR product for each multiplex set were analysed. Additionally, to determine whether the combinations of loci within the multiplex reactions were compatible with each other and being evenly amplified, PCR products from the individually amplified loci were combined to form mixtures containing the four loci in each of the multiplex sets. Pairs of duplexes were also mixed to form four loci mixes. For each set of four loci, three reaction mixes underwent ABI sequencing. The loci 6-FAM Mj7232, 6-FAM Mj4870, NED Mj7132, NED Mj5522, PET Mj0008, PET Mj0247 and PET Mj5331 all showed a high proportion of success, with over 75% of electropherograms showing clear peaks that could be scored as either heterozygous or homozygous. These loci included the six loci used in a study by Villemey *et al.* (2016) and the additional Mj5522. The other eight loci were removed from further analysis due to their low success rates.

The remaining loci were split into two groups of three, with the loci 6-FAM Mj7232, NED Mj5522, and PET Mj0247 making up multiplex I, and 6-FAM Mj4870, NED Mj7132 and PET Mj5331 making up multiplex II. The locus PET Mj0008 was removed from the analysis as further testing

determined that all individuals were homozygous for the locus at 98bp. This indicated that the locus likely had a high frequency of null alleles resulting in homozygote deficit.

Both sets of primers underwent triplex PCRs at 3μM concentrations, followed by 100-fold dilutions and fragment analysis. Electropherograms with distinct peaks that could be scored were produced for multiplex I for more than 75% of DNA samples, across all three loci. Electropherograms from multiplex II showed a greater range of peak heights, with 6-FAM Mj4870 having extremely high levels of fluorescence and NED Mj7132 having very low levels. To resolve this the PCR and fragment analysis were repeated, with altered concentrations of primers used in the multiplex PCR (6-FAM Mj4870 decreased to 1.5μM, NED Mj7132 increased to 5μM and PET Mj5331 increased to 4.5μM). The resultant electropherograms showed more even peak heights across all loci for all samples, allowing more reliable scoring. All 956 individuals across the seven years of sampling were processed with the six remaining microsatellite markers.

Appendix D Chapter 4 – Effects of UK site removal on genetic analysis

D.I Microsatellites, HWE and GE

Table D.1 Composite linkage disequilibrium test outputs for all locus pair combinations across mainland European sites only and with 3 UK sites included.

Sites included	Locus pair		Chi^2	df	p-Value
Mainland Europe	Mj7232	Mj5522	35.307	46	0.874
Mainland Europe	Mj7232	Mj0247	9.218	44	1.000
Mainland Europe	Mj5522	Mj0247	11.421	44	1.000
Mainland Europe	Mj7232	Mj4870	26.240	46	0.992
Mainland Europe	Mj5522	Mj4870	23.815	46	0.997
Mainland Europe	Mj0247	Mj4870	23.062	44	0.996
Mainland Europe	Mj7232	Mj7132	26.769	46	0.990
Mainland Europe	Mj5522	Mj7132	30.894	46	0.957
Mainland Europe	Mj0247	Mj7132	27.414	44	0.976
Mainland Europe	Mj4870	Mj7132	43.216	48	0.669
Mainland Europe	Mj7232	Mj5331	23.129	44	0.996
Mainland Europe	Mj5522	Mj5331	21.471	44	0.998
Mainland Europe	Mj0247	Mj5331	6.969	42	1.000
Mainland Europe	Mj4870	Mj5331	22.350	46	0.999
Mainland Europe	Mj7132	Mj5331	34.732	46	0.888
Mainland Europe + 3 UK	Mj7232	Mj5522	38.163	52	0.924
Mainland Europe + 3 UK	Mj7232	Mj0247	10.044	48	1.000
Mainland Europe + 3 UK	Mj5522	Mj0247	12.504	48	1.000
Mainland Europe + 3 UK	Mj7232	Mj4870	32.479	52	0.985
Mainland Europe + 3 UK	Mj5522	Mj4870	27.363	52	0.998
Mainland Europe + 3 UK	Mj0247	Mj4870	25.079	48	0.997
Mainland Europe + 3 UK	Mj7232	Mj7132	28.816	52	0.996
Mainland Europe + 3 UK	Mj5522	Mj7132	32.767	52	0.983
Mainland Europe + 3 UK	Mj0247	Mj7132	33.181	48	0.949
Mainland Europe + 3 UK	Mj4870	Mj7132	45.951	54	0.774
Mainland Europe + 3 UK	Mj7232	Mj5331	31.035	50	0.984
Mainland Europe + 3 UK	Mj5522	Mj5331	21.217	50	1.000
Mainland Europe + 3 UK	Mj0247	Mj5331	8.465	46	1.000
Mainland Europe + 3 UK	Mj4870	Mj5331	24.913	52	0.999
Mainland Europe + 3 UK	Mj7132	Mj5331	39.770	52	0.893

Table D.2 Locus by populations estimated null allele frequencies. Values in bold exceed 0.2 frequency of null alleles

Sites included	Locus	UK-AU	UK-B	UK-C	SP	FR-A	FR-F	FR-K	FR-M	B-A	B-B	B-C	ND-A	ND-B	L	SZ	DK-A	DK-B	G-A	G-B	G-C	G-D	G-E	SW	E-A	E-B	E-C	E-D
Mainland Europe	Mj7232	-	-	-	0.041	0.181	**0.248**	0.029	0.058	0.140	**0.242**	0.176	0.153	**0.340**	**0.305**	0.000	0.188	0.146	0.082	0.106	0.184	0.128	**0.200**	0.097	**0.261**	0.185	**0.366**	**0.426**
Mainland Europe	Mj5522	-	-	-	0.129	0.000	0.040	0.017	0.000	0.054	0.060	0.013	0.056	0.053	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.038	0.000	0.000	0.000	0.000	0.000	0.034
Mainland Europe	Mj0247	-	-	-	0.070	0.055	0.198	0.081	0.158	0.034	0.072	0.110	0.078	0.000	0.070	0.000	0.176	0.079	0.010	0.073	0.038	0.098	0.192	0.029	0.116	0.051	0.048	0.038
Mainland Europe	Mj4870	-	-	-	0.167	**0.377**	**0.296**	0.173	0.076	**0.227**	0.000	0.354	**0.315**	**0.369**	**0.284**	**0.262**	**0.287**	**0.206**	**0.312**	**0.248**	**0.371**	**0.297**	**0.372**	0.180	**0.384**	**0.221**	**0.321**	**0.358**
Mainland Europe	Mj7132	-	-	-	0.088	0.000	0.000	0.116	0.017	0.006	0.000	0.019	0.000	0.000	0.000	0.068	0.110	0.000	0.127	0.000	0.000	0.000	0.087	0.162	0.000	0.000	0.000	0.074
Mainland Europe	Mj5331	-	-	-	0.000	0.108	0.129	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.024	0.015	0.000	0.000	0.000	0.058	0.015	0.000	0.000	0.078	0.000	0.043	0.000	0.061
Mainland Europe + 3 UK	Mj7232	0.000	0.137	0.079	0.041	0.181	**0.248**	0.029	0.058	0.140	**0.242**	0.176	0.153	**0.340**	**0.305**	0.000	0.188	0.146	0.082	0.106	0.184	0.128	**0.200**	0.097	**0.261**	0.185	**0.366**	**0.426**
Mainland Europe + 3 UK	Mj5522	0.048	0.000	0.000	0.129	0.000	0.040	0.017	0.000	0.054	0.060	0.013	0.056	0.053	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.038	0.000	0.000	0.000	0.000	0.000	0.034
Mainland Europe + 3 UK	Mj0247	0.000	0.085	0.018	0.070	0.055	0.198	0.081	0.158	0.034	0.072	0.110	0.078	0.000	0.070	0.000	0.176	0.079	0.010	0.073	0.038	0.098	0.192	0.029	0.116	0.051	0.048	0.038
Mainland Europe + 3 UK	Mj4870	0.000	0.000	0.000	0.167	**0.377**	**0.296**	0.173	0.076	**0.227**	0.000	**0.354**	**0.315**	**0.369**	**0.284**	**0.262**	**0.287**	**0.206**	**0.312**	**0.248**	**0.371**	**0.297**	**0.372**	0.180	**0.384**	**0.221**	**0.321**	**0.358**
Mainland Europe + 3 UK	Mj7132	0.000	0.045	0.000	0.088	0.000	0.000	0.116	0.017	0.006	0.000	0.019	0.000	0.000	0.000	0.068	0.110	0.000	0.127	0.000	0.000	0.000	0.087	0.162	0.000	0.000	0.000	0.074
Mainland Europe + 3 UK	Mj5331	0.000	0.032	0.005	0.000	0.108	0.129	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.024	0.015	0.000	0.000	0.000	0.058	0.015	0.000	0.000	0.078	0.000	0.043	0.000	0.061

Table D.3 Population-wide expected and observed heterozygosity, and percent difference ((E - O)/E*100), F_{IT}, F_{IS} F_{ST} and R_{ST} at each locus across sites. Bartlett's K-squared for Mainland Europe = 4.824, df = 1, p-value = 0.0281. Bartlett's K-squared for mainland Europe and three UK sites = 4.055, df = 1, p-value = 0.044.

Sites Included	Locus	Number of samples	Number of alleles	H_e	H_o	H_e vs H_o % difference	F_{IT} (p-value)		F_{ST} (p-value)		F_{IS} (p-value)		R_{ST}
Mainland Europe	Mj7232	434	15	0.819	0.724	11.631	0.119	0.000	0.035	0.000	0.086	0.000	0.046
Mainland Europe	Mj5522	427	13	0.828	0.785	5.218	0.054	0.006	0.013	0.666	0.041	0.033	0.021
Mainland Europe	Mj0247	431	29	0.933	0.738	20.912	0.211	0.000	0.024	0.013	0.191	0.000	0.052
Mainland Europe	Mj4870	429	8	0.719	0.207	71.157	0.712	0.000	0.028	0.291	0.704	0.000	0.037
Mainland Europe	Mj7132	427	11	0.782	0.726	7.199	0.074	0.000	0.013	0.653	0.062	0.013	0.005
Mainland Europe	Mj5331	430	21	0.883	0.835	5.439	0.056	0.000	0.015	0.480	0.042	0.014	0.006
Mainland Europe + 3 UK	Mj7232	468	16	0.829	0.722	12.870	0.130	0.000	0.046	0.000	0.088	0.000	0.063
Mainland Europe + 3 UK	Mj5522	471	13	0.830	0.785	5.453	0.057	0.000	0.013	0.674	0.044	0.027	0.021
Mainland Europe + 3 UK	Mj0247	455	29	0.935	0.747	20.113	0.201	0.000	0.023	0.007	0.182	0.000	0.051
Mainland Europe + 3 UK	Mj4870	476	8	0.701	0.222	68.255	0.689	0.000	0.063	0.000	0.668	0.000	0.092
Mainland Europe + 3 UK	Mj7132	473	11	0.779	0.732	6.003	0.063	0.001	0.015	0.496	0.048	0.022	0.008
Mainland Europe + 3 UK	Mj5331	463	21	0.887	0.842	5.026	0.052	0.000	0.017	0.252	0.036	0.022	0.002

D.II Genetic diversity

Table D.4 Sample sizes, genetic diversity, allelic richness, number of private alleles and effective population sizes for *M. jurtina* populations across Europe. All values are estimated on a per population basis.

H_{exp} = expected heterozygosity, A_r = allelic richness, A_p = Private alleles, Ne(1) = effective population size estimated using the heterozygote excess method, Ne(2) = effective population size estimated using linkage disequilibrium method.

Site	Sample Size	Mean H_{exp}	(s.d.)	A_r	A_p	$N_e(1)$	$N_e(2)$
Mainland Europe	434	0.804	(0.097)	-	15	-	-
Mainland Europe + 3 UK Sites	486	0.828	(0.082)	-	12	-	-
UK-AU	14	0.713	(0.269)	5.574	0	∞	∞
UK-B	17	0.738	(0.246)	5.693	0	∞	∞
UK-C	20	0.783	(0.199)	5.990	1	∞	∞
SP	15	0.827	(0.099)	6.035	0	∞	∞
FR-A	20	0.819	(0.068)	5.887	0	∞	∞
FR-B	20	0.818	(0.063)	5.607	0	∞	∞
FR-C	18	0.826	(0.116)	6.186	1	∞	∞
FR-D	20	0.818	(0.112)	6.044	2	∞	∞
B-A	20	0.798	(0.124)	5.801	0	∞	∞
B-B	6	0.712	(0.273)	5.000	0	∞	∞
B-C	18	0.831	(0.066)	5.879	0	∞	∞
ND-A	15	0.811	(0.084)	5.875	0	∞	∞
ND-B	16	0.804	(0.078)	5.606	1	∞	∞
L	18	0.793	(0.110)	5.739	1	∞	∞
SZ	18	0.770	(0.119)	5.406	0	∞	∞
DK-A	19	0.810	(0.077)	5.616	2	∞	∞
DK-B	20	0.806	(0.099)	5.852	1	∞	∞
G-A	23	0.802	(0.090)	5.654	0	∞	∞
G-B	11	0.768	(0.167)	5.335	1	∞	∞
G-C	20	0.817	(0.073)	5.645	0	∞	∞
G-D	22	0.828	(0.079)	6.091	1	∞	∞
G-E	20	0.814	(0.096)	5.895	0	∞	∞
S-W	15	0.815	(0.073)	5.467	0	∞	∞
E-A	19	0.813	(0.074)	5.603	0	∞	∞
E-B	20	0.756	(0.068)	5.077	0	∞	∞
E-C	22	0.817	(0.052)	5.554	0	∞	∞
E-D	20	0.821	(0.076)	5.913	1	∞	∞

D.III Genetic structure, differentiation and gene flow

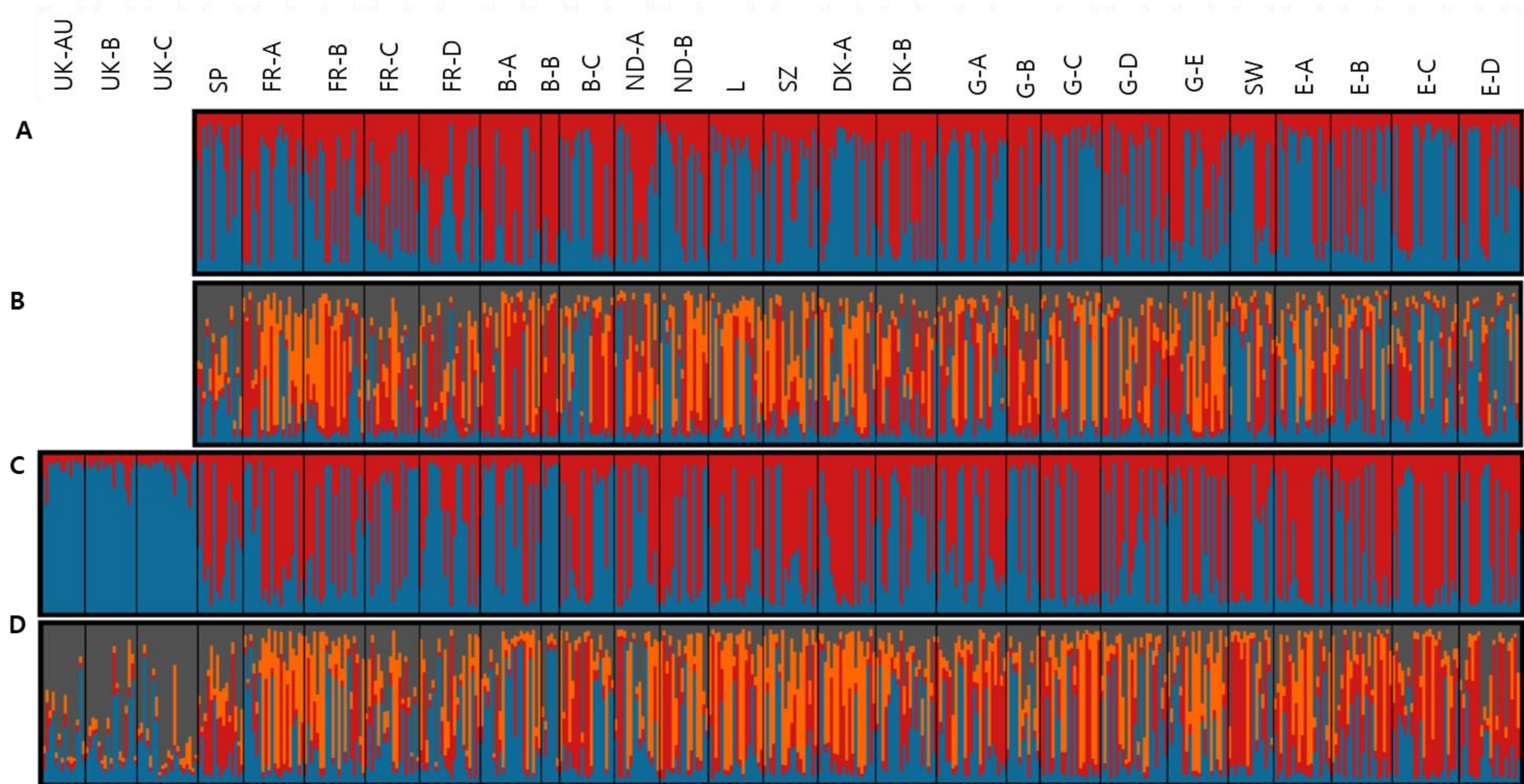


Figure D.1 Bayesian cluster analysis in STRUCTURE. Individual membership coefficients for *M. jurtina* across (A) Mainland European sites where K=2 (B) Mainland European sites where K=4 (C) Mainland European sites and three UK sites where K=2 (D) Mainland European sites and three UK sites where K=4.

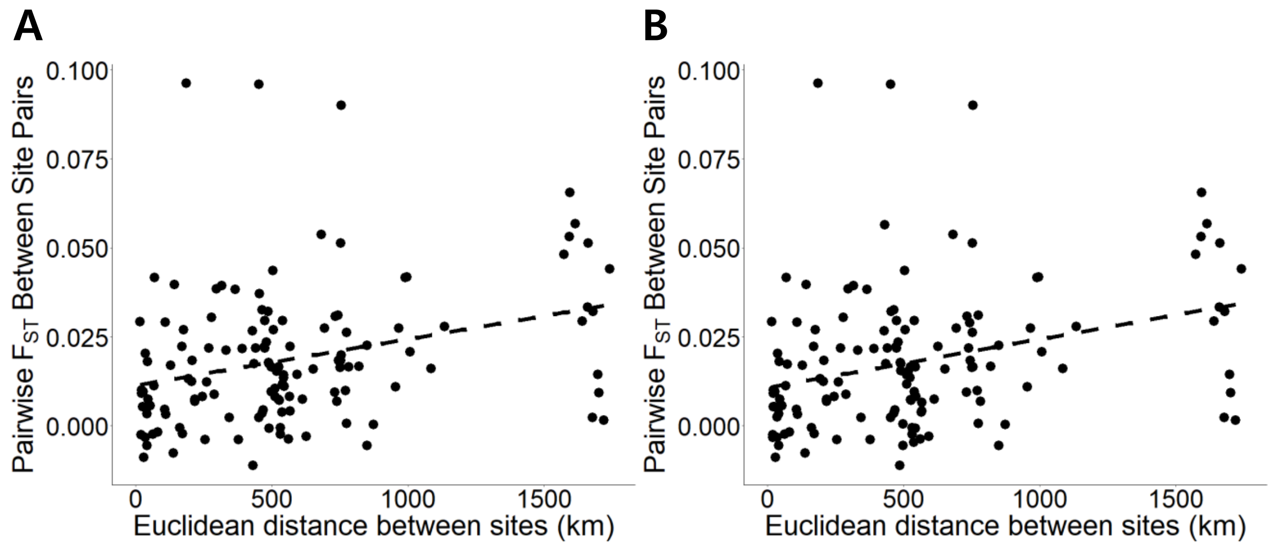


Figure D.2 Pairwise F_{ST} values for *M. jurtina* plotted against Euclidean distances for (A) 24 sites across mainland Europe (μ = 0.023, variance = 0.0004, p = 0.080, Mantel r value =0.183) and (B) 24 sites across mainland Europe and three sites in southern England (μ = 0.030, variance = 0.0007, p = 0.777, Mantel r value =-0.085).

Table D.5a Pairwise F_{ST} values between 24 pairs of sites across mainland Europe. Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 7020 permutations Indicative adjusted nominal level (5%) for multiple comparisons is: 0.000181

	SP	FR-A	FR-B	FR-C	FR-D	B-A	B-B	B-C	ND-A	ND-B	L	SZ	DK-A	DK-B	G-A	G-B	G-C	G-D	G-E	S-W	E-A	E-B	E-C	E-D
SP	*0.000*	NS	*	NS	NS	*	NS	NS	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
FR-A	0.042	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
FR-B	**0.038**	-0.008	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	NS
FR-C	0.030	0.012	0.008	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	*	NS	NS	NS	*	NS	NS	NS	*	*	*	*
FR-D	0.041	0.008	0.009	-0.001	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	*	NS	NS	NS	NS	NS	*	*
B-A	**0.065**	-0.001	0.001	0.002	-0.001	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
B-B	0.097	0.042	0.038	0.039	0.020	0.011	*0.000*	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
B-C	0.021	0.009	0.007	0.007	0.001	0.005	0.040	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
ND-A	**0.066**	-0.004	-0.004	0.027	0.017	-0.002	0.029	0.018	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS
ND-B	**0.052**	0.007	0.002	0.033	0.022	0.017	0.042	0.022	0.018	*0.000*	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
L	**0.064**	-0.002	0.003	0.038	0.039	0.027	0.096	0.031	0.016	0.015	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*
SZ	**0.090**	0.021	0.013	**0.038**	0.024	0.022	**0.096**	0.030	0.017	**0.044**	0.022	*0.000*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*
DK-A	**0.045**	-0.003	0.002	**0.032**	**0.029**	0.018	0.054	0.016	-0.004	0.009	0.015	**0.024**	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
DK-B	**0.051**	0.011	0.011	0.018	0.020	0.008	0.037	0.015	0.003	0.037	0.046	0.041	0.008	*0.000*	NS	NS	NS	NS	NS	NS	*	*	*	*
G-A	**0.041**	0.012	0.008	0.026	0.017	0.014	0.032	-0.003	0.014	0.012	0.023	0.036	0.014	0.034	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS
G-B	**0.081**	0.010	0.004	0.031	0.007	-0.001	-0.011	0.022	0.000	0.013	0.047	0.037	0.017	0.018	0.020	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS
G-C	**0.043**	0.001	-0.005	**0.029**	**0.022**	0.016	0.057	0.007	-0.001	0.004	-0.001	0.016	-0.006	0.022	0.003	0.017	*0.000*	NS	NS	NS	NS	NS	NS	NS
G-D	**0.049**	-0.001	-0.004	0.010	0.001	-0.002	0.022	0.008	-0.008	0.015	0.014	0.020	0.007	0.012	-0.003	0.010	0.006	*0.000*	NS	NS	NS	NS	NS	*
G-E	**0.059**	0.017	0.007	0.031	0.009	0.018	0.027	0.004	0.008	0.015	0.037	0.031	0.015	0.031	0.010	-0.002	0.009	0.007	*0.000*	NS	NS	NS	NS	NS
S-W	**0.044**	0.000	0.011	0.028	0.016	0.017	0.051	-0.006	0.016	0.028	0.023	0.034	0.013	0.012	0.005	0.030	0.002	0.004	0.022	*0.000*	NS	NS	NS	NS
E-A	**0.043**	**0.028**	0.014	**0.035**	0.016	0.032	**0.057**	0.002	0.016	0.025	0.043	0.032	**0.021**	**0.038**	0.008	0.019	0.006	0.017	0.000	0.019	*0.000*	NS	NS	NS
E-B	**0.073**	**0.050**	**0.043**	**0.057**	0.026	0.051	**0.066**	0.009	**0.053**	0.039	**0.069**	0.049	**0.042**	**0.064**	0.013	0.024	0.026	0.038	0.009	0.030	-0.003	*0.000*	NS	NS
E-C	**0.051**	**0.035**	**0.015**	**0.029**	**0.017**	**0.029**	**0.048**	0.002	0.024	**0.023**	**0.041**	0.038	**0.027**	**0.034**	0.013	0.020	0.006	0.018	0.007	0.018	-0.006	0.003	*0.000*	NS
E-D	**0.052**	**0.044**	0.021	**0.037**	**0.032**	0.033	**0.053**	0.015	0.027	0.026	**0.044**	**0.054**	**0.042**	**0.050**	0.019	0.026	0.012	**0.032**	0.026	0.029	0.005	0.029	-0.009	*0.000*

Table D.5b Pairwise F_{ST} values between 24 pairs of sites across mainland Europe and three from the south of England. Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 7020 permutations Indicative adjusted nominal level (5%) for multiple comparisons is: 0.000142.

	UK-AU	UK-B	UK-C	SP	FR-A	FR-B	FR-C	FR-D	B-A	B-B	B-C	ND-A	ND-B	L	SZ	DK-A	DK-B	G-A	G-B	G-C	G-D	G-E	S-W	E-A	E-B	E-C	E-D
UK-AU	0.000	NS	NS	*	*	NS	NS	NS	*	NS	NS	*	*	*	*	*	*	NS	NS	*	NS	NS	*	*	*	*	*
UK-B	-0.003	*0.000*	NS	NS	*	*	NS	*	NS	NS	*	*	*	*	*	*	*	NS	NS	*	*	*	*	*	*	*	*
UK-C	0.003	0.005	*0.000*	*	*	*	*	*	*	NS	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
SP	**0.090**	0.084	**0.067**	*0.000*	NS	NS	NS	NS	*	NS	NS	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
FR-A	**0.070**	**0.068**	**0.054**	0.042	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
FR-B	0.062	**0.061**	**0.047**	0.038	-0.008	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	NS
FR-C	0.051	0.052	**0.034**	0.030	0.012	0.008	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
FR-D	0.033	**0.044**	**0.030**	0.041	0.008	0.009	-0.001	*0.000*	NS	NS	NS	NS	NS	*	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	*	*	*
B-A	**0.039**	0.034	**0.028**	**0.065**	-0.001	0.001	0.002	-0.001	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
B-B	-0.006	0.018	0.008	0.097	0.042	0.038	0.039	0.020	0.011	*0.000*	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS
B-C	0.048	**0.055**	**0.040**	0.021	0.009	0.007	0.007	0.001	0.005	0.040	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
ND-A	**0.064**	**0.054**	**0.043**	**0.066**	-0.004	-0.004	0.027	0.017	-0.002	0.029	0.018	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS
ND-B	**0.060**	**0.053**	**0.053**	**0.052**	0.007	0.002	0.033	0.022	0.017	0.042	0.022	0.018	*0.000*	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
L	**0.119**	**0.108**	**0.094**	**0.064**	-0.002	0.003	0.038	**0.039**	0.027	0.096	0.031	0.016	0.015	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
SZ	**0.115**	**0.117**	**0.089**	**0.090**	0.021	0.013	**0.038**	0.024	0.022	**0.096**	0.030	0.017	**0.044**	0.022	*0.000*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*
DK-A	**0.085**	**0.079**	**0.066**	**0.045**	-0.003	0.002	0.032	0.029	0.018	0.054	0.016	-0.004	0.009	0.015	**0.024**	*0.000*	NS	NS	NS	NS	NS	NS	NS	*	*	*	*
DK-B	**0.063**	**0.056**	**0.039**	**0.051**	0.011	0.011	0.018	**0.020**	0.008	0.037	0.015	0.003	0.037	0.046	0.041	0.008	*0.000*	NS	NS	NS	NS	NS	NS	*	*	*	*
G-A	0.049	0.059	**0.052**	**0.041**	0.012	0.008	0.026	0.017	0.014	0.032	-0.003	0.014	0.012	0.023	0.036	0.014	0.034	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS	NS
G-B	0.028	0.025	**0.023**	**0.081**	0.010	0.004	0.031	0.007	-0.001	-0.011	0.022	0.000	0.013	0.047	0.037	0.017	0.018	0.020	*0.000*	NS	NS	NS	NS	NS	NS	NS	NS
G-C	**0.070**	**0.069**	**0.059**	**0.043**	0.001	-0.005	0.029	0.022	0.016	0.057	0.007	-0.001	0.004	-0.001	0.016	-0.006	0.022	0.003	0.017	*0.000*	NS	NS	NS	NS	NS	NS	NS
G-D	0.056	**0.055**	**0.040**	**0.049**	-0.001	-0.004	0.010	0.001	-0.002	0.022	0.008	-0.008	0.015	0.014	0.020	0.007	0.012	-0.003	0.010	0.006	*0.000*	NS	NS	NS	*	NS	*
G-E	0.044	**0.052**	**0.040**	**0.059**	0.017	0.007	0.031	0.009	0.018	0.027	0.004	0.008	0.015	0.037	0.031	0.015	0.031	0.010	-0.002	0.009	0.007	*0.000*	NS	NS	NS	NS	NS
S-W	**0.071**	**0.071**	**0.066**	**0.044**	0.000	0.011	0.028	0.016	0.017	0.051	-0.006	0.016	0.028	0.023	0.034	0.013	0.012	0.005	0.030	0.002	0.004	0.022	*0.000*	NS	NS	NS	*
E-A	**0.053**	**0.067**	**0.050**	**0.043**	**0.028**	0.014	**0.035**	0.016	**0.032**	0.057	0.002	0.016	0.025	**0.043**	0.032	**0.021**	**0.038**	0.008	0.019	0.006	0.017	0.000	0.019	*0.000*	NS	NS	NS
E-B	**0.066**	**0.087**	**0.079**	**0.073**	**0.050**	**0.043**	**0.057**	**0.026**	**0.051**	**0.066**	0.009	**0.053**	0.039	**0.069**	0.049	**0.042**	**0.064**	0.013	0.024	0.026	**0.038**	0.009	0.030	-0.003	*0.000*	NS	NS
E-C	**0.060**	**0.067**	**0.057**	**0.051**	**0.035**	**0.015**	**0.029**	**0.017**	**0.029**	0.048	0.002	0.024	**0.023**	**0.041**	0.038	**0.027**	**0.034**	0.013	0.020	0.006	0.018	0.007	0.018	-0.006	0.003	*0.000*	NS
E-D	**0.069**	**0.067**	**0.058**	**0.052**	**0.044**	0.021	**0.037**	**0.032**	**0.033**	0.053	0.015	0.027	0.026	**0.044**	**0.054**	**0.042**	**0.050**	0.019	0.026	0.012	**0.032**	0.026	**0.029**	0.005	0.029	-0.009	*0.000*

Appendix E Chapter 4 – Bottleneck analysis results

Table E1 Summary of the bottleneck results detected in *M. jurtina* populations across Europe. Strong evidence of population bottlenecks would be the majority of statistical tests being significant combined with a mode shift.

Population	Change in population size	Mode shift	Number of significant tests out of max of 9
UK-ARN	No	No	2
UK-ARS	No	No	0
UK-AU	No	No	3
UK-B	No	No	0
UK-C	No	No	1
UK-CH	No	No	3
UK-D	No	No	0
UK-HP	No	No	4
UK-LC	No	No	1
UK-LW	No	No	2
UK-MC	No	No	1
UK-PF	No	No	1
UK-SD	No	No	0
UK-TC	No	No	1
UK-WW	No	No	1
SP	Possible Bottleneck	No	6
FR-A	No	No	0
FR-F	Possible Bottleneck	No	5
FR-K	No	No	4
FR-M	No	No	2
B-A	No	No	2
B-B	No	No	0
B-C	No	No	4
ND-A	No	No	1
ND-B	No	No	1
L	No	No	1
SZ	Possible Bottleneck	No	5
DK-A	Possible Bottleneck	No	5
DK-B	No	No	2
G-A	No	No	4
G-B	No	No	4
G-C	Possible Bottleneck	No	5
G-D	No	No	2
G-E	No	No	3
SW	Possible Bottleneck	No	6
E-A	Possible Bottleneck	No	5
E-B	No	No	1
E-C	No	No	4
E-D	Possible Bottleneck	No	5

Table E2 Distribution of allele frequencies in *M. jurtina* across 39 populations across Europe. Alleles are proportioned into ten frequency classes. L shaped distribution of allele frequencies indicate no bottleneck has occurred.

Population	0.0 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0
UK-ARN	0.673	0.218	0.055	0	0	0.036	0	0.018	0	0
UK-ARS	0.592	0.245	0.102	0.02	0.02	0	0	0	0.02	0
UK-AU	0.608	0.235	0.098	0	0.02	0.02	0	0	0.02	0
UK-B	0.7	0.16	0.08	0.02	0	0.02	0	0	0.02	0
UK-C	0.636	0.182	0.127	0.036	0	0	0	0.018	0	0
UK-CH	0.649	0.211	0.105	0	0.018	0	0.018	0	0	0
UK-D	0.556	0.315	0.093	0	0	0.019	0	0	0.019	0
UK-HP	0.607	0.25	0.089	0.018	0.018	0	0.018	0	0	0
UK-LC	0.597	0.323	0.048	0	0.016	0	0	0.016	0	0
UK-LW	0.552	0.328	0.086	0.017	0	0	0	0	0.017	0
UK-MC	0.679	0.25	0.018	0	0.018	0.018	0	0	0.018	0
UK-PF	0.778	0.153	0.028	0.014	0.014	0	0	0	0.014	0
UK-SD	0.574	0.296	0.019	0.093	0	0	0	0	0	0.019
UK-TC	0.529	0.275	0.118	0.059	0	0	0	0	0.02	0
UK-WW	0.519	0.352	0.074	0	0.019	0.019	0	0.019	0	0
SP	0.4	0.4	0.14	0.04	0	0.02	0	0	0	0
FR-A	0.643	0.179	0.071	0.107	0	0	0	0	0	0
FR-F	0.551	0.184	0.184	0.082	0	0	0	0	0	0
FR-K	0.625	0.286	0.036	0.036	0	0.018	0	0	0	0
FR-M	0.564	0.309	0.036	0.073	0	0.018	0	0	0	0
B-A	0.5	0.346	0.077	0.038	0.019	0	0.019	0	0	0
B-B	0.4	0.333	0.067	0.1	0.033	0.033	0	0	0	0.033
B-C	0.577	0.231	0.135	0.058	0	0	0	0	0	0
ND-A	0.604	0.208	0.094	0.038	0.057	0	0	0	0	0
ND-B	0.592	0.224	0.082	0.061	0.041	0	0	0	0	0
L	0.588	0.255	0.098	0	0.02	0.039	0	0	0	0
SZ	0.551	0.286	0.061	0.041	0.02	0.041	0	0	0	0
DK-A	0.551	0.224	0.163	0.02	0.041	0	0	0	0	0
DK-B	0.574	0.278	0.074	0.037	0.019	0.019	0	0	0	0
G-A	0.6	0.236	0.091	0.036	0.036	0	0	0	0	0
G-B	0.575	0.175	0.125	0.075	0	0.025	0.025	0	0	0
G-C	0.52	0.28	0.12	0.06	0.02	0	0	0	0	0
G-C	0.707	0.155	0.069	0.052	0.017	0	0	0	0	0
G-E	0.558	0.308	0.058	0.019	0.058	0	0	0	0	0
SW	0.409	0.295	0.205	0.091	0	0	0	0	0	0
E-A	0.444	0.444	0	0.089	0.022	0	0	0	0	0
E-B	0.556	0.244	0.044	0.067	0.067	0.022	0	0	0	0
E-C	0.62	0.14	0.16	0.08	0	0	0	0	0	0
E-D	0.545	0.291	0.091	0.055	0.018	0	0	0	0	0

Table E3 P-values for all Bottleneck test results. Significant values are shown in bold. IA = Infinite allele model, TPM = Two Phase Model, SMM = Stepwise Mutation Model.

Population	Sign Test			Standard Differences Test			Wilcoxon Test		
	IA	TPM	SMM	IA	TPM	SMM	IA	TPM	SMM
UK-ARN	0.234	0.546	**0.050**	0.076	0.383	0.073	**0.031**	0.688	0.078
UK-ARS	0.218	0.223	0.460	0.067	0.333	0.229	0.156	0.438	1.000
UK-AU	0.507	0.192	**0.007**	0.499	0.116	**0.001**	0.844	0.109	**0.016**
UK-B	0.230	0.465	0.168	0.102	0.419	0.157	0.109	1.000	0.438
UK-C	0.239	0.237	0.480	0.053	0.257	0.233	**0.047**	0.438	0.688
UK-CH	**0.049**	0.527	0.492	**0.029**	0.165	0.377	**0.016**	0.109	1.000
UK-D	0.217	0.253	0.517	0.152	0.488	0.112	0.438	0.438	0.844
UK-HP	**0.041**	0.234	0.186	**0.029**	0.148	0.400	**0.016**	**0.047**	0.563
UK-LC	0.231	0.450	0.188	0.062	0.275	0.243	**0.031**	0.563	0.563
UK-LW	0.224	0.442	0.456	**0.045**	0.178	0.387	**0.047**	0.563	0.844
UK-MC	0.547	0.180	0.209	0.386	0.192	**0.003**	0.688	0.563	0.078
UK-PF	0.237	0.197	0.191	0.082	0.486	**0.011**	0.109	0.844	0.078
UK-SD	0.538	0.536	0.468	0.200	0.472	0.091	0.563	0.688	0.563
UK-TC	0.213	0.452	0.197	0.106	0.354	0.176	**0.047**	0.688	0.156
UK-WW	0.549	0.433	0.202	0.186	0.337	**0.009**	0.156	0.844	0.156
SP	**0.046**	**0.047**	0.526	**0.009**	**0.048**	0.214	**0.016**	**0.016**	0.078
FR-A	0.235	0.541	0.455	0.084	0.478	0.053	0.078	0.688	0.438
FR-F	**0.049**	**0.046**	0.466	**0.013**	0.085	0.472	**0.016**	**0.016**	0.844
FR-K	**0.046**	0.231	0.490	**0.014**	0.090	0.488	**0.016**	**0.047**	1.000
FR-M	0.230	0.232	0.474	**0.022**	0.136	0.393	**0.031**	0.078	1.000
B-A	0.222	0.550	0.471	**0.035**	0.166	0.315	**0.031**	0.109	0.688
B-B	0.445	0.481	0.483	0.320	0.454	0.283	0.563	1.000	0.688
B-C	**0.045**	0.217	0.483	**0.016**	0.092	0.476	**0.016**	**0.047**	1.000
ND-A	0.238	0.438	0.192	0.282	0.304	**0.009**	0.438	1.000	0.109
ND-B	0.240	0.459	0.185	0.069	0.347	0.117	**0.047**	0.688	0.438
L	0.230	0.527	0.496	0.057	0.308	0.177	**0.047**	0.688	0.844
SZ	**0.045**	0.547	**0.005**	0.112	0.463	**0.037**	**0.016**	0.688	**0.016**
DK-A	**0.046**	0.221	0.253	**0.010**	**0.050**	0.342	**0.016**	**0.031**	0.438
DK-B	0.237	0.231	0.188	**0.039**	0.239	0.190	**0.031**	0.109	0.688
G-A	**0.048**	0.223	0.189	**0.021**	0.161	0.258	**0.016**	**0.047**	0.563
G-B	**0.036**	0.226	0.474	**0.036**	0.137	0.390	**0.016**	**0.047**	0.844
G-C	**0.045**	0.241	0.534	**0.006**	**0.048**	0.284	**0.016**	**0.031**	0.563
G-C	0.235	0.533	0.498	**0.036**	0.210	0.194	**0.031**	0.156	0.844
G-E	**0.048**	0.542	0.478	**0.024**	0.120	0.367	**0.016**	0.109	0.844
SW	**0.043**	**0.045**	0.526	**0.006**	**0.034**	0.157	**0.016**	**0.016**	0.156
E-A	**0.046**	0.237	0.231	**0.006**	**0.037**	0.242	**0.016**	**0.031**	0.438
E-B	0.548	0.469	0.192	0.266	0.160	**0.000**	0.563	1.000	0.156
E-C	**0.048**	0.233	0.473	**0.012**	0.094	0.488	**0.016**	**0.031**	1.000
E-D	**0.045**	**0.046**	0.474	**0.020**	0.131	0.379	**0.016**	**0.016**	0.688

Appendix F Chapter 5 – Site characteristics and correlations

Table F1 Landscape attribute data for each of the fifteen sites used in the genetic and drought analyses. Data compiled from Natural England priority habitat maps and a 50m resolution digital elevation map (Morris & Flavin 1990).

Slope Angle = Degrees from horizontal, such that 0 = flat, 90 = vertical. Aspect (East) = Mean Eastness of aspect in landscape around site (Eastness = sin((aspectxPI)/180), such that 1 = due East, -1 = due West). Aspect (North) = Mean Northness of aspect in landscape around site (Northness = cos((aspectxPI)/180), such that 1 = due North, -1 = due South). Altitude = Mean height above sea level (m)

Site	% Chalk grass cover	Slope Angle		Aspect (East)		Aspect (North)		Altitude		Northing	Easting
		Mean	SD	Mean	SD	Mean	SD	Mean	SD		
ARN	9.39	10.77	5.19	-0.26	0.64	0.32	0.65	129.16	49.66	197086	472827
ARS	14.16	9.29	5.04	-0.48	0.51	0.44	0.56	130.88	49.34	196060	472285
AU	14.56	5.39	3.32	0.38	0.77	0.24	0.46	101.40	41.96	183700	454500
B	0.00	3.04	3.03	0.28	0.49	-0.03	0.83	108.23	30.73	165000	450800
C	0.00	1.59	0.89	0.05	0.74	0.46	0.5	110.88	40.92	151900	466300
CH	7.56	8.18	6.42	-0.56	0.59	0.27	0.52	134.80	50.09	206700	484700
D	4.70	8.9	3.87	0.35	0.36	0.18	0.85	141.69	45.81	209500	490000
HP	0.00	0.69	0.47	-0.54	0.75	-0.03	0.39	90.82	45.90	190000	461500
LC	21.17	10.89	3.92	0.31	0.62	0.61	0.39	96.71	39.17	180900	458700
LW	0.00	3.71	1.82	-0.27	0.49	-0.28	0.79	68.79	22.32	192300	456200
MC	0.00	0.45	0.62	0.37	0.75	0.32	0.46	83.89	34.73	174100	463700
PF	0.00	1.27	0.9	0.5	0.33	-0.24	0.77	87.31	29.87	161000	461500
SD	11.46	7.1	3.53	-0.23	0.42	0.09	0.88	117.68	52.23	191500	467500
TC	5.73	4.74	3.21	0.59	0.37	0.34	0.63	144.72	44.29	208800	490400
WW	9.39	1.55	1.42	0.01	0.5	0.61	0.61	80.17	19.97	209631	446434

Table F2 Pearson's Rank Correlation coefficients for site characteristics calculated using all 540 UKBMS sites used in the analyses. Each correlation calculated with 1381 degrees of freedom.

Variable 1	Variable 2	Correlation	p-value	t value
Chalk Area	Slope Mean	0.108	<0.001	4.0315
Chalk Area	Slope SD	0.095	<0.001	3.5544
Chalk Area	Mean Aspect (North)	0.011	0.683	0.409
Chalk Area	Aspect (North) SD	0.015	0.577	0.557
Chalk Area	Mean Aspect (East)	-0.056	0.038	-2.08
Chalk Area	Aspect (East) SD	-0.017	0.522	-0.641
Chalk Area	Mean Altitude	-0.063	0.019	-2.353
Chalk Area	Altitude SD	-0.032	0.234	-1.19
Chalk Area	Northing	-0.048	0.073	-1.796
Chalk Area	Easting	0.081	0.0025	3.029
Mean Slope Angle	Slope SD	0.763	<0.001	43.896
Mean Slope Angle	Mean Aspect (North)	-0.017	0.522	-0.64
Mean Slope Angle	Aspect (North) SD	0.007	0.06	0.271
Mean Slope Angle	Mean Aspect (East)	0.005	0.843	0.199
Mean Slope Angle	Aspect (East) SD	0.017	0.539	0.614
Mean Slope Angle	Mean Altitude	0.483	<0.001	20.524
Mean Slope Angle	Altitude SD	0.582	<0.001	26.618
Mean Slope Angle	Northing	0.064	0.018	2.369
Mean Slope Angle	Easting	-0.387	<0.001	15.638

Table F3 Pearson's Rank Correlation coefficients for site characteristics calculated using the 15 sites used in the genetic and drought analyses. Correlations calculated with 13 degrees of freedom.

Variable 1	Variable 2	Correlation	p-value	t value
Chalk Area	Slope Mean	0.811	<0.001	5.006
Chalk Area	Slope SD	0.667	0.007	3.237
Chalk Area	Mean Aspect (North)	0.452	0.091	1.829
Chalk Area	Aspect (North) SD	-0.265	0.339	-0.992
Chalk Area	Mean Aspect (East)	-0.051	0.857	-0.184
Chalk Area	Aspect (East) SD	0.094	0.739	0.34
Chalk Area	Mean Altitude	0.357	0.192	1.377
Chalk Area	Altitude SD	0.528	0.043	2.242
Chalk Area	Northing	0.225	0.421	0.832
Chalk Area	Easting	0.152	0.588	0.556
Mean Slope Angle	Slope SD	0.889	<0.001	7.012
Mean Slope Angle	Mean Aspect (North)	0.335	0.223	1.281
Mean Slope Angle	Aspect (North) SD	0.067	0.813	0.241
Mean Slope Angle	Mean Aspect (East)	-0.209	0.455	-0.77
Mean Slope Angle	Aspect (East) SD	-0.155	0.582	-0.565
Mean Slope Angle	Mean Altitude	0.613	0.015	2.798
Mean Slope Angle	Altitude SD	0.587	0.022	2.611
Mean Slope Angle	Northing	0.447	0.095	1.803
Mean Slope Angle	Easting	0.455	0.089	1.841

Appendix G Chapter 5 – Phenology

Model outputs for Eqn. 1

Table G1 Linear mixed effects model outputs for Eqn. 1, showing the effects of geology, topography, site abundance and Scaled Northing on measures of *M. jurtina* phenology. Models are repeated for each measure of phenology.

Response Variable	Marginal R^2	Conditional R^2	Fixed Effects	Estimate	Lower CI	Upper CI	Standard Error	Degrees of freedom	t value	p-value
Start Date	0.07	0.61	Intercept	87.1	85.12	89.11	1.01	70.32	85.99	<0.001
			Chalk cover %	0.07	0.02	0.12	0.03	414.36	2.61	0.009
			Slope Angle	0.36	0.22	0.5	0.07	420.69	5	<0.001
			Abundance	0.002	0.001	0.002	0.0002	4563.85	8.97	<0.001
			Northing	1.42	0.92	1.92	0.26	444.28	5.53	<0.001
			Easting	0.1	-0.46	0.66	0.29	449.52	0.36	0.722
			Aspect	0.05	-0.44	0.53	0.25	447.78	0.18	0.854
			Altitude	0.002	-0.008	0.01	0.005	470.35	0.38	0.703
Mean Date	0.12	0.77	Intercept	109.07	107.1	111.05	1.01	95.37	108.17	<0.001
			Chalk cover %	0.14	0.08	0.2	0.03	460.92	4.51	<0.001
			Slope Angle	0.62	0.45	0.79	0.09	467.64	7.2	<0.001
			Abundance	0.001	0.0004	0.001	0.0002	5624.12	4.1	<0.001
			Northing	-0.31	-0.9	0.29	0.3	479.18	-1.01	0.315
			Easting	0.35	-0.32	1.01	0.34	486.01	1.02	0.310
			Aspect	-0.29	-0.87	0.29	0.3	485.15	-0.96	0.336
			Altitude	0	-0.003	0.019	0.01	500.45	1.38	0.169
End Date	0.15	0.71	Intercept	125.67	123.43	127.91	1.14	132.1	109.82	<0.001
			Chalk cover %	0.19	0.12	0.27	0.04	451.12	4.89	<0.001
			Slope Angle	0.81	0.6	1.03	0.11	458.21	7.38	<0.001
			Abundance	0.00001	-0.0004	0.0005	0.0002	5534.54	0.06	0.952
			Northing	-2.03	-2.79	-1.27	0.39	472.11	-5.19	<0.001
			Easting	0.29	-0.56	1.15	0.44	479.23	0.67	0.503
			Aspect	-0.5	-1.25	0.24	0.38	478.09	-1.32	0.188
			Altitude	0.01	-0.003	0.03	0.008	495.19	1.53	0.127

Table G1 continued.

Response Variable	Marginal R^2	Conditional R^2	Fixed Effects	Estimate	Lower CI	Upper CI	Standard Error	Degrees of freedom	t value	p-value
Range	0.13	0.47	Intercept	38.37	36.67	40.06	0.87	146.56	44.24	<0.001
			Chalk cover %	0.13	0.07	0.19	0.03	379.49	4.27	<0.001
			Slope Angle	0.43	0.27	0.6	0.09	383.69	5.11	<0.001
			Abundance	-0.001	-0.002	-0.001	0.0003	3703.51	-4.03	<0.001
			Northing	-3.34	-3.93	-2.75	0.3	412.6	-11.002	<0.001
			Easting	0.16	-0.51	0.83	0.34	414.29	0.47	0.640
			Aspect	-0.52	-1.1	0.06	0.3	413.08	-1.76	0.080
			Altitude	0.01	-0.002	0.02	0.006	437.7	1.68	0.092

Table G2 Linear mixed effects model outputs for Eqn. 1, using the 1st and 99th percentiles for start and end dates, showing the effects of geology, topography, site abundance and Scaled Northing on measures of *M. jurtina* phenology. Models are repeated for each measure of phenology.

Response Variable	Marginal R^2	Conditional R^2	Fixed Effects	Estimate	Lower CI	Upper CI	Standard Error	Degrees of freedom	t value	p-value
Start Date	0.05	0.46	Intercept	73.32	71.13	75.54	1.12	64.86	65.44	<0.001
			Chalk cover %	0.04	-0.01	0.10	0.03	328.12	1.67	0.096
			Slope Angle	0.15	0.01	0.30	0.07	329.50	2.03	0.043
			Abundance	0.002	0.001	0.003	0.0003	2787.80	7.94	<0.001
			Northing	2.28	1.76	2.81	0.27	361.78	8.48	<0.001
			Easting	0.16	-0.43	0.75	0.30	358.55	0.52	0.605
			Aspect	0.17	-0.34	0.69	0.26	358.54	0.65	0.516
			Altitude	-0.004	-0.01	0.01	0.005	383.96	-0.66	0.512
Mean Date	0.12	0.77	Intercept	109.07	107.10	111.05	1.01	95.37	108.17	<0.001
			Chalk cover %	0.14	0.08	0.20	0.03	460.92	4.51	<0.001
			Slope Angle	0.62	0.45	0.79	0.09	467.64	7.20	<0.001
			Abundance	0.0007	0.0004	0.001	0.0002	5624.12	4.10	<0.001
			Northing	-0.31	-0.90	0.29	0.30	479.18	-1.01	0.315
			Easting	0.35	-0.32	1.01	0.34	486.01	1.02	0.310
			Aspect	-0.29	-0.87	0.29	0.30	485.14	-0.96	0.336
			Altitude	0.01	-0.003	0.019	0.01	500.44	1.38	0.169
End Date	0.15	0.64	Intercept	143.89	141.50	146.28	1.22	153.23	117.69	<0.001
			Chalk cover %	0.19	0.11	0.28	0.04	429.32	4.43	<0.001
			Slope Angle	0.87	0.63	1.10	0.12	436.41	7.09	<0.001
			Abundance	-0.0004	-0.001	0.0002	0.0003	5230.21	-1.46	0.145
			Northing	-3.15	-4.00	-2.31	0.43	454.06	-7.26	<0.001
			Easting	0.27	-0.68	1.22	0.49	460.90	0.56	0.575
			Aspect	-0.60	-1.43	0.23	0.43	459.39	-1.40	0.161
			Altitude	0.01	-0.006	0.03	0.008	478.77	1.29	0.199

Table G2. Continued

Response Variable	Marginal R^2	Conditional R^2	Fixed Effects	Estimate	Lower CI	Upper CI	Standard Error	Degrees of freedom	t value	p-value
Range	0.11	0.43	Intercept	70.46	67.58	73.31	1.46	130.95	48.15	<0.001
			Chalk cover %	0.15	0.06	0.25	0.05	348.37	3.13	0.002
			Slope Angle	0.68	0.41	0.95	0.14	351.48	4.93	<0.001
			Abundance	-0.002	-0.003	-0.0008	0.0005	3315.5	-3.81	<0.001
			Northing	-5.31	-6.28	-4.34	0.50	380.76	-10.69	<0.001
			Easting	0.08	-1.01	1.17	0.56	380.78	0.14	0.887
			Aspect	-0.75	-1.69	0.20	0.49	379.98	-1.53	0.126
			Altitude	0.02	-0.004	0.03	0.01	404.12	1.55	0.122

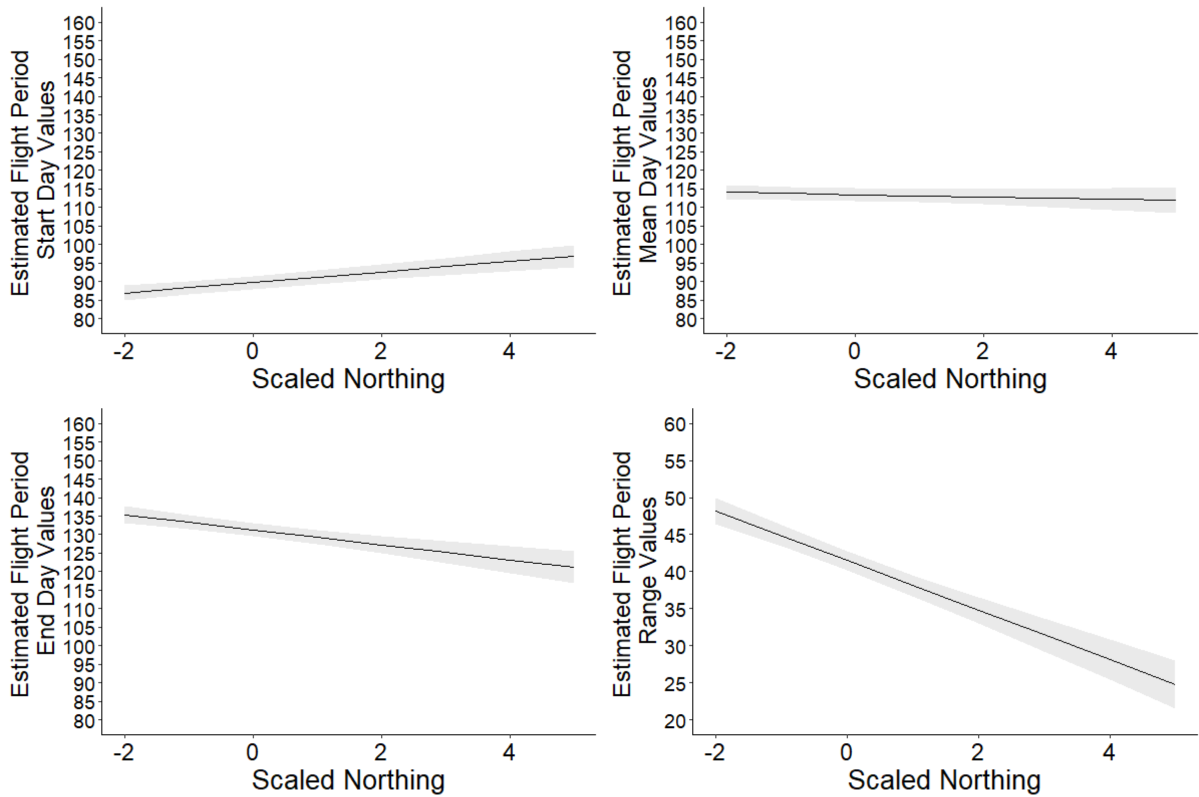


Figure G3 Estimated model values from equation 1 for four measures of phenology for *M. jurtina* in relation to Scaled Northing. Scaled Northing had a significant effect on all measures of phenology, with the exception of mean dates.

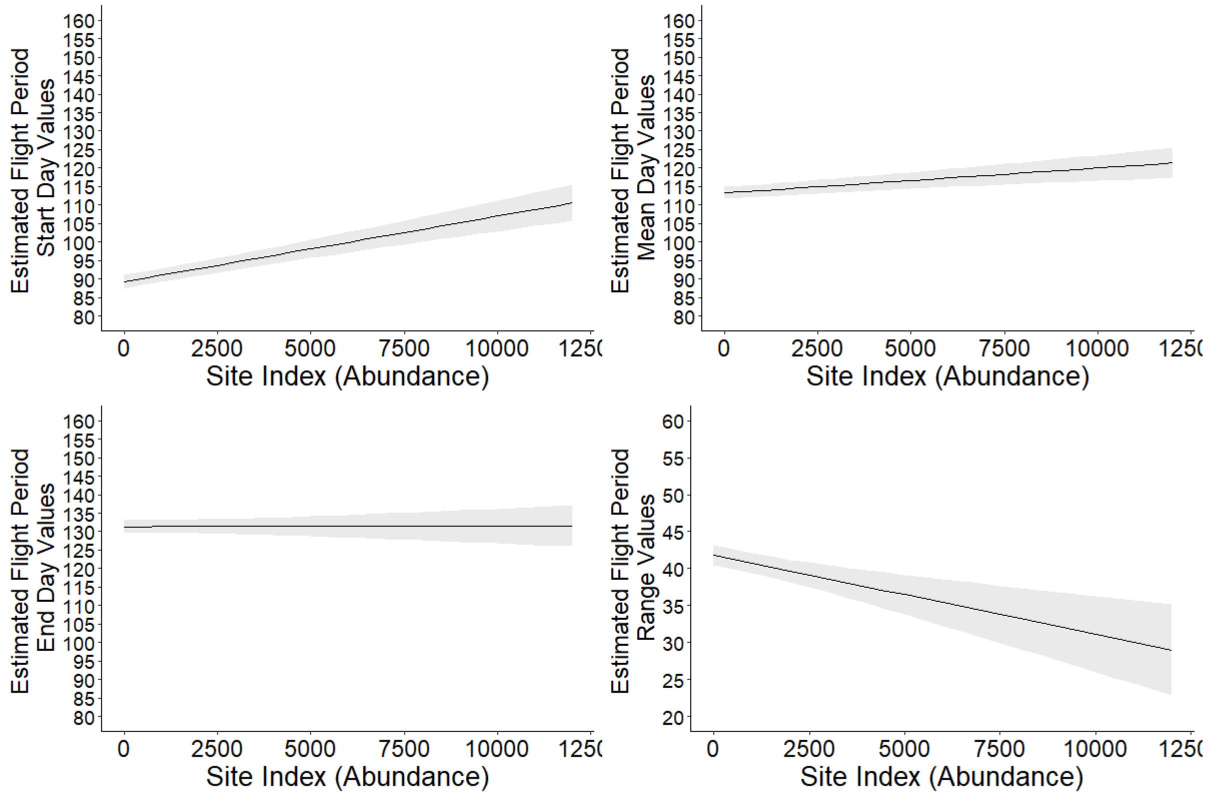


Figure G4 Estimated model values from equation 1 for four measures of phenology for *M. jurtina* in relation to mean butterfly abundance at each site. Abundance had a significant effect on all measures of phenology with the exception of flight period end dates.

Appendix H Chapter 5 – Population genetics

Table H1 Composite linkage disequilibrium test outputs for all locus pair combinations, calculated in Genepop v4.7 (Rousset 2008).

Locus Pair		Chi^2	df	p-Value
Mj7232	Mj5522	30.528	30	0.439
Mj7232	Mj0247	23.029	26	0.631
Mj5522	Mj0247	7.734	26	1.000
Mj7232	Mj4870	30.842	30	0.423
Mj5522	Mj4870	15.522	30	0.986
Mj0247	Mj4870	20.015	26	0.791
Mj7232	Mj7132	16.554	30	0.978
Mj5522	Mj7132	23.531	30	0.793
Mj0247	Mj7132	13.529	26	0.979
Mj4870	Mj7132	23.315	30	0.802
Mj7232	Mj5331	24.854	26	0.527
Mj5522	Mj5331	14.212	26	0.970
Mj0247	Mj5331	13.037	22	0.932
Mj4870	Mj5331	17.472	26	0.894
Mj7132	Mj5331	11.868	26	0.992

Table H2 Locus by populations estimated null allele frequencies for all sites per locus. Values in bold exceed 0.2 frequency of null alleles. Null allele frequencies calculated in Genepop v4.7 (Rousset 2008).

Locus	ARN	ARS	AU	B	C	CH	D	HP	LC	LW	MC	PF	SD	TC	WW	Mean
Mj7232	0.028	0.145	0.000	0.137	0.079	0.096	0.022	0.032	0.086	0.000	0.171	0.050	0.048	0.013	0.032	0.063
Mj5522	0.000	0.055	0.048	0.000	0.000	0.000	0.030	0.055	0.006	0.013	0.029	0.021	0.076	0.074	0.000	0.027
Mj0247	0.042	0.030	0.000	0.085	0.018	0.000	0.041	0.000	0.000	0.037	0.111	0.064	0.068	0.000	0.048	0.036
Mj4870	0.075	**0.205**	0.000	0.000	0.000	0.064	0.086	0.003	0.159	0.151	0.000	0.113	0.150	0.189	0.128	0.088
Mj7132	0.022	0.065	0.000	0.045	0.000	0.000	0.091	0.000	0.000	0.028	0.000	0.000	0.047	0.002	0.000	0.020
Mj5331	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.030	0.028	0.028	0.027	0.000	0.000	0.000	0.010

Table H3 Population-wide expected and observed heterozygosity, and percent difference ((E - O)/E*100), F_{IT}, F_{IS} and F_{ST} at each locus. Bartlett's K-squared: 0.03, df = 1, p-value = 0.8618.
H_e, H_O and percentage differences calculated in PopGenReport (Adamack & Gruber 2014). F_{IT}, F_{ST} and F_{IS} values calculated in Arlequin v 3.5.2.2 (Excoffier & Lischer 2010).

Locus	Number of samples	Number of alleles	H_e	H_o	H_e vs H_o % difference	F_{IT} (p-value)		F_{ST} (p-value)		F_{IS} (p-value)	
Mj7232	285	12	0.798	0.762	-4.488	0.048	0.035	0.002	0.999	0.046	0.049
Mj5522	281	12	0.862	0.809	-6.226	0.064	0.008	0.000	1.000	0.064	0.006
Mj0247	283	31	0.941	0.842	-10.614	0.105	0.000	0.000	1.000	0.105	0.000
Mj4870	282	6	0.370	0.279	-24.596	0.252	0.000	0.000	0.982	0.252	0.000
Mj7132	282	10	0.741	0.752	1.500	-0.013	0.692	0.007	0.911	-0.020	0.770
Mj5331	286	22	0.894	0.902	0.913	-0.007	0.667	0.002	1.000	-0.009	0.692
Mean	283	16	0.768	0.724	-7.252	0.075	-	0.002	-	0.073	-

Table H4 Sample sizes, genetic diversity, allelic richness, number of private alleles and effective population sizes for *M. jurtina* populations in the south of England. All values are estimated on a per population basis.

H_{exp} expected heterozygosity calculated in Arlequin, A_r = allelic richness calculated in FSTAT v2.9.4 (Goudet 1994), A_p = Private alleles calculated in PopGenRport, Ne(1) = effective population size estimated using the heterozygote excess method calculated in NeEstimator V2 (Do *et al.* 2014), Ne(2) = effective population size estimated using linkage disequilibrium method, calculated in NeEstimator.

Site	Sample Size	Mean H_{exp} (s.d.)		A_r	A_p	$N_e(1)$	$N_e(2)$
All Sites	287	0.764	(0.215)	-	9	-	-
Aston Rowant North (ARN)	21	0.760	(0.176)	8.278	0	∞	∞
Aston Rowant South (ARS)	17	0.753	(0.224)	7.765	0	∞	∞
Aston Upthorpe (AU)	14	0.713	(0.269)	8.500	0	∞	∞
Bowdown Forest (B)	17	0.738	(0.246)	8.039	0	∞	∞
Crabtree Plantation (C)	20	0.783	(0.199)	8.444	0	∞	∞
Coombe Hill (CH)	20	0.805	(0.145)	8.690	0	∞	∞
Dancersend (D)	15	0.776	(0.223)	8.806	0	∞	∞
Howbery Park (HP)	20	0.786	(0.163)	8.438	0	∞	∞
Lardon Chase (LC)	20	0.796	(0.176)	9.096	2	∞	∞
Little Whittenham (LW)	20	0.784	(0.229)	8.682	0	∞	∞
Moore Copse (MC)	16	0.736	(0.266)	8.858	1	∞	∞
Pamber Forest (PF)	37	0.771	(0.223)	9.030	5	∞	∞
Swyncombe Down (SD)	15	0.747	(0.279)	8.773	0	∞	∞
The Crong (TC)	15	0.771	(0.213)	8.286	1	∞	∞
Wytham Woods (WW)	20	0.745	(0.192)	8.097	0	∞	∞

Table H5 Pairwise F_{ST} values between fifteen pairs of sites across the south of England in 2017. Values below the diagonal = F_{ST} scores. Values above the diagonal indicate significance level p-values. P-values obtained after 2100 permutations, indicative adjusted nominal level (5%) for multiple comparisons is 0.000476. All values are non-significant.

	ARN	ARS	AU	B	C	CH	D	HP	LC	LW	MC	PF	SD	TC	WW
ARN	-	0.037	0.229	0.333	0.036	0.515	0.459	0.207	0.260	0.014	0.034	0.035	0.073	0.020	0.003
ARS	0.004	-	0.290	0.073	0.060	0.142	0.450	0.387	0.140	0.103	0.336	0.062	0.414	0.053	0.019
AU	0.004	-0.003	-	0.897	0.390	0.388	0.412	0.714	0.211	0.289	0.680	0.522	0.866	0.326	0.330
B	0.013	0.006	-0.003	-	0.098	0.092	0.306	0.896	0.389	0.167	0.487	0.882	0.958	0.451	0.055
C	0.007	-0.003	0.003	0.005	-	0.370	0.275	0.730	0.545	0.766	0.263	0.090	0.788	0.146	0.018
CH	0.001	0.012	0.011	0.020	0.002	-	0.127	0.339	0.719	0.390	0.411	0.123	0.485	0.141	0.027
D	0.001	-0.011	0.002	0.001	-0.003	0.012	-	0.411	0.864	0.843	0.929	0.268	0.619	0.600	0.034
HP	0.003	-0.004	-0.003	-0.009	-0.006	0.003	-0.002	-	0.436	0.600	0.845	0.891	0.947	0.467	0.171
LC	0.006	0.003	0.013	0.004	-0.003	0.003	-0.010	-0.003	-	0.929	0.255	0.255	0.731	0.177	0.105
LW	0.012	0.000	0.008	0.012	-0.007	0.004	-0.008	0.002	-0.008	-	0.359	0.435	0.932	0.031	0.008
MC	0.009	-0.003	-0.005	-0.003	0.001	0.014	-0.012	-0.006	0.005	0.006	-	0.552	0.761	0.380	0.135
PF	0.009	0.000	-0.005	-0.004	0.000	0.011	-0.006	-0.006	0.002	-0.001	-0.006	-	0.917	0.565	0.010
SD	0.014	-0.004	-0.005	-0.007	-0.008	0.008	-0.002	-0.005	0.001	-0.009	-0.003	-0.005	-	0.428	0.149
TC	0.015	0.004	0.003	0.005	-0.004	0.003	-0.002	-0.003	0.000	0.005	0.006	-0.002	-0.002	-	0.451
WW	0.007	0.009	-0.005	0.012	0.011	0.011	0.010	0.000	0.008	0.015	0.003	0.008	0.007	0.005	-

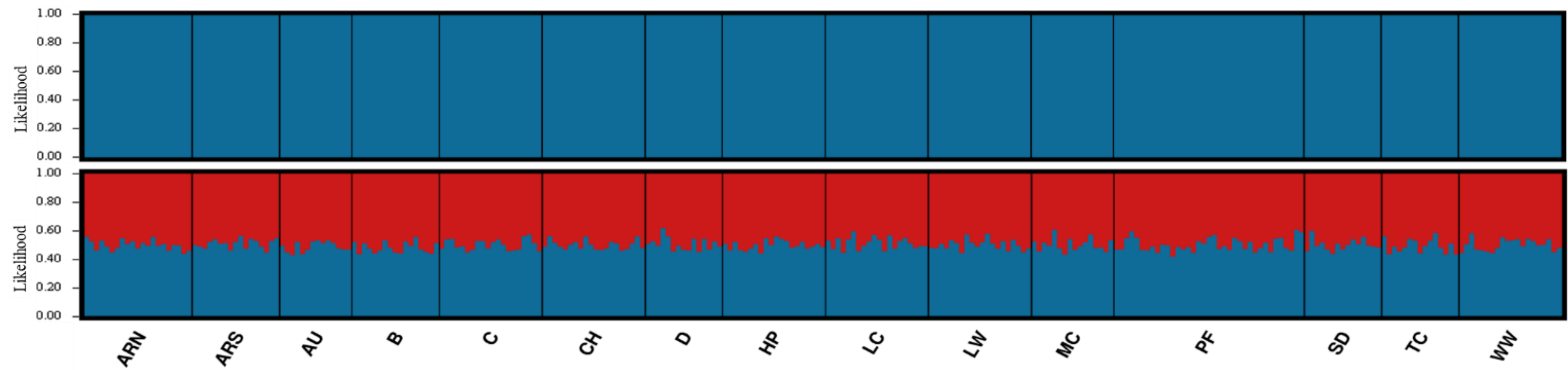


Figure H1 STRUCTURE individual assignment bar plots for *M. jurtina* individuals within the study area. Individuals split by site on the x axis and likelihood of assignment of the individual into genetic clusters on the y. Colours indicate different genetic clusters. Plot A: K = 1, Plot B: K = 2. All individuals have a roughly 50% chance of being assigned into either of the clusters where K=2, indicating no apparent population structuring, i.e. no individuals are more or less likely to be assigned to any K, therefore K = 1

References

A'Hara, S. & Cottrell, J. (2013). Development and characterisation of ten polymorphic microsatellite markers for the pine-tree lappet moth Dendrolimus pini (Lepidoptera: Lasiocampidae). *Conserv. Genet. Resour.*, 5, 1135–1137.

Aarnes, S., Fløystad, I., Schregel, J., Vindstad, O., Jepsen, J., Eiken, H., *et al.* (2015). Identification and evaluation of 21 novel microsatellite markers from the autumnal moth (Epirrita autumnata) (Lepidoptera: Geometridae). *Int. J. Mol. Sci.*, 16, 22541–22554.

Abello, P., Ai, W., Altmann, C., Bernardi, G., Bonato, O., Burchhardt, K., *et al.* (2012). Permanent Genetic Resources added to Molecular Ecology Resources Database 1 June 2010 - 31 July 2010. *Mol. Ecol. Resour.*, 12, 972–974.

Achten, W., Almeida, J. & Muys, B. (2013). Carbon footprint of science: More than flying. *Ecol. Indic.*, 34, 352–355.

Ackerly, D. (2009). Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proc. Natl. Acad. Sci.*, 106, 19699–19706.

Adamack, A. & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods Ecol. Evol.*, 5, 384–387.

Adriaensen, F., Chardon, J., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., *et al.* (2003). The application of "least-cost" modelling as a functional landscape model. *Landsc. Urban Plan.*, 64, 233–247.

Albers, C., Lunter, G., MacArthur, D., McVean, G., Ouwehand, W. & Durbin, R. (2011). Dindel: Accurate indel calls from short-read data. *Genome Res.*, 21, 961–973.

Allan, E., Weisser, W., Weigelt, A., Roscher, C., Fischer, M. & Hillebrand, H. (2011). More diverse plant communities have higher functioning over time due to turnover in complementary dominant species. *Proc. Natl. Acad. Sci.*, 108, 17034–17039.

Amsellem, L., Risterucci, A. & Benrey, B. (2003). Isolation and characterization of polymorphic microsatellite loci in Lobesia botrana Den. & Schiff. (Lepidoptera: Tortricidae). *Mol. Ecol. Notes*, 3, 117–119.

Anderson, C., Epperson, B., Fortin, M., Holderegger, R., James, P., Rosenberg, M., *et al.* (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol. Ecol.*, 19, 3565–3575.

Anderson, S., Dawson, D. & Freeland, J. (2006). Isolation and characterization of highly polymorphic microsatellite loci for the garden tiger moth Arctia caja (Lepidoptera: Arctiidae). *Mol. Ecol. Notes*, 6, 104–106.

Anthony, N., Gelembiuk, G., Raterman, D., Nice, C. & Ffrench-Constant, R. (2001). Brief report Isolation and characterization of microsatellite markers from the endangered Karner blue butterfly Lycaeides melissa samuelis (Lepidoptera). *Hereditas*, 134, 271–273.

Arias, R., Blanco, C., Portilla, M., Snodgrass, G. & Scheffler, B. (2011). First microasatellites from Spodoptera frugiperda (Lepidoptera: Noctuidae) and their potential use for population genetics. *Ann. Entomol. Soc. Am.*, 104, 576–587.

Arunkumar, K., Kifayathullah, L. & Nagaraju, J. (2009). Microsatellite markers for the Indian golden silkmoth, Antheraea assama (Saturniidae: Lepidoptera). *Mol. Ecol. Resour.*, 9, 268–270.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. & Jeffcoat, S. (2001). *The millennium atlas of butterflies in Britain and Ireland*. Oxford University Press, Oxford.

Auffret, A., Plue, J. & Cousins, S. (2015). The spatial and temporal components of functional connectivity in fragmented landscapes. *Ambio*, 44, 51–59.

August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T. & Jepson, P. (2015). Emerging technologies

for biological recording. *Biol. J. Linn. Soc.*, 115, 731–749.

Bailey, J. (2011). From genes to ecosystems: A genetic basis to ecosystem services. *Popul. Ecol.*, 53, 47–52.

Baird, N., Etter, P., Atwood, T., Currey, M., Shiver, A., Zachary, A., *et al.* (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One*, 3, e3376.

Balkenhol, N., Gugerli, F., Cushman, S., Waits, L., Coulon, A., Arntzen, J., *et al.* (2009a). Identifying future research needs in landscape genetics: Where to from here? *Landsc. Ecol.*, 24, 455–463.

Balkenhol, N., Waits, L. & Dezzani, R. (2009b). Statistical approaches in landscape genetics: An evaluation of methods for linking landscape and genetic data. *Ecography (Cop.).*, 32, 818–830.

Balloux, F. & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Mol. Ecol.*, 11, 155–165.

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

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67, 1–47.

Baum, K., Haynes, K., Dillemuth, F. & Cronin, J. (2004). The Matrix Enhances the Effectiveness of Corridors and Stepping Stones. *Ecology*, 85, 2671–2676.

Baxter, S., Hoffman, J., Tregenza, T., Wedell, N. & Hosken, D. (2017). EB Ford revisited: Assessing the long-term stability of wing-spot patterns and population genetic structure of the meadow brown butterfly on the Isles of Scilly. *Heredity (Edinb).*, 118, 322–329.

Beldade, P., Van't Hof, A., Jeronimo, M. & Long, A. (2009). Microsatellite markers associated with genes expressed in developing wings of Bicyclus anynana butterflies. *Mol. Ecol. Resour.*, 9, 1487–1492.

Bélisle, M. (2005). Measuring landscape connectivity: The challenge of behavioral landscape ecology. *Ecology*, 86, 1988–1995.

Beller, E., Spotswood, E., Robinson, A., Anderson, M., Higgs, E., Hobbs, R., *et al.* (2019). Building Ecological Resilience in Highly Modified Landscapes. *Bioscience*, 69, 80–92.

Bengston, S. (1981). Does bird predation influence the spot-number variation in Maniola jurtina (Lepidoptera)? *Biol. J. Linn. Soc.*, 15, 23–27.

Bennie, J., Hill, M., Baxter, R. & Huntley, B. (2006). Influence of slope and aspect on long-term vegetation change in British chalk grasslands. *J. Ecol.*, 94, 355–368.

Bennie, J., Huntley, B., Wiltshire, A., Hill, M. & Baxter, R. (2008). Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland. *Ecol. Modell.*, 216, 47–59.

Berkes, F., Colding, J. & Folke, C. (2003). Navigating Social-Ecological Systems. In: *Navigating Social-Ecological Systems* (eds. Berkes, F., Colding, J. & Folke, C.). Cambridge University Press, Cambridge, pp. 13–16.

Bettaibi, A., Mezghani-Khemakhem, M., Soltani, Z., Makni, H. & Makni, M. (2013). Development of polymorphic microsatellite loci for the tomato leaf miner, Tuta absoluta (Lepidoptera: Gelechiidae). *J. Genet.*, 92, e110-2.

Bezzerides, A., Bezzerides, J. & Eisner, T. (2004). Isolation of five polymorphic microsatellite markers in Utetheisa ornatrix (Lepidoptera: Arctiidae). *Mol. Ecol. Notes*, 4, 566–567.

Biggs, R., Schlüter, M., Biggs, D., Bohensky, E., BurnSilver, S., Cundill, G., *et al.* (2012). Toward Principles for Enhancing the Resilience of Ecosystem Services. *Annu. Rev. Environ. Resour.*, 37, 421–448.

Birch, L. (1948). The Intrinsic Rate of Natural Increase of an Insect Population. *J. Anim. Ecol.*, 17, 15–26.

Boettcher, P., Tixier-Boichard, M., Toro, M., Simianer, H., Eding, H., Gandini, G., *et al.* (2010). Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic

resources. *Anim. Genet.*, 41, 64–77.

Bogdanowicz, S., Mastro, V., Prasher, D. & Harrison, R. (1997). Microsatellite DNA variation among Asian and North American gypsy moths (Lepidoptera: Lymantriidae). *Ann. Entomol. Soc. Am.*, 90, 768–775.

Booy, G., Hendriks, R., Smulders, M., Van Groenendael, J. & Vosman, B. (2000). Genetic diversity and the survival of populations. *Plant Biol.*, 2, 379–395.

Botham, M., Brereton, T., Middlebrook, I., Randle, Z. & Roy, D. (2013). *United Kingdom Butterfly Monitoring Scheme report for 2012*.

Boutin-Ganache, I., Raposo, M., Raymon, M. & Deschepper, C. (2001). M13-Tailed Primers Improve the Readability and Usability of Microsatellite Analyses Performed with Two Different Allele Sizing Methods. *Biotechniques*, 31, 24–26.

ter Braak, C., van Strien, A., Meijer, R. & Verstrael, T. (1994). Analysis of Monitoring Data With Many Missing Values: Which Method? In: *Bird Numbers 1992. Distribution, monitoring and ecological aspects. Proceedings of the 12th International Conference of IBCC and EOAC* (eds. Hagemeijer, E. & Verstrael, T.). SOVON, Beek-Ubbergen, pp. 663–673.

Bradburd, G., Ralph, P. & Coop, G. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution (N. Y).*, 67, 3258–3273.

Bradshaw, A. (1965). Evolutionary Significance of Phenotypic Plasticity in Plants. In: *Advances in Genetics Volume 13* (eds. Caspari, E. & Thoday, J.). Academic Press, London, pp. 115–151.

Brakefield, P. (1979). Spot-number in Maniola jurtina- variation between generations and selection in marginal populations. *Heredity (Edinb).*, 42, 259–266.

Brakefield, P. (1982a). Ecological Studies on the Butterfly Maniola jurtina in Britain. I. Adult Behaviour, Microdistribution and Dispersal. *J. Anim. Ecol.*, 51, 713–726.

Brakefield, P. (1982b). Ecological Studies on the Butterfly Maniola jurtina in Britain . II . Population Dynamics : The Present Position. *J. Anim. Ecol.*, 51, 727–738.

Brakefield, P. (1984). The Ecological Genetics of Quantitative Characters of Maniola jurtina and Other Butterflies. In: *Symposia of the Royal Entomological Society. The Biology of Butterflies* (eds. Vane-Wright, R. & Ackery, P.). Academic Press, London, pp. 167–192.

Brakefield, P. (1987). Geographical variability in, and temperature effects on the phenology of Maniola jurtina and Pyronia tithonus (Lepidoptera, Satyrinar) in England and Wales. *Ecol. Entomol.*, 12, 139–148.

Brakefield, P. & van Noordwijkl, A. (1985). The genetics of spot pattern characters in the meadow brown butterfly Maniola jurtina (Lepidoptera : Satyrinae). *Heredity (Edinb).*, 54, 275–284.

Brakefield, P. & Shreeve, T. (1992). Case studies in evolution. In: *The Ecology of Butterflies in Britain* (ed. Dennis, R.). Oxford University Press, pp. 197–209.

Brereton, T., Cruickshanks, K., Risely, K., Noble, D. & Roy, D. (2011). Developing and launching a wider countryside butterfly survey across the United Kingdom. *J. Insect Conserv.*, 15, 279–290.

Breuker, C., Gibbs, M., Van Dongen, S., Merckx, T. & Van Dyck, H. (2010). The Use of Geometric Morphometrics in Studying Butterfly Wings in an Evolutionary Ecological Context. In: *Morphometrics for Nonmorphometricians. Lecture Notes in Earth Sciences, vol 124* (ed. Elewa, A.). Springer Berlin Heidelberg, Berlin.

Brook, B., Sodhi, N. & Bradshaw, C. (2008). Synergies among extinction drivers under global change. *Trends Ecol. Evol.*, 23, 453–560.

Brooke, M., Lees, D. & Lawman, J. (1985). Spot distribution in the meadow brown butterfly, Maniola jurtina L. (Lepidoptera: Satyridae): South Welsh populations. *Biol. J. Linn. Soc.*, 24, 337–348.

Bruford, M., Davie, N., Dulloo, M., Faith, D. & Walters, M. (2017). Monitoring Changes in Genetic Diversity. In:

The GEO Handbook on Biodiversity Observation Networks (eds. Walters, M. & Scholes, R.). Springer International Publishing AG, Cham, pp. 107–128.

Bruford, M., Ginja, C., Hoffmann, I., Joost, S., Wengel, P., Alberto, F., *et al.* (2015). Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Front. Genet.*, 6, 1–11.

Burke, I. (2010). Travel Trade-Offs for Scientists. *Science (80-.).*, 330, 1476–1478.

Butchart, S., Walpole, M., Collen, B., van Strien, A., Scharlemann, J., Almond, R., *et al.* (2010). Global Biodiversity : Indicators of Recent Declines. *Science (80-.).*, 1164, 1164–1169.

BWARS. (2020). *Bees Wasps & Ants Recording Society*. Available at: https://www.bwars.com/. Last accessed 4 March 2020.

Caldas, A., Hawthorne, D. & Barbosa, P. (2002). Isolation and characterization of microsatellite markers from Zale galbanata (Lepidoptera: Noctuidae) and amplification in other members of the genus. *Mol. Ecol. Notes*, 2, 296–297.

Callen, D., Thompson, A., Shen, Y., Phillips, H., Richards, R., Mulley, J., *et al.* (1993). Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *Am. J. Hum. Genet.*, 52, 922–927.

Cao, L., Wen, J., Wei, S., Liu, J., Yang, F. & Chen, M. (2015). Characterization of novel microsatellite markers for Hyphantria cunea and implications for other Lepidoptera. *Bull. Entomol. Res.*, 105, 273–284.

Cardinale, B., Duffy, J., Gonzalez, A., Hooper, D., Perrings, C., Venail, P., *et al.* (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67.

Cassel-Lundhagen, A., Ronnas, C. & Frauenfelder, N. (2009). Characterization and cross-amplification of 13 microsatellite loci in the northern pine processionary moth, Thaumetopoea pinivora (Lepidoptera: Notodontidae). *Mol. Ecol. Resour.*, 9, 1074–1075.

Cassel, A. (2002). Characterization of microsatellite loci in Coenonympha hero (Lepidoptera: Nymphalidae). *Mol. Ecol. Notes*, 2, 566–568.

Chambers, G. & MacAvoy, E. (2000). Microsatellites: consensus and controversy. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.*, 126, 455–476.

Chappin, F., Kofinas, G. & Folke, C. (2009). A Framework for Understanding Change. In: *Principles of ecosystem stewardship: Resilience-based natural resource management in a changing world* (eds. Chappim, F., Kofinas, G. & Folke, C.). Springer Science + Business Media, New York.

Chapuis, M. & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.*, 24, 621–631.

Chevasco, V., Elzinga, J., Viinikainen, S., Mappes, J. & Galarza, J. (2012). Unexpected genetic variability in the parthenogenetic bag worm moth Dahlica fennicella (Lepidoptera: Psychidae) revealed by novel microsatellite markers. *Conserv. Genet. Resour.*, 4, 159–162.

Clark, N., Lovell, R., Wheeler, B., Higgins, S., Depledge, M. & Norris, K. (2014). Biodiversity, cultural pathways, and human health: A framework. *Trends Ecol. Evol.*, 29, 198–204.

Clarke, B. (1970). Festschrift in Biology. *Science (80-.).*, 169, 1192.

Coates, B., Sumerford, D., Hellmich, R. & Lewis, L. (2010). A Helitron-like transposon superfamily from lepidoptera disrupts (GAAA)n microsatellites and is responsible for flanking sequence similarity within a microsatellite family. *J. Mol. Evol.*, 70, 275–288.

Conradt, L., Bodsworth, E., Roper, T. & Thomas, C. (2000). Non-random dispersal in the butterfly Maniola jurtina: implications for metapopulation models. *Proc. R. Soc. London. Ser. B- Biol. Sci.*, 267, 1505–1510.

Conradt, L., Corbet, S., Roper, T. & Bodsworth, E. (2002). Parasitism by the mite Trombidium breei on four U.K. butterfly species. *Ecol. Entomol.*, 27, 651–659.

Conradt, L. & Roper, T. (2006). Nonrandom movement behavior at habitat boundaries in two butterfly

species: Implications for dispersal. *Ecology*, 87, 125–132.

Conradt, L., Zollner, P., Roper, T., Frank, K. & Thomas, C. (2003). Foray Search: An Effective Systematic Dispersal Strategy in Fragmented Landscapes. *Am. Nat.*, 161, 905–915.

Convention on Biological Diversity. (2011a). *Aichi Biodiversity Targets*. *Strateg. Plan 2011-2020*. Available at: https://www.cbd.int/sp/targets/default.shtml. Last accessed 13 February 2017.

Convention on Biological Diversity. (2011b). *Key Elements of the Strategic Plan 2011-2020, including Aichi Biodiversity Targets*. *Strateg. Plan 2011-2020*. Available at: https://www.cbd.int/sp/elements/. Last accessed 13 February 2017.

Convention on Biological Diversity. (2020a). *Convention on Biological Diversity*. Available at: https://www.cbd.int/. Last accessed 27 February 2020.

Convention on Biological Diversity. (2020b). *Global Biodiversity Outlook 5*. Montreal.

Convention on Biological Diversity. (2020c). *Zero Draft of the Post-2020 Global Biodiversity Framework, Appendix 1*. Available at: https://www.cbd.int/doc/c/3539/9fe5/d7f2e35051986addba4ec258/wg2020-02-03-add1-en.pdf. Last accessed 16 March 2020.

Convention on Biological Diversity. (2020d). *Zero Draft of the Post-2020 Global Biodiversity Framework*. Available at: https://www.cbd.int/doc/c/efb0/1f84/a892b98d2982a829962b6371/wg2020-02-03-en.pdf?fbclid=IwAR2Ajf56fA8szfZP1FcFRIXejAH0Ae9XKuGiBc24c0lGgJIgiiSJ1wata1o. Last accessed 16 March 2020.

Corlett, R. (2017). A Bigger Toolbox: Biotechnology in Biodiversity Conservation. *Trends Biotechnol.*, 35, 55–65.

Costanza, R., D'Arge, R., Groot, R., Farber, S., Grasso, M., Hannon, B., *et al.* (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387, 253–260.

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S., Kubiszewski, I., *et al.* (2014). Changes in the global value of ecosystem services. *Glob. Environ. Chang.*, 26, 152–158.

Courtney, S., Hill, C. & Westerman, A. (1982). Pollen Carried for Long Periods by Butterflies. *Oikos*, 38, 260–263.

Cowley, M., Thomas, C., Roy, D., Wilson, R., Léon.Cortés, J., Gutierrez, D., *et al.* (2001). Density-distribution relationships in Brittish butterflies. I. The effect of mobility and spatial scale. *J. Anim. Ecol.*, 70, 410–425.

Creed, E., Dowdeswell, W., Ford, E. & McWhirter, K. (1959). Evolutionary studies on Maniola jurtina: The English Mainland, 1956-57. *Heredity (Edinb).*, 13, 363–391.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W., *et al.* (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.*, 7, 1008–1018.

Crick, H., Crosher, I., Mainstone, C., Taylor, S., Wharton, A., Langford, P., *et al.* (2020). *Nature Networks Evidence Handbook. Natural England Research Report NERR081*. York.

Crooks, K., Burdett, C., Theobald, D., King, S., Di Marco, M., Rondinini, C., *et al.* (2017). Quantification of habitat fragmentation reveals extinction risk in terrestrial mammals. *Proc. Natl. Acad. Sci.*, 114, 7635–7640.

Culley, T., Stamper, T., Stokes, R., Brzyski, J., Hardiman, N., Klooster, M., *et al.* (2013). An Efficient Technique for Primer Development and Application that Integrates Fluorescent Labeling and Multiplex PCR. *Appl. Plant Sci.*, 1, 1300027.

Dakin, E. & Avise, J. (2004). Microsatellite null alleles in parentage analysis. *Heredity (Edinb).*, 93, 504–509.

Dalecky, A., Bogdanowicz, S., Dopman, E., Bourguet, D. & Harrison, R. (2006). Two multiplex sets of eight and five microsatellite markers for the European corn borer, Ostrinia nubilalis Hubner (Lepidoptera : Crambidae). *Mol. Ecol. Notes*, 6, 945–947.

Daly, D., Waltham, K., Mulley, J., Watts, P., Rosin, A., Kemp, S., *et al.* (2004). Trinucleotide microsatellite loci for the peppered moth (Biston betularia). *Mol. Ecol. Notes*, 4, 179–181.

Dapporto, L. & Bruschini, C. (2012). Invading a refugium: Post glacial replacement of the ancestral lineage of a Nymphalid butterfly in the West Mediterranean. *Org. Divers. Evol.*, 12, 39–49.

Dapporto, L., Bruschini, C., Baracchi, D., Cini, A., Gayubo, S., GonzÁlez, J., *et al.* (2009). Phylogeography and counter-intuitive inferences in island biogeography: Evidence from morphometric markers in the mobile butterfly Maniola jurtina (Linnaeus) (Lepidoptera, Nymphalidae). *Biol. J. Linn. Soc.*, 98, 677–692.

Dapporto, L., Habel, J., Dennis, R. & Schmitt, T. (2011). The biogeography of the western Mediterranean: Elucidating contradictory distribution patterns of differentiation in Maniola jurtina (Lepidoptera: Nymphalidae). *Biol. J. Linn. Soc.*, 103, 571–577.

Dapporto, L., Vodă, R., Dincă, V. & Vila, R. (2014). Comparing population patterns for genetic and morphological markers with uneven sample sizes: An example for the butterfly Maniola jurtina. *Methods Ecol. Evol.*, 5, 834–843.

Davey, J. & Blaxter, M. (2010). RADseq: Next-generation population genetics. *Brief. Funct. Genomics*, 9, 416–423.

Davies, Z., Wilson, R., Coles, S. & Thomas, C. (2006). Changing habitat associations of a thermally constrained species, the silver-spotted skipper butterfly, in response to climate warming. *J. Anim. Ecol.*, 75, 247–256.

Delamaire, S., Esselink, G., Samiei, L., Courtin, C., Magnoux, E., Rousselet, J., *et al.* (2010). Isolation and characterization of six microsatellite loci inthelarch budmoth Zeiraphera diniana (Lepidoptera: Tortricidae). *Eur. J. Entomol.*, 107, 267–269.

Delattre, T., Baguette, M., Burel, F., Stevens, V., Quénol, H. & Vernon, P. (2013a). Interactive effects of landscape and weather on dispersal. *Oikos*, 122, 1576–1585.

Delattre, T., Burel, F., Humeau, A., Stevens, V., Vernon, P. & Baguette, M. (2010). Dispersal mood revealed by shifts from routine to direct flights in the meadow brown butterfly Maniola jurtina. *Oikos*, 119, 1900–1908.

Delattre, T., Vernon, P. & Burel, F. (2013b). An agri-environmental scheme enhances butterfly dispersal in european agricultural landscapes. *Agric. Ecosyst. Environ.*, 166, 102–109.

Delport, W., Ferguson, J. & Bloomer, P. (2005). Characterization of six microsatellite loci in the African wild silk moth (Gonometa postica, Lasiocampidae). *Mol. Ecol. Notes*, 5, 860–862.

Dennis, E., Freeman, S., Brereton, T. & Roy, D. (2013). Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. *Methods Ecol. Evol.*, 4, 637–645.

Dennis, E., Morgan, B., Brereton, T., Roy, D. & Fox, R. (2017). Using citizen science butterfly counts to predict species population trends. *Conserv. Biol.*, 31, 1350–1361.

Dennis, R. (2010). *A Resource-Based Habitat View for Conservation: Butterflies in the British Landscape*. 1st edn. Wiley-Blackwell, London.

Dennis, R. & Shreeve, T. (1996). Migration Records. In: *Butterflies on British and Irish offshore islands: Ecology and Biogeography*. Gem Publishing Company, Wallingford, pp. 43–54.

Department for Environment Food & Rural Affairs. (2010). *Press release 'Making space for nature': a review of England's wildlife sites published today*. *GOV.UK*. Available at: https://www.gov.uk/government/news/making-space-for-nature-a-review-of-englands-wildlife-sites-published-today. Last accessed 13 February 2017.

Dharmarajan, G., Beatty, W. & Rhodes, O. (2013). Heterozygote deficiencies caused by a Wahlund effect: Dispelling unfounded expectations. *J. Wildl. Manage.*, 77, 226–234.

Diacon-Bolli, J., Dalang, T., Holderegger, R. & Bürgi, M. (2012). Heterogeneity fosters biodiversity: Linking

history and ecology of dry calcareous grasslands. *Basic Appl. Ecol.*, 13, 641–653.

Diamond, S., Frame, A., Martin, R. & Buckley, B. (2011). Species' traits predict phenologcal responses to climate change in butterflies. *Ecology*, 92, 1005–1012.

Díaz, S. & Cabido, M. (1997). Plant Functional Types and Ecosystem Function in Relation to Global Change. *J. Veg. Sci.*, 8, 463–474.

Díaz, S., Fargione, J., Chapin, F. & Tilman, D. (2006). Biodiversity loss threatens human well-being. *PLoS Biol.*, 4, e277.

Díaz, S., Purvis, A., Cornelissen, J., Mace, G., Donoghue, M., Ewers, R., *et al.* (2013). Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecol. Evol.*, 3, 2958–2975.

van Dijk, E., Auger, H., Jaszczyszyn, Y. & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.*, 30, 418–426.

DiLeo, M. & Wagner, H. (2016). A Landscape Ecologist's Agenda for Landscape Genetics. *Curr. Landsc. Ecol. Reports*, 1, 115–126.

Diniz-Filho, J., Soares, T., Lima, J., Dobrovolski, R., Landeiro, V., Telles, M., *et al.* (2013). Mantel test in population genetics. *Genet. Mol. Biol.*, 36, 475–485.

Do, C., Waples, R., Peel, D., Macbeth, G., Tillett, B. & Ovenden, J. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol. Ecol. Resour.*, 14, 209–214.

Donohue, I., Petchey, O., Montoya, J., Jackson, A., Mcnally, L., Viana, M., *et al.* (2013). On the dimensionality of ecological stability. *Ecol. Lett.*, 16, 421–429.

Dowdeswell, W. (1961). Experimental studies on natural selection in the butterfly, Maniola jurtina. *Heredity (Edinb).*, 16, 39–52.

Dowdeswell, W. (1962). A further study of the butterfly Maniola jurtina in relation to natural selection by Apanteles tetricus. *Heredity (Edinb).*, 17, 513–523.

Dowdeswell, W. (1981). *The life of the meadow brown*. 1st edn. Butler & Tanner Ltd, London.

Dowdeswell, W., Fisher, R. & Ford, E. (1949). The quantitative study of populations in the Lepidoptera. *Heredity (Edinb).*, 3, 67–84.

Dowdeswell, W. & Ford, E. (1952). The distribution of spot-numbers as an index of geographical variation in the butterfly Maniola jurtina L. (Lepidoptera: Satyridae). *Heredity (Edinb).*, 6, 99–109.

Dowdeswell, W., Ford, E. & McWhirter, K. (1957). Further studies on isolation in the butterfly Maniola jurtina L. *Heredity (Edinb).*, 11, 51–65.

Dowdeswell, W., Ford, E. & McWhirter, K. (1960). Further studies on the evolution of Maniola jurtina in the Isles of Scilly. *Heredity (Edinb).*, 14, 333–364.

Dowdeswell, W. & McWhirter, K. (1967). Stability of spot-distribution in Maniola jurtina throughout its range. *Heredity (Edinb).*, 22, 187–210.

Earl, D. & VonHoldt, B. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.*, 4, 359–361.

Ekblom, R. & Wolf, J. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evol. Appl.*, 7, 1026–1042.

Elmqvist, T., Folke, C., Nyström, M., Peterson, G., Bengtsson, J., Walker, B., *et al.* (2007). Response diversity, ecosystem change, and resilience. *Front. Ecol. Environ.*, 1, 488–494.

Estoup, A., Rousset, F., Michalakis, Y., Cornuet, J., Adriamanga, M. & Guyomard, R. (1998). Comparative analysis of microsatellite and allozyme markers: A case study investigating microgeographic differentiation in brown trout (Salmo trutta). *Mol. Ecol.*, 7, 339–353.

Estrada, E. & Bodin, Ö. (2008). Using Network Centrality Measures to Manage Landscape Connectivity. *Ecol. Appl.*, 18, 1810–1825.

Etherington, T. (2011). Python based GIS tools for landscape genetics: Visualising genetic relatedness and measuring landscape connectivity. *Methods Ecol. Evol.*, 2, 52–55.

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *J. Stat. Softw.*, 76, 1–21.

Etter, P., Bassham, S., Hohenlohe, P., Johnson, E. & Cresko, W. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: *Molecular Methods for Evolutionary Genetics* (eds. Orgogozo, V. & Rockman, M.). Humana Press, New York.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.*, 14, 2611–2620.

Evans, L., Sibly, R., Thorbek, P., Sims, I., Oliver, T. & Walters, R. (2019). Quantifying the effectiveness of agri-environment schemes for a grassland butterfly using individual-based models. *Ecol. Modell.*, 411, 108798.

Evans, L., Sibly, R., Thorbek, P., Sims, I., Oliver, T. & Walters, R. (2020). The importance of including habitat-specific behaviour in models of butterfly movement. *Oecologia*, 193, 249–259.

Excoffier, L. (2001). Analysis of Population Subdivision. In: *Handbook of statistical genetics* (eds. Balding, D., Bishop, M. & Cannings, C.). JohnWiley & Sons, Ltd, Chichester, pp. 271–308.

Excoffier, L. & Lischer, H. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.*, 10, 564–567.

Eycott, A., Stewart, G., Buyung-Ali, L., Bowler, D., Watts, K. & Pullin, A. (2012). A meta-analysis on the impact of different matrix structures on species movement rates. *Landsc. Ecol.*, 27, 1263–1278.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, 34, 487–515.

Fahrig, L. (2017). Ecological Responses to Habitat Fragmentation Per Se. *Annu. Rev. Ecol. Evol. Syst.*, 48, 1–23.

Fahrig, L., Arroyo-Rodríguez, V., Bennett, J., Boucher-Lalonde, V., Cazetta, E., Currie, D., *et al.* (2019). Is habitat fragmentation bad for biodiversity? *Biol. Conserv.*, 230, 179–186.

Falush, D., Stephens, M. & Pritchard, J. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies Daniel. *Genetics*, 164, 1567–1587.

Falush, D., Stephens, M. & Pritchard, J. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes*, 7, 574–578.

Faure, N. & Silvain, J. (2005). Characterization of eight microsatellite loci in the maize stalk borer Busseola fusca Fuller, 1901 (Lepidoptera: Noctuidae). *Mol. Ecol. Notes*, 5, 846–848.

Fauvelot, C. (2005). Isolation and characterization of microsatellites in two tropical butterflies, Drupadia theda and Arhopala epimuta (Lepidoptera: Lycaenidae). *Mol. Ecol. Notes*, 5, 724–726.

Ferrari, J., Lookingbill, T. & Neel, M. (2007). Two measures of landscape-graph connectivity: Assessment across gradients in area and configuration. *Landsc. Ecol.*, 22, 1315–1323.

Fischer, J. & Lindenmayer, D. (2007). Landscape modification and habitat fragmentation: a synthesis. *Glob. Ecol. Biogeogr.*, 16, 265–280.

Fitzgibbon, S., Putland, D. & Goldizen, A. (2007). The importance of functional connectivity in the conservation of a ground-dwelling mammal in an urban Australian landscape. *Landsc. Ecol.*, 22, 1513–1525.

Flanagan, N., Blum, M., Davidson, A., Alamo, M., Albarrán, R., Faulhaber, K., *et al.* (2002). Characterization of microsatellite loci in neotropical Heliconius butterflies. *Mol. Ecol. Notes*, 2, 398–401.

Fletcher, R., Didham, R., Banks-Leite, C., Barlow, J., Ewers, R., Rosindell, J., *et al.* (2018). Is habitat

fragmentation good for biodiversity? *Biol. Conserv.*, 226, 9–15.
Fletcher, R., Maxwell, C., Andrews, J. & Helmey-Hartman, W. (2013). Signal detection theory clarifies the concept of perceptual range and its relevance to landscape connectivity. *Landsc. Ecol.*, 28, 57–67.
Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., *et al.* (2004). Regime Shifts, Resilience, and Biodiversity in Ecosystem Management. *Annu. Rev. Ecol. Evol. Syst.*, 35, 557–581.
Forman, B., Ford, E. & McWhirter, K. (1959). An evolutionary study of the butterfly Maniola jurtina in the north of Scotland. *Heredity (Edinb).*, 13, 353–361.
Fortin, M., James, P., MacKenzie, A., Melles, S. & Rayfield, B. (2012). Spatial statistics, spatial regression, and graph theory in ecology. *Spat. Stat.*, 1, 100–109.
Fox, R., Asher, J., Brereton, T., Roy, D. & Warren, M. (2006). *The State of Butterflies in Britain and Ireland*. NatureBureau, Newbury.
Fox, R., Brereton, T., Asher, J., August, T., Botham, M., Bourn, N., *et al.* (2015). *The State of the UK's Butterflies 2015*.
Fox, R., Warren, M., Brereton, T., Roy, D. & Robinson, A. (2011). A new Red List of British butterflies. *Insect Conserv. Divers.*, 4, 159–172.
Franck, P., Guérin, F., Loiseau, A. & Sauphanor, B. (2005). Isolation and characterization of microsatellite loci in the codling moth Cydia pomonella L. (Lepidoptera, Tortricidae). *Mol. Ecol. Notes*, 5, 99–102.
Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity (Edinb).*, 78, 311–327.
Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biol. Conserv.*, 143, 1919–1927.
Frankham, R., Ballou, J. & Briscoe, D. (2011). Recognition of Genetic Factors in Conservation Biology. In: *Introduction to Conservation Genetics* (eds. Frankham, R., Ballou, J. & Briscoe, D.). Cambridge University Press, Cambridge, pp. 7–8.
Freeland, J., Kirk, H. & Petersen, S. (2011). Molecular Markers in Ecology. In: *Molecular Ecology*. JohnWiley & Sons, Ltd, Chichester, pp. 31–61.
Fuller, R., Smith, G., Sanderson, J., Hill, R. & Thomson, A. (2002). The UK Land Cover Map 2000: Construction of a Parcel-Based Vector Map from Satellite Images. *Cartogr. J.*, 39, 15–25.
Funk, J., Larson, J., Ames, G., Butterfield, B., Cavender-Bares, J., Firn, J., *et al.* (2017). Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biol. Rev.*, 92, 1156–1173.
Furlan, E., Stoklosa, J., Griffiths, J., Gust, N., Ellis, R., Huggins, R., *et al.* (2012). Small population size and extremely low levels of genetic diversity in island populations of the platypus, Ornithorhynchus anatinus. *Ecol. Evol.*, 2, 844–857.
Gagnaire, P., Normandeau, E., Côté, C., Hansen, M. & Bernatchez, L. (2012). The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel. *Genetics*, 190, 725–736.
Galarza, J., Viinikainen, S., Ashrafi, R. & Mappes, J. (2011). First microsatellite panel for the Wood Tiger Moth (Parasemia plantaginis). *Conserv. Genet. Resour.*, 3, 197–199.
Galpern, P., Manseau, M. & Wilson, P. (2012). Grains of connectivity: Analysis at multiple spatial scales in landscape genetics. *Mol. Ecol.*, 21, 3996–4009.
Gardner, M., Fitch, A., Bertozzi, T. & Lowe, A. (2011). Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. *Mol. Ecol. Resour.*, 11, 1093–1101.
Garratt, M., Breeze, T., Jenner, N., Polce, C., Biesmeijer, J. & Potts, S. (2014). Avoiding a bad apple: Insect pollination enhances fruit quality and economic value. *Agric. Ecosyst. Environ.*, 184, 34–40.

Geiger, H. & Shapiro, A. (1992). Genetics, systematics and evolution of holarctic Peris napi species group populations (Lepidoptera, Pieridae), 30, 100–122.

Geijzendorffer, I., Regan, E., Pereira, H., Brotons, L., Brummitt, N., Gavish, Y., *et al.* (2016). Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective. *J. Appl. Ecol.*, 53, 1341–1350.

Geijzendorffer, I. & Roche, P. (2013). Can biodiversity monitoring schemes provide indicators for ecosystem services? *Ecol. Indic.*, 33, 148–157.

Gerber, S., Mariette, S., Streiff, R., Bodenes, C. & Kremer, A. (2000). Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Mol. Ecol.*, 9, 1037–1048.

Gibbs, M., Van Dyck, H. & Breuker, C. (2012). Development on drought-stressed host plants affects life history, flight morphology and reproductive output relative to landscape structure. *Evol. Appl.*, 5, 66–75.

Gibbs, M., Van Dyck, H. & Breuker, C. (2018). Flight-induced transgenerational maternal effects influence butterfly offspring performance during times of drought. *Oecologia*, 186, 383–391.

Glenn, T. & Schable, N. (2005). Isolating microsatellite DNA loci. *Methods Enzymol.*, 395, 202–222.

Goslee, S. & Urban, D. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *J. Stat. Softw.*, 22.

Goudet, J. (1994). FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *J. Hered.*, 86, 485–486.

Goulson, D. (1993a). Allozyme variation in the butterfly, Maniola-jurtina (Lepidoptera, Satyrinae) (L) - Evidence for selection. *Heredity (Edinb).*, 71, 386–393.

Goulson, D. (1993b). The evolutionary significance of bimodal emergence in the butterfly, Maniola jurtina (Lepidoptera: Satyrinae) (L.). *Biol. J. Linn. Soc.*, 49, 127–139.

Grace, T., Kambhampati, S. & Subramanyam, B. (2005). Characterization of polymorphic microsatellite markers in the Indian meal moth, Plodia interpunctella (Hübner). *Mol. Ecol. Notes*, 5, 321–322.

Grant, J. & Bogdanowicz, S. (2006). Isolation and characterization of microsatellite markers from the panic moth, Saucrobotys futilalis L. (Lepidoptera: Pyralidae: Pyraustinae). *Mol. Ecol. Notes*, 6, 353–355.

Gratton, P. & Sbordoni, V. (2009). Isolation of novel microsatellite markers for the clouded Apollo (P. mnemosyne Linnaeus, 1758; Lepidoptera, Papilionidae). *Conserv. Genet.*, 10, 1141–1143.

Graur, D. (1985). Gene Diversity in Hymenoptera. *Evolution (N. Y).*, 31, 190–199.

Greenbaum, G., Templeton, A., Zarmi, Y. & Bar-David, S. (2014). Allelic richness following population founding events - A stochastic modeling framework incorporating gene flow and genetic drift. *PLoS One*, 9, 1–23.

Grimm, V. & Wissel, C. (1997). Babel, or the ecological stability discussions: an inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia*, 109, 323–334.

Groenewald, C., Born, C., Chown, S. & Jansen van Vuuren, B. (2011). Development of a microsatellite library for the flightless moth Pringleophaga marioni Viette (Lepidoptera: Tineidae). *Conserv. Genet. Resour.*, 3, 291–294.

Gross, N., Robson, T., Lavorel, S., Albert, C., LeBagusse-Pinguet, Y. & Guillemin, R. (2008). Plant response trains mediate the effects of subalpine grasslands on soil moisture. *New Phytol.*, 180, 652–662.

Gruber, B. & Adamack, A. (2015). landgenreport: A new R function to simplify landscape genetic analysis using resistance surface layers. *Mol. Ecol. Resour.*, 15, 1172–1178.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., *et al.* (2011). Current trends in microsatellite genotyping. *Mol. Ecol. Resour.*, 11, 591–611.

Gunderson, L. & Allen, C. (2010). Why Resilience? Why Now? In: *Foundations of Ecological Resilience* (eds. Gunderson, L., Allen, C. & Holling, C.). Island Press, Washington DC, pp. XIII–XXV.

Habel, J., Dieker, P. & Schmitt, T. (2009). Biogeographical connections between the Maghreb and the Mediterranean peninsulas of southern Europe. *Biol. J. Linn. Soc.*, 98, 693–703.

Habel, J., Finger, A., Meyer, M., Schmitt, T. & Assmann, T. (2008). Polymorphic microsatellite loci in the endangered butterfly Lycaena helle (Lepidoptera: Lycaenidae). *Eur. J. Entomol.*, 105, 361–362.

Habel, J., Rödder, D., Lens, L. & Schmitt, T. (2013). The genetic signature of ecologically different grassland Lepidopterans. *Biodivers. Conserv.*, 22, 2401–2411.

Habel, J., Schmitt, T. & Müller, P. (2005). The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: The phylogeography of the Marbled White butterfly (Satyrinae, Lepidoptera). *J. Biogeogr.*, 32, 1489–1497.

Haeler, E., Fiedler, K. & Grill, A. (2014). What prolongs a butterfly's life?: Trade-offs between dormancy, fecundity and body size. *PLoS One*, 9, e111955.

Hajjar, R., Jarvis, D. & Gemmill-Herren, B. (2008). The utility of crop genetic diversity in maintaining ecosystem services. *Agric. Ecosyst. Environ.*, 123, 261–270.

Hale, M., Burg, T. & Steeves, T. (2012). Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. *PLoS One*, 7, e451710.

Hale, M., Lurz, P., Shirley, M., Rushton, S., Fuller, R. & Wolff, K. (2001). Impact of landscape management on the genetic structure of red squirrel populations. *Science (80-.).*, 293, 2246–2248.

Hamm, C., Aggarwal, D. & Landis, D. (2010). Evaluating the impact of non-lethal DNA sampling on two butterflies, Vanessa cardui and Satyrodes eurydice. *J. Insect Conserv.*, 14, 11–18.

Handford, P. (1973). Patterns of Variation in a Number of Genetic Systems in Maniola jurtina : The Isles of Scilly. *Proc. R. Soc. London. Ser. B- Biol. Sci.*, 183, 285–300.

Hanks, E. & Hooten, M. (2013). Circuit Theory and Model-Based Inference for Landscape Connectivity. *J. Am. Stat. Assoc.*, 108, 22–33.

Hanley, N., Breeze, T., Ellis, C. & Goulson, D. (2015). Measuring the economic value of pollination services: Principles, evidence and knowledge gaps. *Ecosyst. Serv.*, 14, 124–132.

Hanski, I. (2015). Correspondence - Habitat fragmentation and species richness. *J. Biogeogr.*, 42, 989–993.

Harper, G., Maclean, N. & Goulson, D. (2003). Microsatellite markers to assess the influence of population size, isolation and demographic change on the genetic structure of the UK butterfly Polyommatus bellargus. *Mol. Ecol.*, 12, 3349–3357.

Harper, G., Piyapattanakorn, S., Goulson, D. & Maclean, N. (2000). Isolation of microsatellite markers from the Adonis blue butterfly (Lysandra bellargus). *Mol. Ecol.*, 9, 1948–1949.

Harrison, P., Berry, P., Simpson, G., Haslett, J., Blicharska, M., Bucur, M., *et al.* (2014). Linkages between biodiversity attributes and ecosystem services: A systematic review. *Ecosyst. Serv.*, 9, 191–203.

Harvey, P. & Pagel, M. (1991). *The comparative method for studying adaptation*. *Comp. Method Evol. Biol.* Oxford University Press, Oxford.

Hayhow, D., Burns, F., Eaton, M., Al Fulaij, N., August, T., Babey, L., *et al.* (2016). *State of Nature Report*. *State Nat. Partnersh.*

Hector, A. & Bagchi, R. (2007). Biodiversity and ecosystem multifunctionality. *Nature*, 448, 188–190.

Helyar, S., Hemmer-Hansen, J., Bekkevold, D., Taylor, M., Ogden, R., Limborg, M., *et al.* (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Resour.*, 11, 123–136.

Hevia, V., Martín-López, B., Palomo, S., García-Llorente, M., de Bello, F. & González, J. (2017). Trait-based approaches to analyze links between the drivers of change and ecosystem services: Synthesizing existing evidence and future challenges. *Ecol. Evol.*, 7, 831–844.

Hijmans, R. (2019). geosphere: Spherical Trigonometry R package version 1.5-10.

Hill, W. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.*, 38, 209–216.

Hindle, B., Kerr, C., Richards, S. & Willis, S. (2015). Topographical variation reduces phenological mismatch between a butterfly and its nectar source. *J. Insect Conserv.*, 19, 227–236.

Hiron, M., Pärt, T., Siriwardena, G. & Whittingham, M. (2018). Species contributions to single biodiversity values under-estimate whole community contribution to a wider range of values to society. *Sci. Rep.*, 8, 1–7.

Hoban, S., Hauffe, H., Pérez-Espona, S., Arntzen, J., Bertorelle, G., Bryja, J., *et al.* (2013). Bringing genetic diversity to the forefront of conservation policy and management. *Conserv. Genet. Resour.*, 5, 593–598.

Hochachka, W., Fink, D., Hutchinson, R., Sheldon, D., Wong, W. & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends Ecol. Evol.*, 27, 130–137.

Hoffman, E. & Blouin, M. (2004). Historical data refute recent range contraction as cause of low genetic diversity in isolated frog populations. *Mol. Ecol.*, 13, 271–276.

Holderegger, R., Kamm, U. & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: Implications for landscape genetics. *Landsc. Ecol.*, 21, 797–807.

Holderegger, R. & Wagner, H. (2008). Landscape genetics. *Bioscience*, 58, 199–207.

Holling, C. (1961). Principles of Insect Predation. *Annu. Rev. Entomol.*, 6, 163–182.

Holling, C. (1973). Resilience and stability of ecological systems. *Annu. Rev. Ecol. Syst.*, 4, 1–23.

Holling, C. (1996). Engineering within ecological constraints. In: *Engineering within ecological constraints* (ed. Schulze, P.). National Academies Press, Washington DC, pp. 31–44.

Honnay, O., Coart, E., Butaye, J., Adriaens, D., Van Glabeke, S. & Roldán-Ruiz, I. (2006). Low impact of present and historical landscape configuration on the genetics of fragmented Anthyllis vulneraria populations. *Biol. Conserv.*, 127, 411–419.

Hubisz, M., Falush, D., Stephens, M. & Pritchard, J. (2009). Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.*, 9, 1322–1332.

Hudson, M. (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resour.*, 8, 3–17.

Hughes, A., Inouye, B., Johnson, M., Underwood, N. & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecol. Lett.*, 11, 609–623.

Hundsdoerfer, A., Sanetra, M., Corbeil, D. & Stuckas, H. (2010). Eleven hawkmoth microsatellite loci of canary island hyles tithymali (Lepidoptera). *Conserv. Genet. Resour.*, 2, 241–244.

Hutchinson, W., Van Oosterhout, C., Rogers, S. & Carvalho, G. (2003). Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (Gadus morhua). *Proc. R. Soc. London. Ser. B- Biol. Sci.*, 270, 2125–2132.

Ibrahim, K., Yassin, Y. & Elguzouli, A. (2004). Polymerase chain reaction primers for polymorphic microsatellite loci in the African armyworm, Spodoptera exempta (Lepidoptera: Noctuidae). *Mol. Ecol. Notes*, 4, 653–655.

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Bonn.

IPCC. (2018). *Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to*. Geneva.

Isaac, N. & Pocock, M. (2015). Bias and information in biological records. *Biol. J. Linn. Soc.*, 115, 522–531.

Isaac, N., van Strien, A., August, T., de Zeeuw, M. & Roy, D. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods Ecol. Evol.*, 5, 1052–1060.

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W., Reich, P., *et al.* (2011). High plant diversity is needed to maintain ecosystem services. *Nature*, 477, 199–202.

Ishiguro, N. & Tsuchida, K. (2006). Polymorphic microsatellite loci for the rice stem borer, Chilo suppressalis (Walker) (Lepidoptera: Crambidae). *Appl. Entomol. Zool.*, 41, 565–568.

Ives, A., Gross, K. & Klug, J. (1999). Stability and variability in competative communities. *Science (80-.).*, 286, 542–544.

Ji, Y., Hua, Y., Liu, Y., Liu, H., Shi, C., Li, D., *et al.* (2005a). Ten polymorphic microsatellite markers developed in the masson pine moth Dendrolimus punctatus Walker (Lepidoptera: Lasiocampidae). *Mol. Ecol. Notes*, 5, 911–913.

Ji, Y., Wu, Y. & Zhang, D. (2005b). Novel polymorphic microsatellite markers developed in the cotton bollworm Helicoverpa armigera (Lepidoptera: Noctuidae). *Insect Sci.*, 12, 331–334.

Ji, Y., Zhang, D., Hewitt, G., Kang, L. & Li, M. (2003). Polymorphic microsatellite loci for the cotton bollworm Helicoverpa armigera (Lepidoptera: Noctuidae) and some remarks on their isolation. *Mol. Ecol. Notes*, 3, 102–104.

Jiang, W., Chen, M., He, H., Wu, Y., Zhu, J. & Yu, W. (2016). Eleven microsatellite loci developed from the Chinese Hesperid butterflies, Polytremis fukia (Lepidoptera: Hesperiidae), and cross-amplification in other species of Polytremis. *Appl. Entomol. Zool.*, 1–4.

Jiang, W., Jianqing, Z., Ling, Z., Minghai, C., Chao, S. & Yu, W. (2014). Isolation and characterization of microsatellite loci in Polytremis nascens (Lepidoptera: Hesperiidae) and their cross-amplification in related species. *Appl. Entomol. Zool.*, 49, 177–181.

Jombart, T. (2008). adegenet : a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Jombart, T., Devillard, S., Dufour, A. & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity (Edinb).*, 101, 92–103.

Jost, L., Chao, A. & Chazdon, R. (2011). Compositional similarity and beta diversity. In: *Biological diversity: frontiers in measurement and assessment* (eds. Magurran, A. & McGill, B.). Oxford University Press, Oxford, pp. 66–84.

Kadoya, T. (2009). Assessing functional connectivity using empirical data. *Popul. Ecol.*, 51, 5–15.

Kalinowski, S. (2004). Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conserv. Genet.*, 5, 539–543.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I., Leadley, P., Bönisch, G., *et al.* (2011). TRY - a global database of plant traits. *Glob. Chang. Biol.*, 17, 2905–2935.

Keyghobadi, N. (2007). The genetic implications of habitat fragmentation for animals. *Can. J. Zool.*, 85, 1049–1064.

Keyghobadi, N., Crawford, L. & Maxwell, S. (2009). Successful analysis of AFLPs from non-lethally sampled wing tissues in butterflies. *Conserv. Genet.*, 10, 2021–2024.

Keyghobadi, N., Roland, J. & Strobeck, C. (1999). Influence of landscape on the population genetic structure of the alpine butterfly Parnassius smintheus (Papilionidae). *Mol. Ecol.*, 8, 1481–1495.

Keyghobadi, N., Roland, J. & Strobeck, C. (2002). Brief report: Isolation of novel microsatellite loci in the Rocky Mountain apollo butterfly , Parnassius smintheus. *Hereditas*, 136, 247–250.

Kim, K., Coates, B., Hellmich, R., Sumerford, D. & Sappington, T. (2008). Isolation and characterization of microsatellite loci from the European corn borer, Ostrinia nubilalis (Hübner) (Insecta: Lepidoptera:

Crambidae). *Perm. Genet. Resour.*, 8, 409–411.

Kim, M., Kim, H., Kwon, D. & Lee, S. (2012). Isolation and characterization of microsatellite loci from Spodoptera exigua (Lepidoptera: Noctuidae). *Appl. Entomol. Zool.*, 47, 149–152.

Kivelä, M., Arnaud-Haond, S. & Saramäki, J. (2015). EDENetworks: A user-friendly software to build and analyse networks in biogeography, ecology and population genetics. *Mol. Ecol. Resour.*, 15, 117–122.

Kleijn, D., Winfree, R., Bartomeus, I., Carvalheiro, L., Henry, M., Isaacs, R., *et al.* (2015). Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nat. Commun.*, 6, 1–8.

Klütsch, C., Misof, B. & Naumann, C. (2003). Characterization of microsatellite loci for Reissita simonyi (Rebel, 1899) (Lepidoptera, Zygaenidae). *Mol. Ecol. Notes*, 3, 528–531.

Kohler, M., Devaux, C., Grigulis, K., Leitinger, G., Lavorel, S. & Tappeiner, U. (2017). Plant functional assemblages as indicators of the resilience of grassland ecosystem service provision. *Ecol. Indic.*, 73, 118–127.

Kopelman, N., Mayzel, J., Jakobsson, M., Rosenberg, N. & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.*, 15, 1179–1191.

Koscinski, D., Crawford, L., Keller, H. & Keyghobadi, N. (2011). Effects of different methods of non-lethal tissue sampling on butterflies. *Ecol. Entomol.*, 36, 301–308.

Kreuzinger, A., Fiedler, K., Letsch, H. & Grill, A. (2015). Tracing the radiation of Maniola (Nymphalidae) butterflies: New insights from phylogeography hint at one single incompletely differentiated species complex. *Ecol. Evol.*, 5, 46–58.

Lacy, R. (1997). Importance of Genetic Variation to the Viability of Mammalian Populations. *Am. Soc. Mammal.*, 78, 320–335.

Laikre, L. (2010). Genetic diversity is overlooked in international conservation policy implementation. *Conserv. Genet.*, 11, 349–354.

Laikre, L., Hoban, S., Bruford, M., Segelbacher, G., Allendorf, F., Gajardo, G., *et al.* (2020). Post-2020 goals overlook genetic diversity. *Science (80-.).*, 367, 1083–1085.

Landguth, E., Cushman, S., Schwartz, M., McKelvey, K., Murphy, M. & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Mol. Ecol.*, 19, 4179–4191.

Larsen, T., Williams, N. & Kremen, C. (2005). Extinction order and altered community structure rapidly disrupt ecosystem functioning. *Ecol. Lett.*, 8, 538–547.

Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisting the Holy Grail. *Funct. Ecol.*, 16, 545–556.

Lavorel, S., Grigulis, K., Lamarque, P., Colace, M., Garden, D., Girel, J., *et al.* (2011). Using plant functional traits to understand the landscape distribution of multiple ecosystem services. *J. Ecol.*, 99, 135–147.

Lavorel, S., McIntyre, S., Landsberg, J. & Forbes, T. (1997). Plant functional classifications: from general groups to specific groups based on response to disturbance. *Trends Ecol. Evol.*, 12, 474–478.

Lawson, C., Bennie, J., Hodgson, J., Thomas, C. & Wilson, R. (2014). Topographic microclimates drive microhabitat associations at the range margin of a butterfly. *Ecography (Cop.).*, 37, 732–740.

Lawson Handley, L. (2015). How will the "molecular revolution" contribute to biological recording? *Biol. J. Linn. Soc.*, 115, 750–766.

Lawton, J., Brotherton, P., Brown, V., Elphick, C., Fitter, A., Forshaw, J., *et al.* (2010). *Making space for nature: A review of England's wildlife Sites and ecological network*. *Rep. to Defra*.

Lebeau, J., Wesselingh, R. & Van Dyck, H. (2016). Floral resource limitation severely reduces butterfly survival, condition and flight activity in simplified agricultural landscapes. *Oecologia*, 180, 421–427.

Lebeau, J., Wesselingh, R. & Van Dyck, H. (2017). Flower use of the butterfly Maniola jurtina in nectar-rich and nectar-poor grasslands: a nectar generalist with a strong preference? *Insect Conserv. Divers.*, 10, 258–270.

Lebeau, J., Wesselingh, R. & Van Dyck, H. (2018). Impact of floral nectar limitation on life-history traits in a grassland butterfly relative to nectar supply in different agricultural landscapes. *Agric. Ecosyst. Environ.*, 251, 99–106.

Leberg, P. (1996). Applications of Allozyme Electrophoresis in Conservation Biology. In: *Molecular Genetic Approaches in Conservation* (eds. Smith, T. & Wayne, R.). Oxford University Press, Oxford, pp. 87–103.

Leberg, P. (2002). Estimating allelic richness: Effects of sample size and bottlenecks. *Mol. Ecol.*, 11, 2445–2449.

Lebigre, C., Turlure, C. & Schtickzele, N. (2015). Characterisation of sixteen additional polymorphic microsatellite loci for the spreading but locally rare european butterfly, Brenthis ino (Lepidoptera: Nymphalidae). *Eur. J. Entomol.*, 112, 389–392.

Leclercq, S., Rivals, E. & Jarne, P. (2010). DNA slippage occurs at microsatellite loci without minimal threshold length in humans: A comparative genomic approach. *Genome Biol. Evol.*, 2, 325–335.

Lefcheck, J., Byrnes, J., Isbell, F., Gamfeldt, L., Griffin, J., Eisenhauer, N., *et al.* (2015). Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat. Commun.*, 6, 1–7.

Legagneux, P., Casajus, N., Cazelles, K., Chevallier, C., Chevrinais, M., Guéry, L., *et al.* (2018). Our House Is Burning: Discrepancy in Climate Change vs. Biodiversity Coverage in the Media as Compared to Scientific Literature. *Front. Ecol. Evol.*, 5, 1–6.

Legrand, D., Chaput-Bardy, A., Turlure, C., Dubois, Q., Huet, M., Schtickzelle, N., *et al.* (2014). Isolation and characterization of 15 microsatellite loci in the specialist butterfly Boloria eunomia. *Conserv. Genet. Resour.*, 6, 223–227.

Li, Y., Korol, A., Fahima, T. & Nevo, E. (2004). Microsatellites within genes: Structure, function, and evolution. *Mol. Biol. Evol.*, 21, 991–1007.

Lianyong, W. & Eagles, P. (2009). Some theoretical considerations: From landscape ecology to waterscape ecology. *Acta Ecol. Sin.*, 29, 176–181.

Litt, M. & Luty, J. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.*, 44, 397–401.

Liu, Y., Hou, M. & Hao, L. (2009). Twelve nuclear microsatellite loci for rice stem borer (Chilo suppressalis W.). *Mol. Ecol. Resour.*, 9, 180–182.

Liu, Y., Wu, K. & Guo, Y. (2006). Isolation and characterization of polymorphic microsatellite loci from the pink bollworm Pectinophora gossypiella (Saunders) (Lepidoptera: Gelechiidae). *Mol. Ecol. Notes*, 6, 847–849.

Loreau, M. & de Mazancourt, C. (2013). Biodiversity and ecosystem stability: A synthesis of underlying mechanisms. *Ecol. Lett.*, 16, 106–115.

Luck, G., Lavorel, S., McIntyre, S. & Lumb, K. (2012). Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *J. Anim. Ecol.*, 81, 1065–1076.

Lumley, L., Davis, C. & Sperling, F. (2009). Isolation and characterization of eight microsatellite loci in the spruce budworm species Choristoneura fumiferana and Choristoneura occidentalis, and cross-species amplification in related tortricid moths. *Conserv. Genet. Resour.*, 1, 501–504.

Maclean, I., Mosedale, J. & Bennie, J. (2019). Microclima: An r package for modelling meso- and microclimate. *Methods Ecol. Evol.*, 10, 280–290.

Manel, S. & Holderegger, R. (2013). Ten years of landscape genetics. *Trends Ecol. Evol.*, 28, 614–621.

Manel, S., Schwartz, M., Luikart, G. & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology

and population genetics. *Trends Ecol. Evol.*, 18, 189–197.
Manni, F., Guérard, E. & Heyer, E. (2004). Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Hum. Biol.*, 76, 173–190.
Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.*, 27, 209–220.
Marí-Mena, N., Valade, R., Magnoux, E., Augustin, S. & Lopez-Vaamonde, C. (2008). Microsatellites for the highly invasive Cameraria ohridella (Lepidoptera: Gracillariidae) and their amplification in related species. *Eur. J. Entomol.*, 105, 945–948.
Martin, L., Blossey, B. & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Front. Ecol. Environ.*, 10, 195–201.
Mason, S., Hill, J., Thomas, C., Powney, G., Fox, R., Brereton, T., *et al.* (2018). Population variability in species can be deduced from opportunistic citizen science records: a case study using British butterflies. *Insect Conserv. Divers.*, 11, 131–142.
Massardo, D., Roratto, P., Kronforst, M. & Moreira, G. (2012). Development of a microsatellite library for the passion flower butterfly Dione moneta Hübner (Lepidoptera: Nymphalidae: Heliconiinae). *Conserv. Genet. Resour.*, 4, 719–724.
Matisziw, T., Alam, M., Trauth, K., Inniss, E., Semlitsch, R., McIntosh, S., *et al.* (2015). A Vector Approach for Modeling Landscape Corridors and Habitat Connectivity. *Environ. Model. Assess.*, 20, 1–16.
Matthysen, E. (2005). Density-dependent dispersal in birds and mammals. *Ecography (Cop.).*, 28, 403–416.
Mayr, E. (1954). Change of genetic environment and evolution. In: *Evolution as a Process* (eds. Huxley, J., Hardy, A. & Ford, E.). Allen and Unwin, London, pp. 156–180.
McCarthy, M., Moore, J., Morris, W., Parris, K., Garrard, G., Vesk, P., *et al.* (2013). The influence of abundance on detectability. *Oikos*, 122, 717–726.
McGill, B., Dornelas, M., Gotelli, N. & Magurran, A. (2015). Fifteen forms of biodiversity trend in the anthropocene. *Trends Ecol. Evol.*, 30, 104–113.
McPhearson, M. & Moller, S. (2006). Reagents and Intrumentation. In: *PCR*. Taylor & Francis Group, Abingdin, pp. 23–63.
McRae, B. & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proc. Natl. Acad. Sci.*, 104, 19885–90.
McRae, B., Dickson, B., Keitt, T. & Shah, V. (2008). Using Circuit Theory to Model Connectivity in Ecology, Evolution and Conservation. *Ecology*, 89, 2712–2724.
McWhirter, K. (1957). A further analysis of variability in Maniola jurtina L. *Heredity (Edinb).*, 11, 359–371.
Meglécz, E., Petenian, F., Danchin, E., D'Acier, A., Rasplus, J. & Faure, E. (2004). High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: Parnassius apollo and Euphydryas aurinia. *Mol. Ecol.*, 13, 1693–1700.
Meglécz, E. & Solignac, M. (1998). Microsatellite loci for Parnassius mnemosyne (Lepidoptera). *Hereditas*, 128, 179–180.
Meirmans, P. & Hedrick, P. (2011). Assessing population structure: FST and related measures. *Mol. Ecol. Resour.*, 11, 5–18.
Mende, M., Stuckas, H. & Hundsdoerfer, A. (2011). Eight New Microsatellite Loci of the Western Palearctic Hyles euphorbiae Complex (Lepidoptera, Sphingidae). *Ann. Zool. Fennici*, 48, 142–146.
Merriam, G. (1984). Connectivity: a fundamental ecological characteristic of landscape pattern. In: *Methodology in landscape ecological research and planning: proceedings, 1st seminar, International Association of Landscape Ecology* (eds. Brandt, J. & Agger, P.). Roskilde, pp. 5–15.

Miao, X., Xu, S., Li, M., Li, M., Ng, J., Dai, F., *et al.* (2005). Simple sequence repeat-based consensus linkage map of Bombyx mori. *Proc. Natl. Acad. Sci.*, 102, 16303–16308.

Middleton-Welling, J., Wade, R., Dennis, R., Dapporto, L. & Shreeve, T. (2018). Optimising trait and source selection for explaining occurrence and abundance changes: A case study using British butterflies. *Funct. Ecol.*, 00, 1–11.

Mikheyev, A., Vo, T., Wee, B., Singer, M. & Parmesan, C. (2010). Rapid microsatellite isolation from a butterfly by de novo transcriptome sequencing: Performance and a comparison with AFLP-derived distances. *PLoS One*, 5, 3–6.

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. *Ecosystems*. Island Press, Washington DC.

Miller-Rushing, A., Høye, T., Inouye, D. & Post, E. (2010). The effects of phenological mismatches on demography. *Philos. Trans. R. Soc. B Biol. Sci.*, 365, 3177–3186.

Miller-Rushing, A., Primack, R. & Bonney, R. (2012). The history of public participation in ecological research. *Front. Ecol. Environ.*, 10, 285–290.

Mills, S., Oliver, T., Bradbury, R., Gregory, R., Brereton, T., Kühn, E., *et al.* (2017). European butterfly populations vary in sensitivity to weather across their geographical ranges. *Glob. Ecol. Biogeogr.*, 26, 1374–1385.

Minor, E. & Urban, D. (2008). A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv. Biol.*, 22, 297–307.

Mira, Ó., Martínez, J., Dawson, D., Tinaut, A. & Sánchez-Prieto, C. (2014). Twenty new microsatellite loci for population structure and parentage studies of Parnassius apollo nevadensis (Lepidoptera; Papilionidae). *J. Insect Conserv.*, 18, 771–779.

Monmonier, M. (1973). Maximum-Difference Barriers : An Alternative Numerical Regiodization Method. *Geogr. Anal.*, 5, 245–261.

Morecroft, M., Bealey, C., Beaumont, D., Benham, S., Brooks, D., Burt, T., *et al.* (2009). The UK Environmental Change Network: Emerging trends in the composition of plant and animal communities and the physical environment. *Biol. Conserv.*, 142, 2814–2832.

Mori, A., Furukawa, T. & Sasaki, T. (2013). Response diversity determines the resilience of ecosystems to environmental change. *Biol. Rev.*, 88, 349–364.

Morris, D. & Flavin, R. (1990). A digital terrain model for hydrology. In: *Proc 4th International Symposium on Spatial Data Handling*. Zurich, pp. 250–262.

Mortimer, S., Hollier, J. & Brown, V. (1998). Interactions between plant and insect diversity in the restoration of lowland calcareous grasslands in southern Britain. *Appl. Veg. Sci.*, 1, 101–114.

Mullis, K. & Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.*, 155, 335–350.

Nakahama, N., Izuno, A., Arima, K. & Isagi, Y. (2015). Development of microsatellite markers for two endangered grassland butterflies, Melitaea ambigua and M. protomedia (Nymphalidae), using Ion Torrent next-generation sequencing. *Conserv. Genet. Resour.*, 7, 525–527.

Natural England. (2019). Priority Habitat Inventory (England). Open Government Licence.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution (N. Y).*, 29, 1–10.

Newbold, T., Hudson, L., Hill, S., Contu, S., Lysenko, I., Senior, R., *et al.* (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520, 45–50.

van Noordwijk, C., Flierman, D., Remke, E., WallisDeVries, M. & Berg, M. (2012). Impact of grazing

management on hibernating caterpillars of the butterfly Melitaea cinxia in calcareous grasslands. *J. Insect Conserv.*, 16, 909–920.

Nussey, D., Coltman, D., Coulson, T., Kruuk, L., Donald, A., Morris, S., *et al.* (2005). Rapidly declining fine-scale spatial genetic structure in female red deer. *Mol. Ecol.*, 14, 3395–3405.

O'Connor, R. (2014). The Ecology and Conservation of the Adonis Blue (Polyommatus bellargus) and the Chalkhill Blue (Polyommatus coridon) Butterflies in the UK. University of Oxford.

Oliver, T., Heard, M., Isaac, N., Roy, D., Procter, D., Eigenbrod, F., *et al.* (2015). Biodiversity and Resilience of Ecosystem Functions. *Trends Ecol. Evol.*, 30, 673–684.

Oliver, T. & Morecroft, M. (2014). Interactions between climate change and land use change on biodiversity: Attribution problems, risks, and opportunities. *Wiley Interdiscip. Rev. Clim. Chang.*, 5, 317–335.

Oliver, T. & Roy, D. (2015). The pitfalls of ecological forecasting. *Biol. J. Linn. Soc.*, 115, 767–778.

Oliver, T., Roy, D., Hill, J., Brereton, T. & Thomas, C. (2010). Heterogeneous landscapes promote population stability. *Ecol. Lett.*, 13, 473–484.

Van Oosterhout, C., Hutchinson, W., Wills, D. & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*, 4, 535–538.

Orozco-Terwengel, P., Corander, J. & Schlötterer, C. (2011). Genealogical lineage sorting leads to significant, but incorrect Bayesian multilocus inference of population structure. *Mol. Ecol.*, 20, 1108–1121.

Ouin, A., Martin, M. & Burel, F. (2008). Agricultural landscape connectivity for the meadow brown butterfly (Maniola jurtina). *Agric. Ecosyst. Environ.*, 124, 193–199.

De Palma, A., Dennis, R., Brereton, T., Leather, S. & Oliver, T. (2017). Large reorganizations in butterfly communities during an extreme weather event. *Ecography (Cop.).*, 40, 577–585.

Palo, J., Varvio, S., Hanski, I. & Väinölä, R. (1995). Developing microsatellite markers for insect population structure: Complex variation in a checkerspot butterfly. *Hereditas*, 123, 295–300.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Pavinato, V., Martinelli, S., de Lima, P., Zucchi, M. & Omoto, C. (2013a). Microsatellite markers for genetic studies of the fall armyworm, Spodoptera frugiperda. *Genet. Mol. Res.*, 12, 370–380.

Pavinato, V., Silva-Brandão, K., Monteiro, M., Zucchi, M., Pinheiro, J., Dias, F., *et al.* (2013b). Development and characterization of microsatellite loci for genetic studies of the sugarcane borer, Diatraea saccharalis (Lepidoptera: Crambidae). *Genet. Mol. Res.*, 12, 1631–1635.

Pereira, H., Ferrier, S., Walters, M., Geller, G., Jongman, R., Scholes, R., *et al.* (2013). Essential Biodiversity Variables. *Science (80-.).*, 339, 277–278.

Perera, O., Blanco, C., Ballard, L., Silva-Brandão, K., Domingues, F. & Abel, C. (2011). Evaluation of Anonymous and Expressed Sequence Tag-Derived Polymorphic Microsatellite Markers in the Tobacco Budworm, Heliothis virescens (Lepidoptera: Noctuidae). *Southwest. Entomol.*, 36, 287–294.

Perera, O., Blanco, C., Scheffler, B. & Abel, C. (2007). Characteristics of 13 polymorphic microsatellite markers in the corn earworm, Helicoverpa zea (Lepidoptera: Noctuidae). *Mol. Ecol. Notes*, 7, 1132–1134.

Petenian, F., Meglécz, E., Genson, G., Rasplus, J. & Faure, E. (2005). Isolation and characterization of polymorphic microsatellites in Parnassius apollo and Euphydryas aurinia (Lepidoptera). *Mol. Ecol. Notes*, 5, 243–245.

Peterman, W. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods Ecol. Evol.*, 9, 1638–1647.

Peterman, W., Winiarski, K., Moore, C., Carvalho, C., Gilbert, A. & Spear, S. (2019). A comparison of popular approaches to optimize landscape resistance surfaces. *Landsc. Ecol.*, 34, 2197–2208.

Pickett, S. & Cadenasso, M. (1995). Landcape ecology: Spatial heterogeneity in ecological systems. *Science (80-.).*, 269, 331–334.

Pimm, S. (1984). The complexity and stability of ecosystems. *Nature*, 307, 321–326.

Pimm, S., Jenkins, C., Abell, R., Brooks, T., Gittleman, J., Joppa, L., *et al.* (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science (80-.).*, 344, 1246752.

Piry, S., Luikart, G. & Cornuet, J. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered.*, 90, 502–503.

Pocock, M., Roy, H., Preston, C. & Roy, D. (2015). The Biological Records Centre: A pioneer of citizen science. *Biol. J. Linn. Soc.*, 115, 475–493.

Pocock, M., Tweddle, J., Savage, J., Robinson, L. & Roy, H. (2017a). The diversity and evolution of ecological and environmental citizen science. *PLoS One*, 12, 1–17.

Pocock, M., Tweddle, J., Savage, J., Robinson, L. & Roy, H. (2017b). The diversity and evolution of ecological and environmental citizen science. *PLoS One*, 12, 1–17.

Pollard, E. (1977). A method for assessing changes in the abundance of butterflies. *Biol. Conserv.*, 12, 115–134.

Pollard, E., Lakhani, K. & Rothery, P. (1987). The Detection of Density-Dependence from a Series of Annual Censuses. *Ecology*, 68, 2046–2055.

Pollard, E. & Yates, T. (1993). *Monitoring Butterflies for Ecology and Conservation*. Chapman and Hall, London.

Poulsen, N., Nielsen, E., Schierup, M., Loeschcke, V. & Grønkjær, P. (2006). Long-term stability and effective population size in North Sea and Baltic Sea cod (Gadus morhua). *Mol. Ecol.*, 15, 321–331.

Powney, G., Broaders, L. & Oliver, T. (2012). Towards a measure of functional connectivity: Local synchrony matches small scale movements in a woodland edge butterfly. *Landsc. Ecol.*, 27, 1109–1120.

Powney, G. & Isaac, N. (2015). Beyond maps: A review of the applications of biological records. *Biol. J. Linn. Soc.*, 115, 532–542.

Powney, G., Preston, C., Purvis, A., Van Landuyt, W. & Roy, D. (2014). Can trait-based analyses of changes in species distribution be transferred to new geographic areas? *Glob. Ecol. Biogeogr.*, 23, 1009–1018.

Powney, G., Roy, D., Chapman, D., Brereton, T. & Oliver, T. (2011). Measuring functional connectivity using long-term monitoring data. *Methods Ecol. Evol.*, 2, 527–533.

Prasad, M., Muthulakshmi, M., Madhu, M., Archak, S., Mita, K. & Nagaraju, J. (2005). Survey and analysis of microsatellites in the silkworm, Bombyx mori: Frequency, distribution, mutations, marker potential and their conservation in heterologous species. *Genetics*, 169, 197–214.

Pritchard, J., Stephens, M. & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155, 945–959.

Proença, V., Martin, L., Pereira, H., Fernandez, M., McRae, L., Belnap, J., *et al.* (2017). Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. *Biol. Conserv.*, 213, 256–263.

Pruett, C. & Winker, K. (2008). The Effects of Sample Size on Population Genetic Diversity Estimates in Song Sparrows Melospiza melodia. *J. Avian Biol.*, 39, 252–256.

Prunier, J., Kaufmann, B., Fenet, S., Picard, D., Pompanon, F., Joly, P., *et al.* (2013). Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: Towards a better assessment of functional connectivity using an individual-based sampling scheme. *Mol. Ecol.*, 22, 5516–5530.

Pudovkin, A., Zaykin, D. & Hedgecock, D. (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics*, 144, 383–387.

Putman, A. & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for

population genetic studies. *Ecol. Evol.*, 4, 4399–4428.

Quinn, R., Gaston, K. & Roy, D. (1998). Coincidence in the distributions of butterflies and their foodplants. *Ecography (Cop.).*, 21, 279–288.

R Core Team. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna*. *R Found. Stat. Comput.* Available at: https://www.r-project.org/. Last accessed 16 February 2017.

Rambaut, A., Drummond, A., Xie, D., Baele, G. & Suchard, M. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.*, 67, 901–904.

Reddy, K., Abraham, E. & Nagaraju, J. (1999). Microsatellites in the silkworm, Bombyx mori: abundance, polymorphism, and strain characterization. *Genome*, 42, 1057–65.

Reineke, A., Assaf, H., Kulanek, D., Mori, N., Pozzebon, A. & Duso, C. (2015). A novel set of microsatellite markers for the European grapevine moth Lobesia botrana isolated using next-generation sequencing and their utility for genetic characterization of populations from Europe and the Middle East. *Bull. Entomol. Res.*, 105, 408–416.

Renkonen, O. (1938). Statisch-okologische Untersuchungen uber die terrestrische Kaferwelt der finnischen Bruchmoore. *Ann. Zool. Soc*, 6, 1–231.

Richard, M., Villemey, A., Stevens, V., Blanvillain, G., Dardenne, S. & Baguette, M. (2015). Fifteen new polymorphic microsatellite loci for the meadow brown butterfly, Maniola jurtina. *Biochem. Syst. Ecol.*, 63, 165–169.

Ricketts, T.H. (2001). The matrix matters: effective isolation in fragmented landscapes. *Am. Nat.*, 158, 87–99.

Robinson, J. & Harris, S. (1999). Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective. In: *Which DNA marker for which purpose?* (ed. Gillet, E.). Göttingen, pp. 1–22.

Rodriguez Gonzalez, J., del Barrio, G. & Duguy. (2008). Assessing functional landscape connectivity for disturbance propagation on regional scales-A cost-surface model approach applied to surface fire spread. *Ecol. Modell.*, 211, 121–141.

Roland, J., Keyghobadi, N. & Fownes, S. (2000). Alpine Parnassius Butterfly Dispersal : Effects of Landscape and Population Size. *Ecology*, 6, 1642–1653.

Rothery, P. & Roy, D. (2001). Application of generalized additive models to butterfly transect count data. *J. Appl. Stat.*, 28, 897–909.

Rousselet, J., Magnoux, E. & Kerdelhué, C. (2004). Characterization of five microsatellite loci in the Pine Processionary Moth Thaumetopoea pityocampa (Lepidoptera Notodontidae Thaumetopoeinae). *Mol. Ecol. Notes*, 4, 213–214.

Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.*, 8, 103–106.

Rowland, C., Morton, R., Carrasco, L., McShane, G., O'Neil, A. & Wood, C. (2017). Land Cover Map 2015 (vector, GB).

Roy, D., Oliver, T., Botham, M., Beckmann, B., Brereton, T., Dennis, R., *et al.* (2015). Similarities in butterfly emergence dates among populations suggest local adaptation to climate. *Glob. Chang. Biol.*, 21, 3313–3322.

Roy, D., Rothery, P., Moss, D., Pollard, E. & Thomas, J. (2001). Butterfly numbers and weather: predicting historical trends in abundance and the effects of climate change. *J. Anim. Ecol.*, 70, 201-217.

Roy, D. & Sparks, T. (2000). Phenology of British butterflies and climate change. *Glob. Chang. Biol.*, 6, 407–416.

Saarinen, E., Daniels, J. & Maruniak, J. (2009). Development and characterization of polymorphic microsatellite

loci in the endangered Miami blue butterfly (Cyclargus thomasi bethunebakeri). *Mol. Ecol. Resour.*, 9, 242–244.

Sala, O., Chapin, F., Armesto, J., Berlow, E., Bloomfield, J., Dirzo, R., *et al.* (2000). Global Biodiversity Scenarios for the Year 2100. *Science (80-.).*, 287, 1770–1774.

Sarhan, A. (2006). Isolation and characterization of five microsatellite loci in the Glanville fritillary butterfly (Melitaea cinxia). *Mol. Ecol. Notes*, 6, 163–164.

Sauné, L., Abella, F. & Kerdelhué, C. (2015). Isolation, characterization and PCR multiplexing of 17 microsatellite loci in the pine processionary moth Thaumetopoea pityocampa (Lepidoptera, Notodontidae). *Conserv. Genet. Resour.*, 7, 755–757.

Scali, V. & Masetti, M. (1973). The Population Structure of Maniola jurtina (Lepidoptera : Satyridae): The Sex-Ratio Control. *J. Anim. Ecol.*, 42, 773–778.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109, 365–371.

Schmeller, D., Henry, P., Julliard, R., Gruber, B., Clobert, J., Dziock, F., *et al.* (2009). Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conserv. Biol.*, 23, 307–316.

Schmitt, T., Giessl, A. & Seitz, A. (2003). Did Polyommatus icarus (Lepidoptera: Lycaenidae) have distinct glacial refugia in southern Europe? Evidence from population genetics. *Biol. J. Linn. Soc.*, 80, 529–538.

Schmitt, T. & Hewitt, G. (2004). The genetic pattern of population threat and loss: A case study of butterflies. *Mol. Ecol.*, 13, 21–31.

Schmitt, T., Röber, S. & Seitz, A. (2005). Is the last glaciation the only relevant event for the present genetic population structure of the meadow brown butterfly Maniola jurtina (Lepidoptera: Nymphalidae)? *Biol. J. Linn. Soc.*, 85, 419–431.

Schneider, C., Dover, J. & Fry, G. (2003). Movement of two grassland butterflies in the same habitat network: The role of adult resources and size of the study area. *Ecol. Entomol.*, 28, 219–227.

Schooley, R. & Branch, L. (2011). Habitat quality of source patches and connectivity in fragmented landscapes. *Biodivers. Conserv.*, 20, 1611–1623.

Schroeder, H., Arens, P. & Smulders, M. (2009). Autosomal and sex-linked microsatellite loci in the green oak leaf roller Tortrix viridana L. (Lepidoptera, Tortricidae). *Mol. Ecol. Resour.*, 9, 809–811.

Schumaker, N. (1996). Using Landscape Indices to Predict Habitat Connectivity. *Ecology*, 77, 1210–1225.

Schwartz, M., Luikart, G. & Waples, R. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends Ecol. Evol.*, 22, 25–33.

Scott, K., Lange, C., Scott, L. & Gahan, L. (2004). Isolation and characterization of microsatellite loci from Helicoverpa armigeraH{ü}bner (Lepidoptera: Noctuidae). *Mol. Ecol. Notes*, 4.

Scrucca, L. (2013). GA: A package for genetic algorithms in R. *J. Stat. Softw.*, 53, 1–37.

Segelbacher, G., Cushman, S., Epperson, B., Fortin, M., Francois, O., Hardy, O., *et al.* (2010). Applications of landscape genetics in conservation biology: Concepts and challenges. *Conserv. Genet.*, 11, 375–385.

Selkoe, K. & Toonen, R. (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecol. Lett.*, 9, 615–629.

Shennan, I., Lambeck, K., Flather, R., Horton, B., McArthur, J., Innes, J., *et al.* (2000). Modelling western North Sea palaeogeographies and tidal changes during the Holocene. *Geol. Soc. London, Spec. Publ.*, 166, 299–319.

Shirk, A., Wallin, D., Cushman, S., Rice, C. & Warheit, K. (2010). Inferring landscape effects on gene flow: A new model selection framework. *Mol. Ecol.*, 19, 3603–3619.

Shreeve, T. (1989). The extended flight period of Maniola jurtina (Lepidoptera: Satyridae) on chalk downland: seasonal changes of the adult phenotype and evidence for a population of mixed origins.

Entomologist, 108, 202–215.

Shreeve, T., Dennis, R. & Williams, W. (1995). Uniformity of wing spotting of Maniola jurtina (L.) in relation to environmental heterogeneity. *Nota lepid*, 18, 77–92.

Silva, C., Cordeiro, E. & Corrêa, A. (2019). Isolation and characterization of microsatellite markers for soybean looper (Lepidoptera: Noctuidae). *J. Insect Sci.*, 19.

Silvertown, J. (2009). A new dawn for citizen science Silvertown 2009. *Trends Ecol. Evol.*, 24, 467–471.

Sinama, M., Dubut, V., Costedoat, C., Gilles, A., Junker, M., Malausa, T., *et al.* (2011). Challenges of microsatellite development in Lepidoptera: Euphydryas aurinia (Nymphalidae) as a case study. *Eur. J. Entomol.*, 108, 261–266.

Singh, K., Hosken, D., Wedell, N., Ffrench-Constant, R., Bass, C., Baxter, S., *et al.* (2020). De novo genome assembly of the meadow brown butterfly, Maniola jurtina. *G3 Genes, Genomes, Genet.*, 10, 1477–1484.

Siriwardena, G., Baillie, S., Buckland, S., Fewster, R., Marchant, J. & Wilson, J. (1998). Trends in the abundance of farmland birds: A quantitative comparison of smoothed Common Birds Census indices. *J. Appl. Ecol.*, 35, 24–43.

Slatkin, M. (1985). Rare Alleles as Indicators of Gene Flow. *Evolution (N. Y).*, 39, 53–65.

Slatkin, M. (1995). A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics*, 139, 457–462.

Smith, T., Green, C. & Sherman, C. (2015). Patterns of connectivity and population structure of the southern calamary Sepioteuthis australis in southern Australia. *Mar. Freshw. Res.*, 66, 942–947.

Sork, V. & Smouse, P. (2006). Genetic analysis of landscape connectivity in tree populations. *Landsc. Ecol.*, 21, 821–836.

Spear, S., Balkenhol, N., Fortin, M., McRae, B. & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Mol. Ecol.*, 19, 3576–3591.

Stevenson-Holt, C., Watts, K., Bellamy, C., Nevin, O. & Ramsey, A. (2014). Defining landscape resistance values in least-cost connectivity models for the invasive grey squirrel: A comparison of approaches using expert-opinion and habitat suitability modelling. *PLoS One*, 9, e112119.

Storfer, A., Murphy, M., Evans, J., Goldberg, C., Robinson, S., Spear, S., *et al.* (2007). Putting the "landscape" in landscape genetics. *Heredity (Edinb).*, 98, 128–42.

Storfer, A., Murphy, M., Spear, S., Holderegger, R. & Waits, L. (2010). Landscape genetics: Where are we now? *Mol. Ecol.*, 19, 3496–3514.

Van Strien, A., Plantenga, W., Soldaat, L., Van Swaay, C. & WallisDeVries, M. (2008). Bias in phenology assessments based on first appearance data of butterflies. *Oecologia*, 156, 227–235.

Suding, K. & Goldstein, L. (2008). Testing the Holy Grail framework: using functional traits to predict ecosystem change. *New Phytol.*, 180, 559–562.

Suding, K., Lavorel, S., Chapin, F., Cornelissen, J., Díaz, S., Garnier, E., *et al.* (2008). Scaling environmental change through the community-level: A trait-based response-and-effect framework for plants. *Glob. Chang. Biol.*, 14, 1125–1140.

Suggitt, A., Gillingham, P., Hill, J., Huntley, B., Kunin, W., Roy, D., *et al.* (2011). Habitat microclimates drive fine-scale variation in extreme temperatures. *Oikos*, 120, 1–8.

Suggitt, A., Stefanescu, C., Páramo, F., Oliver, T., Anderson, B., Hill, J., *et al.* (2012). Habitat associations of species show consistent but weak responses to climate. *Biol. Lett.*, 8, 590–593.

Sutherland, C., Fuller, A. & Royle, J. (2015a). Modelling non-Euclidean movement and landscape connectivity in highly structured ecological networks. *Methods Ecol. Evol.*, 6, 169–177.

Sutherland, W., Roy, D. & Amano, T. (2015b). An agenda for the future of biological recording for ecological

monitoring and citizen science. *Biol. J. Linn. Soc.*, 115, 779–784.

Van Swaay, C., Dennis, E., Schmucki, R., Sevilleja, C., Balalaikins, M., Botham, M., *et al.* (2019). *The EU Butterfly Indicator for Grassland species: 1990-2017: Technical Report*.

Van Swaay, C., Nowicki, P., Settele, J. & Van Strien, A. (2008). Butterfly monitoring in Europe: Methods, applications and perspectives. *Biodivers. Conserv.*, 17, 3455–3469.

Van Swaay, C., Van Strien, A., Harpke, A., Fontaine, B., Stefanescu, C., Roy, D., *et al.* (2013). *The European Grassland Butterfly Indicator: 1990–2011*. *EEA Tech. Rep.*

Talloen, W., Van Dyck, H. & Lens, L. (2004). The cost of melanization: Butterfly wing coloration under environmental stress. *Evolution (N. Y).*, 58, 360–366.

Tan, S., Chen, X., Zhang, A. & Li, D. (2001). Isolation and characterization of DNA microsatellite from cotton bollworm (Helicoverpa armigera, Hübner). *Mol. Ecol. Notes*, 1, 243–244.

Tauber, A. (1970). Glaziale Reliktformen der Gattung Maniola (Lep. Satyridae) in Vorderasien. *Zeitschrift der Arbeitsgemeinschaft der österreichischen Entomol.*, 22, 101–119.

Tay, W., Behere, G., Batterham, P. & Heckel, D. (2010). Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evol. Biol.*, 10, 144.

Taylor, L. (1961). Aggregation, Variance and the Mean. *Nature*, 189, 732–735.

Taylor, P., Fahrig, L., Henein, K. & Merriam, G. (1993). Connectivity Is a Vital Element of Landscape Structure. *Oikos*, 68, 571–573.

Thomas, J. (2005). Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philos. Trans. R. Soc. B Biol. Sci.*, 360, 339–357.

Thomas, J. & Lewington, R. (2010). *The Butterflies of Britain and Ireland*. *Butterflies Britain Irel.* 2nd edn. British Wildlife Publishing.

Thomson, G. (1971). The possible existence of temporal sub-speciation in Maniola jurtina (L.) (Lep., Satyridae). *Entomol. Rec. J. Var.*, 83, 87–90.

Thomson, G. (1973). Geographical variation of Maniola jurtina. *Tijdschr. voor Entomol.*, 116, 185–227.

Thomson, G. (1987). Enzyme variation at morphological boundaries in Maniola and related genera. University of Stirling.

Tilman, D. (1999). The Ecological Consequences Of Changes In Biodiversity: A Search For General Principles. *Ecology*, 80, 1455–1474.

Tischendorf, L. & Fahrig, L. (2000). On the usage and measurement of landscape connectivity. *Oikos*, 90, 7–19.

Tittensor, D., Walpole, M., Hill, S., Boyce, D., Britten, G., Burgess, N., *et al.* (2014). A mid-term analysis of progress toward international biodiversity targets. *Science (80-.).*, 346, 241–245.

Tolman, T. & Lewington, R. (1997). *Collins Field Guide Butterflies of Britain and Europe*. Harper Collins, London.

Troll, C. (1939). Luftbildplan und ökologische Bodenforschung. Ihr zweckmäßiger Einsatz für die wissenschaftliche Erforschung und praktische Erschließung wenig bekannter Länder. *Zeitschrift der gesellschaft für Erdkd. zu Berlin*, 7, 241–298.

Turner, M. (2005). Landscape Ecology: What Is the State of the Science? *Annu. Rev. Ecol. Evol. Syst.*, 36, 319–344.

UK National Ecosystem Assesment. (2011). The UK National Ecosystem Assessment Synthesis of the Key Findings. *Unep-Wcmc*, Cambridge.

UKBMS. (2016). *UKBMS methods*. *ukbms.org*. Available at: http://www.ukbms.org/Methods.aspx. Last accessed 10 February 2017.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., *et al.* (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.*, 40, e115.

Urban, D. & Keitt, T. (2001). Landscape Connectivity : A Graph-Theoretic Perspective. *Ecology*, 82, 1205–1218.

Valdes, A., Slatkin, M. & Freimer, N. (1993). Allele frequencies at microsatellite loci: The stepwise mutation model revisited. *Genetics*, 133, 737–749.

Väli, Ü., Einarsson, A., Waits, L. & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol. Ecol.*, 17, 3808–3817.

Van't Hof, A., Brakefield, P., Saccheri, I. & Zwaan, B. (2007). Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly Bicyclus anynana, with implications for other Lepidoptera. *Heredity (Edinb).*, 98, 320–328.

Van't Hof, A., Zwaan, B., Saccheri, I., Daly, D., Bot, A. & Brakefield, P. (2005). Characterization of 28 microsatellite loci for the butterfly Bicyclus anynana. *Mol. Ecol. Notes*, 5, 169–172.

Vandewoestijne, S., Turlure, C. & Baguette, M. (2012). Development of novel microsatellite markers for a specialist species of lepidoptera, Boloria aquilonaris (Nymphalidae), based on 454 sequences. *Eur. J. Entomol.*, 109, 129–134.

Vanhala, T., Watts, K., A'Hara, S. & Cottrell, J. (2014). Population genetics of Formica aquilonia wood ants in Scotland: The effects of long-term forest fragmentation and recent reforestation. *Conserv. Genet.*, 15, 853–868.

Venkatesan, T., Sridhar, V., Tomason, Y., Jalali, S., Behere, G., Shanthi, R., *et al.* (2016). Use of expressed sequence tag microsatellite markers for population genetic research of Helicoverpa armigera (Lepidoptera: Noctuidae) from India. *Can. Entomol.*, 148, 187–100.

Vieira, M., Santini, L., Diniz, A. & Munhoz, C. (2016). Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.*, 39, 312–328.

Vila, M., Auger-Rozenberg, M., Goussard, F. & Lopez-Vaamonde, C. (2009a). Effect of non-lethal sampling on life-history traits of the protected moth Graellsia isabelae (Lepidoptera: Saturniidae). *Ecol. Entomol.*, 34, 356–362.

Vila, M., Latasa, T., Pino, J. & Verhulst, G. (2009b). Characterization of ten polymorphic microsatellite markers for the endemic Chapman's ringlet , Erebia palarica (Lepidoptera : Nymphalidae). *Eur. J. Entomol.*, 106, 485–490.

Vila, M., Marí-Mena, N., Yen, S. & Lopez-Vaamonde, C. (2010). Characterization of ten polymorphic microsatellite markers for the protected Spanish moon moth Graellsia isabelae (Lepidoptera: Saturniidae). *Conserv. Genet.*, 11, 1151–1154.

Villemey, A., Peterman, W., Richard, M., Ouin, A., van Halder, I., Stevens, V., *et al.* (2016). Butterfly dispersal in farmland: a replicated landscape genetics study on the meadow brown butterfly (Maniola jurtina). *Landsc. Ecol.*, 31, 1629–1641.

Visser, M., Holleman, L. & Gienapp, P. (2006). Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia*, 147, 164–172.

Vodă, R. (2015). Biodiversity and comparative phylogeography of western Mediterranean butterflies. Universitat Autonoma de Barcelona.

Voetdijk, B., Bakker, A. & Breeuwer, J. (2007). Microsatellite markers for testing host plant-related population differentiation in the moth genus Yponomeuta. *Mol. Ecol. Notes*, 7, 66–68.

Vogt, P., Ferrari, J., Lookingbill, T., Gardner, R., Riitters, K. & Ostapowicz, K. (2009). Mapping functional connectivity. *Ecol. Indic.*, 9, 64–71.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., *et al.* (1995). AFLP: a new technique

for DNA finterprinting. *Nucleic Acids Res.*, 23, 4407–4414.

Wahlund, S. (1928). Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas*, 11, 65–106.

Walker, B., Holling, C., Carpenter, S. & Kinzig, A. (2004). Resilience, adaptability and transformability in social-ecological systems. *Ecol. Soc.*, 9.

Wallner, W. (1987). Factors Affecting Insect Population Dynamics: Differences Between Outbreak and Non-Outbreak Species. *Annu. Rev. Entomol.*, 32, 317–340.

Walther, G., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T., *et al.* (2002). Ecological responses to recent climate change. *Nature*, 416, 389–395.

Wang, J. (2004). Application of the one-migrant-per-generation rule to conservation and management. *Conserv. Biol.*, 18, 332–343.

Wang, Y., Cao, L., Zhu, J. & Wei, S. (2016). Development and Characterization of Novel Microsatellite Markers for the Peach Fruit Moth Carposina sasakii (Lepidoptera: Carposinidae) Using Next-Generation Sequencing. *Int. J. Mol. Sci.*, 17, 362.

Waples, R. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.*, 7, 167–184.

Waples, R. & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evol. Appl.*, 3, 244–262.

Wardill, T., Scott, K., Graham, G. & Zalucki, M. (2004). Isolation and characterization of microsatellite loci from Chiasmia assimilis (Warren, 1899) (Lepidoptera: Geometridae). *Mol. Ecol. Notes*, 4, 358–360.

Watts, K., Eycott, A., Handley, P., Ray, D., Humphrey, J. & Quine, C. (2010). Targeting and evaluating biodiversity conservation action within fragmented landscapes: An approach based on generic focal species and least-cost networks. *Landsc. Ecol.*, 25, 1305–1318.

Watts, K. & Handley, P. (2010). Developing a functional connectivity indicator to detect change in fragmented landscapes. *Ecol. Indic.*, 10, 552–557.

Weir, B. (1996). *Genetic Data Analysis II*. 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Weir, B. & Cockerham, C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N. Y).*, 38, 1358.

Wiens, J. (2000). Reconstructing phylogenies from allozyme data: Comparing method performance with congruence. *Biol. J. Linn. Soc.*, 70, 613–632.

Wiens, J., Stenseth, N., Horne, B. & Ims, R. (1993). Ecological Mechanisms in Landscape Ecology. *Oikos*, 66, 369–380.

Williams, B., Brawn, J. & Paige, K. (2002). Highly polymorphic microsatellite loci for Speyeria idalia (Lepidoptera : Nymphalidae). *Mol. Ecol. Notes*, 2, 87–88.

Williams, J., Kubelik, A., Livak, K., Rafalski, J. & Tingey, S. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, 18, 6531–6535.

Wilson, G. & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177–1191.

Wilson, R., Davies, Z. & Thomas, C. (2010). Linking habitat use to range expansion rates in fragmented landscapes: A metapopulation approach. *Ecography (Cop.).*, 33, 73–82.

Winiarski, K., Peterman, W. & McGarigal, K. (2020). Evaluation of the R package 'resistancega': A promising approach towards the accurate optimization of landscape resistance surfaces. *Mol. Ecol. Resour.*, 00, 1–14.

Wood, B. & Pullin, A. (2002). Persistence of species in a fragmented urban landscape: the importance of dispersal ability and habitat availability for grassland butterflies. *Biodivers. Conserv.*, 11, 1451–1468.

Wright, S. (1943). Isolation By Distance. *Genetics*, 145, 1219–1228.

Wright, S. (1965). The interprestation of population structure by F-statistics with special regard to systems of mating. *Evolution (N. Y).*, 19, 395–420.

Wu, H., Wan, P., Huang, M. & Lei, C. (2019). Microsatellites reveal strong genetic structure in the common cutworm, Spodoptera litura. *J. Integr. Agric.*, 18, 636–643.

Wu, J. & Hobbs, R. (2002). Key issues and research priorities in landscape ecology: An idiosyncratic synthesis. *Landsc. Ecol.*, 17, 355–365.

Yachi, S. & Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proc. Natl. Acad. Sci.*, 96, 1463–1468.

Young, A., Boyle, T. & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *Trends Ecol. Evol.*, 11, 413–418.

Zakharov, E. & Hellmann, J. (2007). Characterization of 17 polymorphic microsatellite loci in the Anise swallowtail, Papilio zelicaon (Lepidoptera: Papilionidae), and their amplification in related species. *Mol. Ecol. Notes*, 7, 144–146.

Zakharov, E., Hellmann, J. & Romero-Severson, J. (2007). Microsatellite loci in the Propertius duskywing, Erynnis propertius (Lepidoptera: Hesperiidae), and related species. *Mol. Ecol. Notes*, 7, 266–268.

Zeisset, I., Als, T., Settele, J. & Boomsma, J. (2005). Microsatellite markers for the large blue butterflies Maculinea nausithous and Maculinea alcon (Lepidoptera: Lycaenidae) and their amplification in other Maculinea species. *Mol. Ecol. Notes*, 5, 165–168.

Zeller, K., McGarigal, K. & Whiteley, A. (2012). Estimating landscape resistance to movement: A review. *Landsc. Ecol.*, 27, 777–797.

Zhang, D. (2004). Lepidopteran microsatellite DNA : redundant but promising. *Trends Ecol. Evol.*, 19, 507–509.

Zhdanova, O. & Pudovkin, A. (2008). Nb_HetEx: A program to estimate the effective number of breeders. *J. Hered.*, 99, 694–695.

Zhou, Y., Gu, H. & Dorn, S. (2005). Isolation of microsatellite loci in the codling moth, Cydia pomonella (Lepidoptera: Tortricidae). *Mol. Ecol. Notes*, 5, 226–227.

Zhu, J., Yang, S., Li, Q. & Li, Y. (2011). Isolation and characterization of microsatellite loci for Rhyacionia leptotubula (Lepidoptera : Tortricidae). *African J. Microbiol. Res.*, 5, 4026–4028.

Zima, J., Lestina, D., Jansta, P., Petru, V. & Tropek, R. (2014). Isolation and characterisation of microsatellite markers for Mylothris jacksoni knutsoni (Lepidoptera: Pieridae), an endemic butterfly of the Gulf of Guinea Highlands. *Conserv. Genet. Resour.*, 6, 763–764.

Zima, J., Leština, D. & Konvička, M. (2013). Characterization of ten polymorphic microsatellite markers for an endangered butterfly Argynnis niobe and their cross-species utility in the closely related species A. adippe (lepidoptera: Nymphalidae). *Eur. J. Entomol.*, 110, 383–387.