Hybridisation and genetic structure of woodland specialist ants in fragmented habitat

Josie Monaghan

Doctor of Philosophy

University of York

Biology

October 2022

Abstract

High genetic diversity within populations allows them the potential to adapt to environmental changes. Hybridisation and gene flow between species is common in nature, has a role in a variety of evolutionary processes, and can influence genetic diversity. Mound-building red wood ants (*Formica rufa* group) are ecologically dominant, keystone species within woodlands across Eurasia. They are known to hybridise extensively where their species ranges overlap, making their evolutionary relationships highly reticulate and of great interest to biologists. As woodland specialists and poor dispersers they are particularly susceptible to the effects of habitat fragmentation, which causes isolation and reduction in gene flow between populations. This decreases the adaptive potential of a population and increases the risk of local extinction. Three red wood ant species currently occur in the British Isles: *F. rufa*, *F. aquilonia*, *F. lugubris*. European wood ant populations are well studied, in contrast there is a paucity of broad scale data for their British conspecifics. To address this substantial knowledge gap, I characterised morphological and genetic variation in populations across Britain. Introgression of mitochondrial haplotypes into *F. lugubris* morphospecies suggested gene flow between species where they co-occur. However, genomic data from 135 nests indicated a picture of more sporadic hybridisation events followed by backcrossing into parental species. British forests are highly fragmented after millennia of human activity. However, I found little evidence of habitat fragmentation affecting genetic diversity. This may reflect a resilience to such habitat change, or a lack of statistical power and future work will be able to address this question. The data presented in this thesis support a picture of largely robust species boundaries with rare hybridisation events between species pairs, and initial modelling of fragmentation effects suggests no immediate threat to British wood ants.

Contents

List of Figures

List of Tables

Acknowledgments

Well, I couldn't have got that done on my own.

First and foremost I must thank my extremely patient supervisors – Elva, Kanchon, Joan and Kevin – for all their guidance, wisdom and (sometimes slightly intimidating) academic insight. I must single out Elva for special appreciation after five years of unflinching support in the face of quite a lot of bumps in the road. I cannot state in strong enough words that I would not have completed this thesis without her. Further sincere thanks to Kanchon, especially for helping me untangle the web of timeout errors that was my RAD chapter, and for doing so with such good humour. I greatly appreciated all the technical advice and incisive wit of my external supervisors Joan and Kevin, and external advisor Jonna, it really helped keep me on track.

I am very grateful to my Thesis Advisory Panel members Julia Ferrari and Dan Jeffares for not only their invaluable comments and feedback, but also for teaching me the "so what?" test for my results! That saved me many an hour of pointless thinking time. Their honest but generous assessment of progress really helped me get back on track after difficult periods.

Also, they reminded me to go on holiday from time to time.

I'd like to sincerely thank NERC and my CASE partners at Forest Research for funding both this research and my continued existence. Also huge thanks to the NERC Environmental Omics Facility (NEOF) for their very generous funding of the RAD library preparation and sequencing.

I had a lot of help from some wonderful people with various parts of my project. Gavin

Horsburgh at NEOF is peerless in his ability to wrangle students into something approaching lab competence. His help was instrumental in getting my RAD data during my time at NEOF in Sheffield. Stuart A'Hara from Forest Research gave me invaluable advice on coaxing DNA from ants. Sandor Piross was of enormous help with the modelling work in Chapter 4.

Monica Bandeira, David Nelmes and Fiona Scaife, three of the many unsung heroes of the Biology Department, deserve far more thanks than I can give them. Thank you for putting up with my incessant emails.

I am indebted to all my office mates and lab group members for the sciencey support and, far more importantly, all the morale-boosting lunches and office hijinks when the going got tough.

Finally, all the long-suffering creatures I call my friends and family, who have all helped in big ways and small. Without my Grandparents Sue and Mike I wouldn't have had the guts to undertaken this journey in the first place, so thank you. My biggest thanks go to my wonderful Mum and my best mate Anne, for both scrabbling around Scotland in search of ants and providing me with all the moral support (and tea) that I could ask for during write up.

Declaration

This thesis has not previously been accepted for any degree and is not being concurrently submitted in candidature for any degree other than Doctor of Philosophy of the University of York. This thesis is the result of my own investigations, except where otherwise stated. All other sources are acknowledged by explicit references.

Chapter 1: General Introduction

Written by Josie Monaghan (JM) with input from Elva Robinson (EJHR)

Chapter 2: Morphological and mitochondrial data

Experimental design was by JM, EJHR and Kanchon Dasmahapatra (KKD) with input from Joan Cottrell (JC) and Jonna Kulmuni (JK). Sampling, morphological and laboratory work and data analysis were performed by JM. The Chapter was written by JM with input from EJHR and KKD.

Chapter 3: RAD sequencing

Experimental design was by JM, EJHR and KKD with input from JC and JK. Sampling, morphological and laboratory work were performed by JM. The ddRAD library preparation and sequencing were supported by the NERC Environmental Omics Facility, where the work was carried out by JM with training and guidance from Gavin Horsburgh. ddRAD library processing and analysis was carried out by JM with with input from KKD. The Chapter was written by JM with input from EJHR and KKD.

The output from Chapters 2 and 3 combined is being prepared for publication.

Chapter 4: Woodland fragmentation

Experimental design was by JM, EJHR and Kevin Watts (KW). Data collection was

carried out by JM and EJHR with input from KW. Linear modelling was carried out by JM. Chapter was written by JM with input from EJHR.

Chapter 5: General Discussion

Chapter was written by JM with input from EJHR.

Chapter 1

Introduction

1.1 General introduction

Maintenance of genetic diversity, or polymorphism, is essential for the long-term survival of a species or population. Higher diversity increases adaptive potential and reduces the risk of population extinction in response to environmental change. Levels of genetic diversity vary across taxa, time and genomes themselves (see Ellegren and Galtier, 2016). A wide variety of factors contribute to genetic diversity, including effective population size, population structure and connectivity, and gene flow within and between species. Characterisation of the genetic variation and its structure across populations allows us to reconstruct demographic and evolutionary histories, phylogenetic relationships between species, and the conservation status of species or populations. This enables us to both explore the natural world and the evolutionary processes that shape its diverse taxa, and measure our impact on it as anthropogenic change affects almost every ecosystem on earth.

Formica rufa group wood ants are an extremely well studied social insect system, however, there remain substantial gaps in our knowledge. They are keystone woodland spe-

cialists known to hybridise in parts of their species ranges in Eurasia. In this thesis, I will characterise the morphological and genetic variation amongst the wood ant populations in the UK, which are considerably less well studied than their European counterparts. I will use these data to assess the evidence for hybridisation within British populations where two species occur in sympatry. Furthermore, I will investigate whether British populations are negatively affected by the extensive woodland habitat destruction in the UK.

1.2 Hybridisation and introgression

Humans like to categorise things, and we can be resistant to anything that disturbs our peaceful categorisation. This may explain why, despite long being of interest to biologists, it took so long for hybridisation between related species to be recognised as an important force in evolution. Despite early recognition by botanists (Anderson and Stebbins, 1954), the debate as to the role of hybridisation more broadly amongst taxa still raged into the twenty first century (Seehausen, 2004). Until quite recently the prevailing view was that gene flow would limit diversification of taxa (Mayr, 1963; see Seehausen, 2004), however, and it is now known to affect at least 10% of animal and 25% of plant species (Mallet, 2005). Improvements to DNA sequencing technology and reduction in the cost of generating large sequence data sets have allowed the role of hybridisation to be explored across diverse taxa with renewed vigour. Whole genome sequencing is giving us extraordinary new insight into reticulate evolution in whole groups of eukarotes (see Mallet et al. (2016) for review).

Hybridisation has shaped the genomes of extant taxa to a surprising extent (Moran et al., 2021). This includes our own species, where between 2 and 5% of the genome in some populations is derived from ancient gene flow with our extinct relatives, the Neanderthals and Denisovans (Sankararaman et al., 2014). The findings are even more dramatic in

other taxa such as the famously reticulate *Heliconius* butterfly species complex and rapidly speciating cichlid fish, where perhaps more than 10% of the genome of some species are a consequence of ancient hybridisation (Martin et al., 2013; Meier et al., 2017).

Hybridisation plays a major evolutionary role by both creating species boundaries (e.g. polyploid speciation in plants; Soltis and Soltis, 2017) and acting against divergence (Mayr, 1963; Seehausen, 2004; Abbott et al., 2013). Perhaps less immediately intuitive is the concept of speciation with gene flow, where hybridisation boosts genetic variation and facilitates species divergence (Seehausen, 2004; Mallet, 2007; Abbott et al., 2013). This process can drive species radiations with adaptive introgression. It is important to note, however, that even in systems where adaptive introgression occurs locally, generally there is broad selection against hybrids (e.g. in *Heliconius*, Consortium, 2012; see Moran et al., 2021 for review). However, this pattern of selection against hybrids may be weaker when populations are colonising new environments (Meier et al., 2017).

Disentangling the highly reticulate nature of species relationships in groups that hybridise, and distinguishing gene flow from other causes of reticulations (e.g. incomplete lineage sorting), are non-trivial questions (Degnan and Rosenberg, 2009). Rates of gene flow vary across the genome as well as across taxa, with differences across the genome expected during both speciation with gene flow (Wu, 2001; Yang, 2010) and secondary contact after isolation (Barton and Gale, 1993). Adaptive introgression, where gene flow leads to an increase in fitness of the recipient, could result in highly localised genetic signals of hybridisation limited to the loci under selection (Martin et al., 2013). A variety of other processes (including but not limited to selection, genetic drift and recombination) interact and make patterns of divergence noisy and challenging to interpret (Martin et al., 2013).

Eusocial insects, and especially ants, hybridise extensively (Feldhaar et al., 2008; Seifert,

2019)(see below). The differences that eusocial life history traits (e.g. colony organisation, queen number) cause in gene flow compared to nonsocial organisms are beginning to be understood (Purcell et al., 2016). Characterisation of further hybridising ant systems will ultimately help shed light on this fascinating line of questioning.

1.3 Habitat fragmentation

Anthropogenic land use change and habitat fragmentation are key drivers of biodiversity loss, and thought to negatively affect most taxa on earth (Fischer and Lindenmayer, 2007). Human mediated degradation of continuous habitats into smaller, isolated remnants exposes populations or species within that fragmented landscape to increased extinction risk. The estimated extinction rate of British nonmarine taxa is 1-5% per century, mainly driven by habitat loss (Hambler et al., 2011). This rate is predicted to increase over the coming century (Hambler et al., 2011).

Habitat fragmentation can have adverse effects on populations in a variety of often interconnected ways, including: changes or disruption in the flow of food and other resources in an ecosystem; exposure to potentially harmful biotic or abiotic factors such as predators (Saunders et al., 1991); increased susceptibility to stochastic demographic or rare abiotic events; and reduced population size and dispersal ability. The decrease in both population size and connectivity between individuals in a population lowers the effective population size (N_e), and causes both increased differentiation between populations and reduced genetic variation within the fragmented populations (Frankham et al., 2002). Loss of genetic variation essentially reduces the potential for a population to respond to future environmental change. and, especially when paired with an inbreeding depression caused by fragmentation, can result in local extinction (Vanhala et al., 2014).

Forest habitats have been particularly affected by human activity. Over 50% of temperate forest cover has been fragmented or removed by humans, and European forests have

suffered the most reduction (Wade et al., 2003). The net rate of forest loss has slowed in recent decades, but an "area about the size of Libya" (178 million hectares) was lost between 1990 and 2020 (FAO, 2020). In the United Kingdom, around 13% of land is currently classified as woodland (Forestry Commission, 2022). Much of this forest cover is now highly fragmented; for example, tiny habitat patches of <2 hectares making up 75% of total patches in England (Watts, 2006). This dramatic loss of habitat connectivity is likely to have especially adverse effects on woodland specialist species who are highly adapted to a particular niches and are not resilient to rapid change.

Formica rufa group wood ants are habitat specialists with very poor dispersal, particularly in their British species ranges where at least two species of three found new nests by budding and thus can only move short distances (Stockan and Robinson, 2016). Despite expectations of woodland fragmentation having negative effects on genetic diversity, recent evidence in wood ants from Britain suggests none (see below). Further genetic data from a wider array of wood ant species, fragmentation levels, and population structures would allow us to see if this pattern applies more broadly across British wood ants. If a similar pattern of resilience is found more broadly, this may help inform us about how woodland specialists are resisting the detrimental effects of habitat fragmentation. If the opposite is found, any data would be invaluable in assessing the conservation status of keystone red wood ants across the UK.

1.4 Red wood ants

Mound-building red wood ants of the *Formica rufa* group (Hymenoptera: Formicidae) are ecologically significant woodland specialists found in temperate and boreal woodland across Eurasia (Stockan and Robinson, 2016). They build large, often clearly-visible mound structures using plant material from the surrounding area and it is perhaps this visibility that has made them so widely studied for so long. They have important roles in

forest ecosystems including nutrient cycling and energy flow, habitat modification, and seed dispersal (Frouz et al., 2016). Their main source of calorific intake is the honeydew excreted by aphid species occupying nearby trees, with wood ant colonies potentially collecting honeydew from different species of aphid across multiple trees (Rosengren and Sundström, 1991; Domisch et al., 2016). They also have an effect on local invertebrate communities, both in trees where they predate many taxa (protecting aphids in the process), and within and around their nests where a wide variety of dependent species can be found (Härkönen and Sorvari, 2014; Parmentier et al., 2014). Some wood ant species form supercolonies of interconnected nest mounds that contain millions of individuals and cover square kilometres (Robinson et al., 2016). Red wood ants are keystone species when present in forest ecosystems, and represent an excellent system with which to assess the effects of habitat change on both genetic diversity and hybridisation between species.

1.4.1 Hybridisation and phylogenetic complexity

In 1955, I.H.H. Yarrow exclaimed "[s]ince the day when Linnaeus described an ant as *Formica rufa*, the identity of this species has been an enigma and it has been the centre of a colossal nomenclatorial tangle ever since" (Yarrow, 1955). Almost seven decades and countless hours from dedicated specialists later and his phrase "chaotic state of affairs" might still have a ring of truth. Until recently the *F. rufa* group was generally accepted to include eight species: *F. rufa* Linnaeus, 1781, *F. polyctena* Foerster, 1850, *F. lugubris* Zetterstedt, 1838, *F. aquilonia* Yarrow, 1955, *F. paralugubris* Seifert, 1996, *F. pratensis* Retzius, 1793, *F. frontalis* Santschi, 1919, and *F. truncorum* Fabricus, 1804 (Goropashnaya et al., 2012; Stockan and Robinson, 2016)(Fig. 1.1). However, a recent taxonomic revision (Seifert, 2021) based on morphological character determinant analysis and genetic data defines 13 species, including the above plus *F. kupyanskayae* Bolton, 1955, *F. dusmeti* Emery, 1909, *F. sinensis* Wheeler, 1913, and two new

species *F. helvetica* SEIFERT, 2021 and *F. ussuriensis* SEIFERT, 2021.

The phylogeny shown in Figure 1.1 is based on mitochondrial sequence data (a 1441 bp region including the cytochrome *b* gene; Goropashnaya et al., 2012). The *F. rufa* group clustered tightly, and was largely consistent with a previous phylogeny of the eight species hitherto comprising the group (Goropashnaya, Fedorov and Pamilo, 2004). The topology difference in the position of the *F. truncorum*/*F. frontalis* clade was ascribed to low sequence variation in the *F. rufa* group subgenus compared with the others (Goropashnaya et al., 2012). This low level of mitochondrial sequence variation is characteristic of *F. rufa* group species, and their social behaviours and colony structures (*i.e.* highly polygynous and polydomous forms; see below) are likely to contribute to this pattern (Rosengren and Pamilo, 1983; Gyllenstrand et al., 2005; Goropashnaya et al., 2012). Problems with only using a mitochondrial (mtDNA; or indeed any single marker) data to delineate species phylogenies arise, however, when species hybridisise. Mismatch between mtDNA haplotype and the species identification based on morphometric analysis has indicated that mtDNA alone cannot be used to reliably infer species relationships in the *F. rufa* group (Seifert and Goropashnaya, 2004). This study built upon previous work that identified hybridisation among *F. rufa* group species based on morphological and genetic data, and characterisation of hybridisation in red wood ants continues to be a focus of study.

Around 19% of Central European ant species hybridise (see Seifert, 2019), almost double the estimated figure of 10% for animal taxa in general stated above (Mallet, 2005). Compare with this the 46% of *F. rufa* group species thought to hybridise (Seifert, 2021) and the highly reticulate nature of red wood ant species relationships becomes clear. Hybridisation between *F. rufa* group species occurs widely where species co-occur across Eurasia, and the wealth of morphological and molecular data demonstrating this is continues to grow.

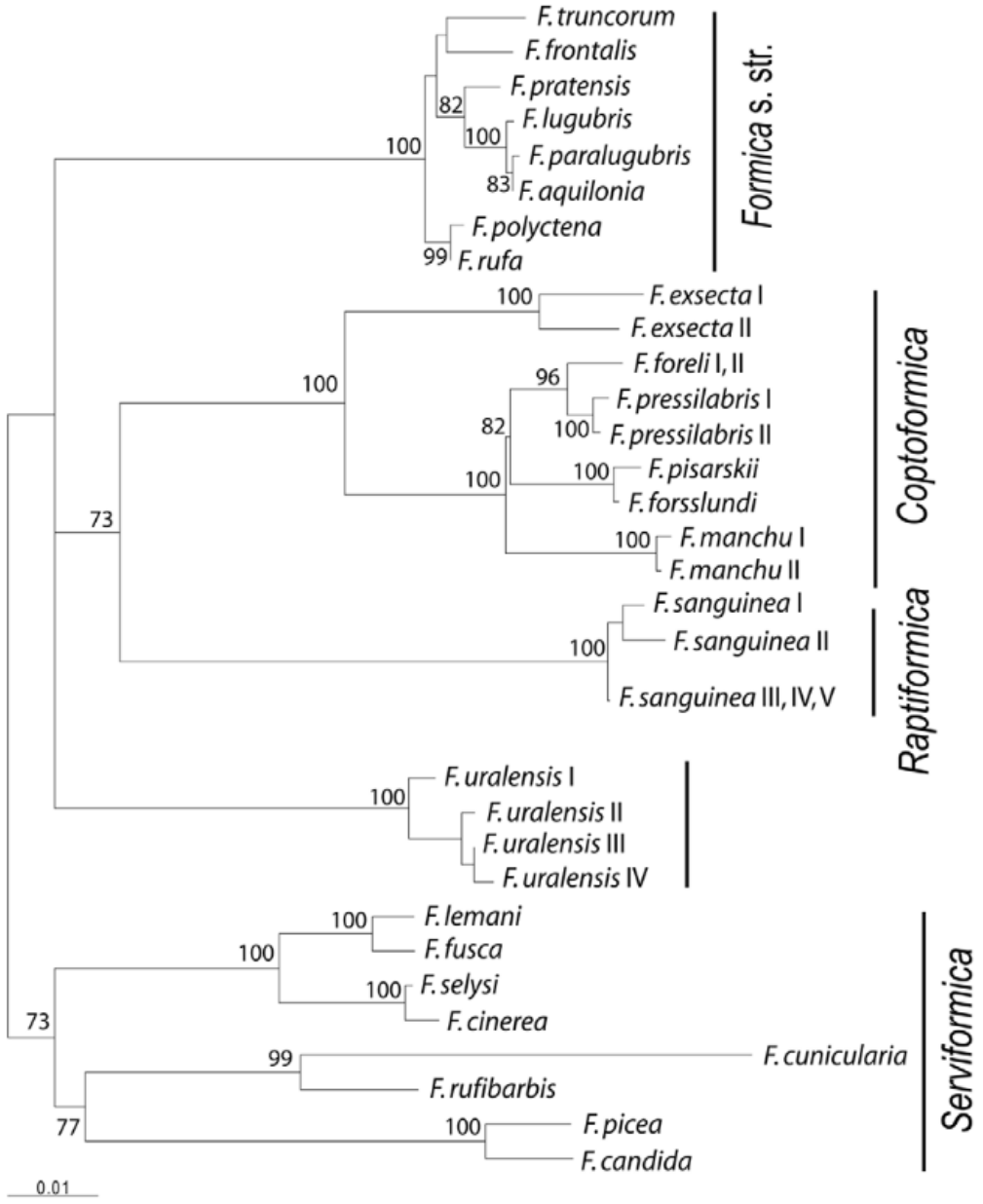

Figure 1.1: Maximum likelihood phylogeny of 32 mitochondrial haplotypes across *Formica* species taken from Goropashnaya et al. (2012). The *Formica* s. str. group at the top show the eight red wood ant species described in-text and do not include the additional five species classified as *F. rufa* group species in the taxonomic reclassification of Seifert (2021). Bootstrap values above 70 are shown on nodes.

Hybridisation in species pairs *F. aquilonia* × *F. polyctena* and *F. rufa* × *F. polyctena* is particularly common, where the frequency of hybrids over the whole range of the three parental species probably exceeds 5% for both pairings (Seifert, 2021). *Formica aquilonia* × *F. polyctena* hybrids in Finland have been well characterised using morphological, mtDNA and nuclear microsatellite markers (Kulmuni et al., 2010; Kulmuni and Pamilo, 2014; Beresford et al., 2017) and, most recently, whole genome analysis to infer demographic and speciation histories (Kulmuni et al., 2020; Portinha et al., 2022). Analysis of whole-genome data for 20 samples supported a hypothesis of species divergence with asymmetric gene flow between *F. aquilonia* and *F. polyctena*, alongside evidence of contemporary gene flow where they co-occurred (Portinha et al., 2022). *Formica rufa* × *F. polyctena* hybridisation in Central Europe (Seifert, 1991; Seifert et al., 2010) is so extensive that these species are difficult to justify taxonomically (Seifert, 2021).

Further species pairs have been found to hybridise, briefly: *F. aquilonia* × *F. lugubris* in Siberia (Goropashnaya, Fedorov and Pamilo, 2004; Bernasconi et al., 2011; Seifert, 2021); *F. pratensis* × *F. lugubris* in the Pyrenees (Seifert and Goropashnaya, 2004; Seifert, 2021), and *F. rufa* × *F. lugubris* from the Peak District in England (Seifert, 2021). In addition to the extensive hybridisation, cryptic species have been found using genetic and morphological data (Seifert, 1996; Bernasconi et al., 2011) where *F. paralugubris* is thought to have originated through hybridisation between *F. aquilonia* and *F. lugubris* (Seifert, 2021).

Even the brief outline above makes it very clear that speciation in the *F. rufa* group is complex and highly reticulate due to hybridisation and gene flow. Sex-specific introgression and hybrid fitness effects (e.g. Kulmuni et al., 2010; Kulmuni and Pamilo, 2014) introduce further layers of fascinating complexity, though unfortunately there is not space to discuss them here. Genetic and morphological characterisation of potentially hybridising populations of red wood ants in thus far unsampled extents of their

range will shed further light on the reticulate phylogeny of the *F. rufa* group.

Furthermore, it may provide new data with regard to the effects of colony structure on hybridisation in these species as, for example, the levels of polygyny (queens present in a colony) are slightly different in the British ranges for wood ants compared to their continental conspecifics. To briefly summarise (as we discuss social structures in British populations further below): polygynous *F. rufa* are found in England (Donisthorpe, 1927; Collingwood, 1979) whereas they are most commonly found to be monogynous in northern and central Europe (Punttila, 1996); this contrast is more pronounced in *F. lugubris*, which is highly polygynous in Britain but largely monogynous in other parts of its European distribution (Punttila, 1996; Breen, 1979); however, *F. aquilonia* is polygynous across its European range including Scotland. This picture is simplified (see below), but it is clear British populations of *F. rufa* and *F. lugubris* exhibit different social behaviours than conspecifics in at least part of their European distributions, evidencing the social polymorphism widely described in red wood ants.

1.4.2 Social behaviour and colony structure

Formica rufa group species vary in their social behaviours both between and within species (Stockan and Robinson, 2016). These social behaviours affect relatedness amongst workers within a nest (see Sundström et al., 2005), and the effective population size (N_e) (Pamilo and Crozier, 1997). Populations show different levels of polygyny and polydomy (multiple socially connected but spatially distinct nests exist in a colony; DEBOUT et al., 2007), and can be characterised on a spectrum between the two behavioural extremes of M-Type (monogynous and monodomous) and P-Type (highly polygynous and polydomous) colonies (Rosengren and Pamilo, 1983). The most extreme forms of P-type behaviour are supercoloniality and ultimately unicoloniality (Bourke and Franks, 1995). A number of *F. rufa* group species are super- or unicolonial including *F. paralugubris*, *F. aquilonia*, and *F. lugubris*, where this colony structure is characterised by very re-

duced worker aggression and freedom of movement among nests (e.g. Holzer et al., 2006). Unicolonial ants (unknowingly) trade off ecological dominance for the constant risk of collapse (see Helanterä et al., 2009).

The colony structure affects the dispersal behaviours of queens (Seppä and Pamilo, 1995). Gynes from monogynous colonies establish new colonies independently by flight, whereas in polygynous species or populations new nests are often founded by budding (Keller, 1991) and female reproductives may not leave the colony at all (Rosengren and Pamilo, 1983). Monogynous colony foundation in red wood ants can only be achieved when a queen parasitises the nest of a different species of ant who raise her brood, usually *F. fusca* (subgenus *Serviformica*) in Western Europe (Dekoninck et al., 2014). A number of factors influence dispersal strategies (e.g. habitat structure, resource access, and host nest presence) and both types offer different advantages, for example flight can allow longer-distance dispersal whereas propagation by budding allows for locally dominant polydomous colony formation (Seifert, 1991; Punttila, 1996, and citations therein;Seifert et al., 2010). Adopting either strategy also has its downsides, as monogynous and monodomous nest propagation may be more vulnerable to failure whilst the highly limited dispersal distance resulting from polygynous budding behaviours leave populations susceptible to habitat fragmentation (Seifert, 1991). The evolution of dispersal behaviours in ants is complex and there are many questions as yet unanswered (Hakala et al., 2019). The genetic basis for social behaviours such as these remain poorly understood. However, recent work suggested control of polygyny within *Formica* species (including the red wood ant *F. truncorum*) is contributed to by an ancient supergene (Brelsford et al., 2020).

The literature on ant social behaviours could fill a library and it is unfortunate to not have the space to discuss these topics as they deserve here. We offer a little more discussion of social structures in relation to habitat fragmentation in the following sec-

tions.

1.4.3 Effects of habitat fragmentation

As with many habitat specialists, wood ants are sensitive to local habitat characteristics and are not robust to rapid environmental changes. A variety of individual and nest level life history traits can be be affected by environmental factors. For example, *Formica polyctena* nest size increased along a latitudinal gradient, but was also affected by local scale forestry characteristics such as tree shading (Juhász et al., 2020). Human mediated habitat disturbance, such as forest felling, can impact wood ants population health in a variety of ways, including the removal of protective vegetation and destruction of food sources (Mäki-Petäys et al., 2005). The effects of clear felling on a population of *F. aquilonia* from central Finland has been extensively investigated, and was found to cause: reduced worker size, body-fat content, and nest temperature (Sorvari and Hakkarainen, 2009); decreased nest mound quality (Sorvari et al., 2016) and overwinter nest survival (Sorvari et al., 2011), and; increased between-colony aggression (Sorvari and Hakkarainen, 2004). The same body size reduction effect was not found in *F. aquilonia* queens or gynes from clear-cuts compared with the forest interior, suggesting strong selection on queen body size (Haatanen and Sorvari, 2013). However, the "quality" of queens may be reduced (e.g. decreased production of sexual offspring) in low-quality habitats (Sorvari and Hakkarainen, 2005, 2007). A similar pattern of environmental effects on worker size but not queen size was found in *F. truncorum* (Bargum et al., 2004). Studies of other ant taxa are mixed, where worker size in polyandrous ant species *Cataglyphis cursor* (Fournier et al., 2008) was not found to be heritable, but worker size was found to be genetically determined to at least some extent in *Leptothorax acervorum* (Heinze et al., 2003) and *Temnothorax curvispinosus* (Linksvayer, 2006). The heritability of queen size seems to vary over ant taxa (see Haatanen and Sorvari, 2013). Overall, this suggests that although traits such as worker and queen size are not affected by

environmental factors and habitat disturbance in all species, they are in wood ants.

As with other taxa, habitat fragmentation results in genetic differentiation between populations and a loss of variation within populations in wood ants (Mäki-Petäys et al., 2005). Haplodiploidy reduces the effective population size of social insect populations, a reduction that would be enhanced by fragmentation, leaving them susceptible to inbreeding depression (Pamilo and Crozier, 1997). A contrasting view is that haplodiploid organisms are more resistant to inbreeding depression as they have a reduced genetic load compared with diploids (see de la Filia et al. (2015) for summary). Fragmentation of woodland habitat in Peshki, Russia caused population decline and changes in spatial distribution of *F. aquilonia* and *F. lugubris* nests over three decades (Mäki-Petäys et al., 2005). Overall, the population changes were asymmetric and more pronounced in the more highly polygynous *F. aquilonia*, largely due to the social structure (Mäki-Petäys et al., 2005). These data are not in agreement with the results for a number of British studies, which suggested resilience to fragmentation effects conferred by polygynous colony structure (see below). This discordance brings into focus the importance of understanding demographic and life history traits of local populations when considering conservation action in changing landscapes.

Habitat fragmentation was found to increase the abundance of hybrid *F. polyctena* × *F. rufa* in Germany (Seifert et al., 2010). This may be a result of selective advantage for hybrids in fragmented habitats as they combine the dispersal strategies of parental species (Seifert et al., 2010).

The data briefly summarised above serve to highlight both the importance of multiple forms of data for assessing the effects of habitat fragmentation on populations, and the intriguing interplay between response to habitat loss and hybridisation between species.

1.4.4 British red wood ants

Three *F. rufa* group species occur in Britain (see Fig. 1.2): *F. rufa* in England and Wales, *F. aquilonia* in Scotland, and *F. lugubris* across all three countries (although there is a gap of more than one hundred kilometres between the Scottish and English distributions). The pattern of mitochondrial haplotypes across Eurasia is consistent with relatively recent bottlenecks followed by population expansions, probably within the Holocene after populations could emerge from glacial refugia (Goropashnaya, Fedorov and Pamilo, 2004; Goropashnaya, Fedorov, Seifert and Pamilo, 2004). It is likely these species colonised Britain from these refugia after the last glacial maximum, and were subsequently cut off from continental populations as the current coastline formed around 9000 years ago (and doggerland was lost; Walker et al., 2020). The locations of these refugia and routes via which colonisation events may have occurred are not yet known, leaving a tantalising knowledge gap.

Red wood ants can exhibit striking within-species polymorphism in social behaviours across their geographical distributions (Stockan et al., 2016). *Formica aquilonia* are associated with P-Type colony behaviours throughout Eurasia, including Scottish populations, and colonies contain many coexisting queens (are highly polygynous) and comprise large networks of discrete but connected nests (are highly polydomous) (Stockan and Robinson, 2016). In contrast with this consistency *F. lugubris*, and to a lesser extent *F. rufa*, demonstrate behavioural polymorphism across their distributions (Sundström et al., 2004). As such, Sundström et al. (2004) describe *F. aquilonia* as obligately polygynous, and both *F. lugubris* and *F. rufa* as facultatively polygynous. British populations of *F. lugubris* behave similarly to *F. aquilonia* with P-Type colony structures (Gyllenstrand and Seppä, 2003; Ellis and Robinson, 2015; Procter, 2016), which differs from some northern European populations where they are found to be monogynous and monodomous, e.g Finland (Punttila, 1996; Punttila and Kilpeläinen, 2009), Ireland (Breen,

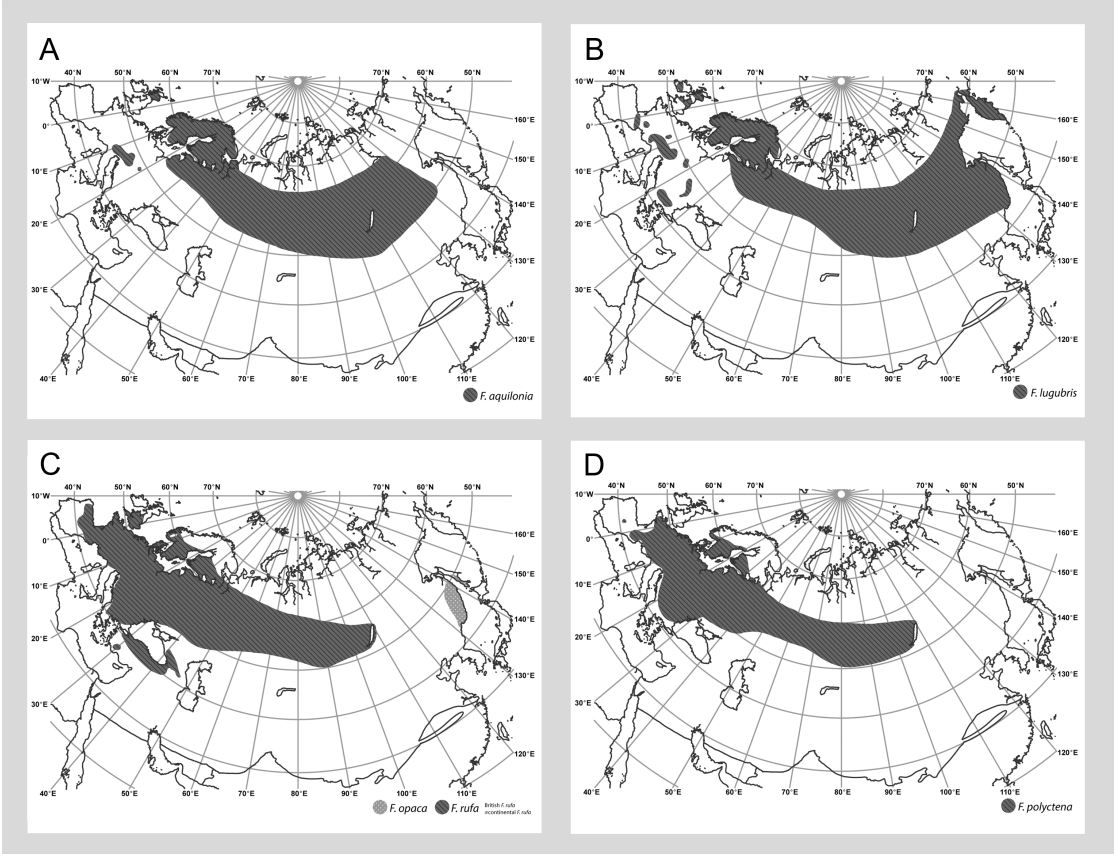

Figure 1.2: The Eurasian distributions of (A) *F. aquilonia*, (B) *F. lugubris* (C) *F. rufa*, and (D) *F. polyctena*. Maps credit to Bernhard Seifert and Phil Roberts, as published in Stockan and Robinson (2016).

1979), and the Swiss Alps where colonies range from mono- to polygynous Bernasconi et al. (2005). *Formica rufa* populations in Britain are polygynous (Donisthorpe, 1927; Collingwood, 1979; Seifert, 2021), though with many fewer queens per nest than *F. aquilonia* and *F. lugubris*, and typically monodomous with some evidence of temporary polydomy in Lancashire (N. England; Robinson and Robinson, 2008) and Kent (S. England; Welch, 1978). In contrast, *F. rufa* is M-type (monogynous and monodomous) in Finland (Punttila, 1996) and largely monogynous in Germany (Seifert, 1991; Seifert et al., 2010) and Sweden (Gyllenstrand et al., 2004). However, an isolated population in Belgium was found to comprise mainly polygynous and polydomous colonies (up to

11 nests per colony; Dekoninck et al., 2014). These differences in colony structure between British species of wood ant result in different dispersal behaviours. The levels of polygyny in *F. aquilonia* and *F. lugubris* result in propagation by budding, whilst *F. rufa* undertakes colony founding flight and parasitises *F. fusca* nests.

Forests in the British Isles are extremely fragmented after millennia of human-mediated land use change. Compared to their counterparts in continental Europe, there are quite limited data characterising the effects this may have had on red wood ant populations in the British Isles. When compared to fragmented populations in Finland, *F. lugubris* populations from highly fragmented habitat in the Peak District (northern England) showed a similar level of genetic variation between them and no evidence of an inbreeding depression (despite the English population being much more isolated; Gyllenstrand and Seppä, 2003). Similarly, populations of *F. aquilonia* from Scottish woodlands that have undergone extreme anthropogenic habitat loss and fragmentation were found to have higher genetic variation across ten microsatellite loci than European conspecifics and no evidence of inbreeding (Vanhala et al., 2014). Interestingly, there was some evidence recent afforestation reconnecting some fragments had restored avenues of gene flow between previously isolated populations (Vanhala et al., 2014). *Formica lugubris* populations across a highly fragmented region of woodland in northern England (the North York Moors; hereafter NYM) also found evidence that afforestation between ancient woodland fragments by commercial pine plantations was restoring gene flow between populations previously limited to the ancient patches (Procter, 2016). These populations were also found to be more genetically diverse than expected with no evidence of inbreeding depression caused by the fragmented landscape (Procter, 2016). In contrast with these data, very little genetic variation was found in an at-risk *F. lugubris* population in the Republic of Ireland (Mäki-Petäys and Breen, 2007). The *F. lugubris* there had a largely monogynous colony structure and was found to have undergone an inbreeding depression, unlike conspecifics and *F. aquilonia* from the UK (Mäki-Petäys

and Breen, 2007). It may be that polygyny helps maintain a sufficiently high effective population size, allowing some populations to remain resilient to the effects of forest fragmentation (Gyllenstrand and Seppä, 2003). Whole-genome sequencing of three *F. aquilonia* samples from Scotland found they had a lower mean expected heterozygosity (gene diversity) and observed heterozygosity than continental *F. aquilonia* and *F. polyctena*, though these values were calculated from sample sets from varying numbers of populations so were not all comparable (Portinha et al., 2022). Further investigation encompassing all three species and across a wider variety of woodland landscapes would help shed light on the effects habitat change has on British wood ants.

As discussed above, many *F. rufa* group species hybridise where they come into contact, even to the point of a complete breakdown of species barriers (*i.e.* between *F. rufa* and *F. polyctena*; Seifert, 2021). There is considerable overlap between the species distributions across the UK, resulting in regions where two species occur in sympatry: *F. lugubris* and *F. rufa* in Wales, and *F. lugubris* and *F. aquilonia* in Scotland. There is no reason to exclude the possibility of gene flow where the species co-occur, especially given the fact hybridisation has been found between both these species pairs previously, whether elsewhere in their ranges (*F. aquilonia* × *F. lugubris* from Siberia; Goropashnaya, Fedorov and Pamilo, 2004; Bernasconi et al., 2011; Seifert, 2021 or in British populations themselves (*F. rufa* and *F. lugubris* from the Peak District in England; Seifert, 2021. Furthermore, mitochondrial haplotype diversity was higher than expected in *F. lugubris* populations across the fragmented NYM woodland landscape in England, which might be explained by historical hybridisation as *F. rufa* also occurred in this region until recently (Procter, 2016). Indeed, *F. rufa* may still occur in remnant populations and it is not possible to know without an extensive survey. Hybridisation has also been suggested as one explanation for specimens found to be morphologically intermediate between *F. aquilonia* and *F. lugubris* during population surveys in Scotland (Macdonald, M. 2018, pers. comm.)

Discriminant analysis on a variety of morphological characters indicates the majority of British *F. rufa* nests are not "pure" *F. rufa* as compared to samples from Europe, and are thought to be descended from *F. rufa* × *F. polyctena* hybrids from the continent (Seifert et al., 2010; Seifert, 2021)(see *F. polyctena* species distribution in Fig. 1.2 D). *Formica rufa* populations in England and Wales are also more polygynous than many populations of their European conspecifics, which provides more support for the hybrid origin hypothesis as *F. polyctena* are polygynous (Seifert et al., 2010; Stockan and Robinson, 2016). *F. polyctena* is not currently present in Britain (Collingwood, 1979; Seifert, 2021), and there are no records suggesting its presence historically. However, this is not enough to completely preclude the possibility there were "pure" *F. polyctena* present before it's description in 1850 but they have since hybridised so much they cannot be distinguished. Comparison of genetic data from British and continental *F. rufa* alongside continental *F. polyctena* may help shed light on the possible hybrid origin of British *F. rufa*.

In summary, the British red wood ants comprise two species pairs that may be hybridising (or have done so historically) and one species that may be of hybrid origin. This presents an interesting avenue of investigation to be explored with novel data. Further genetic and morphological data could help to illuminate the potentially complex species relationships in British *F. rufa* group species. Furthermore, their keystone status and poor dispersal capabilities makes these species attractive study organisms for the impact of habitat fragmentation on genetic diversity and its conservation implications. Finally, characterisation of these species on a broader scale than previously possible will allow us to more fully compare them with their European counterparts.

1.5 Thesis outline

This thesis will comprise of a series of studies characterising variation in British *F. rufa* group wood ants. In Chapter 2, I will present a comparison of morphological species identification and mitochondrial sequence data to look for signals of discordance that may suggest underlying population structure and/or gene flow between species. This will allow us to put the interesting results previously found in the North York Moors (Procter, 2016) into a multispecies and broader geographical context. In Chapter 3, I will further investigate genetic diversity in these populations using reduced representation genome sequencing in the form of restriction associated DNA sequencing (RAD) libraries. Comparison of these data to our previous findings and those of other wood ant study systems will allow even more in depth assessment of population structure and demography, and add a nuclear sequence dataset to any inferences of gene flow. In Chapter 4, I will investigate whether forest fragmentation can predict the genetic diversity found in the RAD dataset from Chapter 3 and assess whether British *F. rufa* group species are negatively impacted by anthropogenic habitat loss.

Chapter 2

Inferring hybridisation from mitochondrial and morphological mismatch in British wood ants

2.1 Abstract

Mound-building red wood ant (*Formica rufa* group) species play a significant role in forest ecosystems across the Palearctic. Morphological and genetic data have shown extensive hybridisation between many *F. rufa* group species across Europe. Three red wood ant species currently occur in the British Isles with varying distributions: *F. rufa* in southern England and Wales, *F. aquilonia* in Scotland, *F. lugubris* in northern England, Wales and Scotland. Previous local-scale data from English populations of *F. lugubris* have hinted gene flow between species. To explore whether there are signals of hybridisation between British red wood ants we sampled 110 nests across the three species, including areas where pairs of species overlap in their range. Individuals were identified to morphospecies based on commonly used morphological characters,

and a 793bp fragment of the COI-II mitochondrial gene was sequenced. We show that there is considerable mismatch between the morphological species ID and mitochondrial (mtDNA) haplogroup of nests. Introgression of *F. rufa* and *F. aquilonia* mtDNA haplotypes into morphologically identified *F. lugubris* was found in a number of woodland sites where species pairs come into contact. This suggests gene flow between species at various locations where species occur in sympatry, either historically or currently. We also suggest neither the morphological identification measures nor the mitochondrial sequencing we used are sufficient to correctly infer nest species in Scottish red wood ants if used alone.

2.2 Introduction

Hybridisation between species is common in nature. Gene flow and subsequent introgression of genetic material causes reticulated patterns in species phylogenies and conflict between data from different genetic and phenotypic markers.

Mound-building red wood ants (*Formica rufa* group) are woodland specialists with an important role in forest ecosystems including nutrient cycling, habitat modification, and seed dispersal (Stockan and Robinson, 2016; see Introduction for details on the *F. rufa* group). The species group is dispersed widely across the Palearctic and characterised by morphological variability, diverse and polymorphic social behaviours, and widespread hybridisation between species, all of which makes defining robust species challenging (Goropashnaya et al., 2012; Stockan et al., 2016; Seifert, 2021). Furthermore, genetic, morphological and behavioural variation may be higher within species than between them, and these differences are not consistent across the geographic range of a given species (Seifert, 2021).

Hybridisation between *F. rufa* group species occurs widely in Eurasia, and Seifert (2021) suggests introgression in up to six of the 13 species recognised in that study. Hybridis-

ation between a various species pairs has been inferred using morphological characters alone or in concert with and mitochondrial or nuclear microsatellite sequence data e.g. *F. aquilonia* × *F. polyctena* hybrids occur in Finland (Kulmuni et al., 2010; Kulmuni and Pamilo, 2014; Beresford et al., 2017), *F. rufa* × *F. polyctena* in Central Europe (Seifert, 1991; Seifert et al., 2010), *F. aquilonia* × *F. lugubris* in Siberia (Goropashnaya, Fedorov and Pamilo, 2004; Bernasconi et al., 2011; Seifert, 2021), *F. pratensis* × *F. lugubris* in the Pyrenees (Seifert and Goropashnaya, 2004; Seifert, 2021). Most relevant to the following study is ancient hybridisation between *F. rufa* and *F. lugubris* inferred from a *F. lugubris* sample from the Peak District in England (Seifert, 2021). Furthermore, cryptic species have been found using genetic and morphological data (Seifert, 1996; Bernasconi et al., 2011) further highlighting the taxonomic uncertainty within the species complex. Indeed, *F. paralugubris* is thought to have originated through hybridisation between *F. aquilonia* and *F. lugubris* (Seifert, 2021).

Three red wood ant species currently inhabit the British Isles as the north-western limit of their range: *F. rufa*, *F. lugubris*, and *F. aquilonia*. They have different but partially overlapping distributions in the British Isles, including some regions with sympatry between pairs of species *i.e.* *F. rufa* and *F. lugubris* in Wales and limited regions in northern England, and *F. lugubris* and *F. aquilonia* in Scotland. Limited genetic data are currently available for populations of red wood ants on the British Isles. A study comparing fragmented, isolated *F. lugubris* populations in the Peak District (Northern England) with a fragmented but non-isolated Finnish population found similar genetic variation between the two and no evidence of an inbreeding depression in the isolated British population (Gyllenstrand and Seppä, 2003). This conflicts with data from a similar population in the Republic of Ireland, which found very little genetic variation, a largely monogynous colony structure, and evidence of an inbreeding depression in an at-risk *F. lugubris* population (Mäki-Petäys and Breen, 2007)). This highlights the effect that social behaviour can have on genetic variation and potential future viability

of populations. Scottish *F. aquilonia* populations are genetically diverse and show no evidence of inbreeding, despite extensive habitat fragmentation over time (Vanhala et al., 2014). The conservation focus of these studies mean they are restricted to specific populations, and the lack of genetic data comparing all three species or on a wider geographical scale represents a substantial knowledge gap. Further work is required to put these findings into context with other British and European red wood ants, and both untangle the history of these populations and look for evidence of current hybridisation where species are sympatric.

More divergence was found between mitochondrial haplotypes within the long-established populations of *F. lugubris* inhabiting a highly fragmented region of woodland in the North York Moors National Park (hereafter "NYM") than between NYM *F. lugubris* and Scottish *F. aquilonia* (Procter, 2016). This level of genetic variation was higher than anticipated (Procter, 2016) and aligned with the findings of Gyllenstrand and Seppä (2003) outlined above, suggesting fragmentation and interrupted gene flow may not be as detrimental to genetic variation as previously thought, and polygyny maintains a sufficiently high effective population size (Gyllenstrand and Seppä, 2003) and/or gene flow has been re-established between previously isolated populations. *F. rufa* remnant populations are known to have persisted in some regions of the NYM until at least 2013 (Procter, 2016), though no *F. rufa* nests have been detected since then (E. J. H. Robinson, pers. comm. 2022). The presence of *F. rufa* - historic or otherwise - in the NYM means the high genetic variation found there may be explained by historic gene flow between species (Procter, 2016). Another hypothesis is that there might be another undescribed *F. rufa* group cryptic species present in England (Procter, 2016). Comparison of data from these unusually genetically diverse populations to others across the British Isles and between all three species may help shed light on red wood ant population history, in the NYM and beyond.

As these unexpected levels of variation were found in mitochondrial haplotype data (Procter, 2016) from certain British wood ant populations, we focus here on mitochondrial sequencing (for nuclear data, see Chapter 3). The mitochondrial genome is inherited matrilineally and does not recombine, so it cannot be used to infer gene flow between species if the sole source of data. However, a mismatch between mitochondrial haplotypes and another form of species identification could allow hybridisation to be detected. To look for signals of hybridisation between British red wood ant species we will generate both a mitochondrial sequence dataset and characterise species based on morphological data, which is commonly used to distinguish ant species (Stockan and Robinson, 2016; Seifert, 2021). A mismatch rate of 11-14% between mtDNA haplotype and morphological phenotype was found in five *F. rufa* group species across their Palearctic ranges (*F. rufa*, *F. polyctena*, *F. pratensis*, *F. lugubris*, and *F. aquilonia*; Seifert and Goropashnaya (2004)), confirming the efficacy of such an approach.

The distribution of the three red wood ant populations across the British Isles provide an excellent system in which to look for signs of gene flow between species because all three species occur both singly and in sympatry with a second species in different parts of their ranges. Furthermore, *F. lugubris* overlaps with both the other species, which may result in signals of varying gene flow across its British range if indeed hybridisation occurs. Our objectives in generating new morphological and mitochondrial sequence data sets for British red wood ants are two-fold: Firstly, to put the genetic diversity observed in the NYM *F. lugubris* populations into context with other red wood ant populations across Britain and ascertain whether it is indeed unusually high, and if so, why? This will allow us to determine whether hybridisation may have contributed to this diversity. Secondly, to investigate if there is conflict between mitochondrial haplotypes and morphological data within populations across the British Isles, which may suggest current or historic gene flow between species.

2.3 Methods

2.3.1 Sampling

We sampled adult workers from 241 *Formica rufa* group nests from 34 woodland regions across the British Isles between 2018 and 2021, 162 of which were morphologically inspected. Of these, 114 were selected for mtDNA sequencing (see Fig. 2.3 for nest locations) based on their location, the species composition of the site and surrounding sites, and the morphological species identity of the nest. Sampling sites were selected based on expected species presence using distribution data from the National Biodiversity Network Atlas (NBN Atlas, 2018). To investigate within species genetic diversity and potential introgression, we included five categories of sampling site based on species composition: Three with a single species present and a further two where a species pair may come into contact (*F. rufa* and *F. lugubris* in Wales, or *F. aquilonia* and *F. lugubris* in Scotland) and gene flow between species is currently possible. In some sampling locations, the woodlands were arranged such that con- or heterospecific populations are separated but may remain within dispersal distance of each other (*i.e.* sexuals would be able to meet during mating flight thus potentially allowing hybridisation between species). Additional workers were included from four *F. lugubris* nests from Longshaw Estate in the Peak District, four *F. lugubris* nests from the North York Moors, and two nests (one *F. rufa* and one *F. polyctena*) from Belgium. All specimens were stored in 100% ethanol at –8°C.

2.3.2 Morphological species identification

The species of the British nests we sampled ($n = 159$) and those from Longshaw Estate ($n = 4$) were identified using morphological characters described previously (Skinner and Allen, 1996; Hughes, 2006; Stockan et al., 2016). We selected two numerical and two binary morphological measures based on their scope to discriminate between the three

species found in the British Isles, and the feasibility of inspecting a suitable number of nests within the limited time frame imposed by the project plan (we discuss method limitations and potential species misidentification below). To ensure no bias during identification, we separated ten workers from every nest and assigned them a blind-code before examining three to ten workers per nest under a stereomicroscope, and recording the number of standing setae (erect hairs, hereafter simply "hairs") on the head and thorax (the "nCH" and "nPn" counts respectively described in Stockan et al., 2016). We used the mean nest hair count ranges described in (Stockan et al., 2016) to distinguish the species based on the above, however, distinguishing between *F. aquilonia* and *F. lugubris* is difficult using these characters alone as there is a considerable overlap between the expected hair count ranges for these species. As such, if a specimen was not identified as *F. rufa* based on nCH and nPn alone, we used the position and relative abundance of erect hairs on the mesopleuron and petiole to label the worker as "*F. lugubris*–like" or "*F. aquilonia*–like" based on the description in Hughes (2006) (see Appendix A Fig. A.1 for schematic of our morphological data collection process). At the nest level, the species of a nest was based on the mean values for both the hair counts and labelling across all workers measured from that nest. Where identification was relatively simple and species was clear (e.g. any *F. rufa* nest, or very hairy *F. lugubris*) fewer individuals were required. Once all nests had been assigned a species identification, we un-blinded the data. Nests where species identification was uncertain based on these characters were noted as potential intermediates and included in sequencing work (see Beresford et al., 2017). The workers used in the morphological identification process were retained and stored separately from the rest of the specimens sampled from that nest.

A complication in identifying *F. rufa* nests arises from the possibility that British *F. rufa* may be hybrids of *F. rufa* and *F. polyctena* (Seifert et al., 2010; Seifert, 2021). *F. polyctena* is not found in Britain, but *F. rufa* × *F. polyctena* do exhibit slightly different morphology to *F. rufa*, which, if observed in our data, may support the hybrid

hypothesis. We further explore the origins of British *F. rufa* using genomic data in Chapter 3.

2.3.3 Mitochondrial sequencing

We extracted total genomic DNA from 123 whole workers using the QIAGEN DNeasy® Blood & Tissue Kit following the manufacturer's protocol for insects, with an initial homogenisation step using microtube pestles and overnight proteinase K digestion at 56°C. Worker DNA from 6 nests from Finland (extracted previously; Beresford et al., 2017) was included in the sequencing, and as morphological inspection was not possible we used the species identities assigned by Beresford et al. (2017): one *F. aquilonia*, one *F. polyctena* and four likely *F. aquilonia* × *F. polyctena* hybrids.

A 793 bp fragment of the mitochondrial gene *cytochrome oxidase I* (COI) was amplified in 127 samples using primers COI-IIa and COI-IIb (Holzer et al., 2009) with the Qiagen Alltaq® Master Mix Kit in 20µl reaction volumes, including 1µl template DNA, 5µl Alltaq® Master Mix, and 0.5µl of each primer (10µM). PCR conditions followed the programme based on advice from the Qiagen Alltaq® Master Mix Kit instructions: 95°C for 2 min (one cycle), 95°C for 5 s, 52°C for 15 s, and 72°C for 10 s (35 cycles), followed by a final extension step of 72°C for 5 min. PCR products were purified and Sanger sequenced with COI-IIa primer using the Eurofins PlateSeq service. We checked Chromatograms visually using BioEdit (Hall, 1999). We aligned these data using the CLUSTAL W accessory (multiple alignment with 1000 bootstrap replicates; Thompson et al., 1994) in BioEdit (Hall, 1999) alongside additional sequences: (i) 43 samples from Procter (2016) comprising 39 *F. lugubris* and two *F. rufa* from the North York Moors (see note below), one *F. aquilonia* from Scotland, and one *F. exsecta* used here as an outgroup, and; (ii) three GenBank accessions including one *F. rufa* and one *F. polyctena* from Denmark (MT862420.1 and MT862419.1, respectively), and one *F. paralugubris* from Switzerland (EU600788.1). We generated maximum likelihood (ML)

trees in MEGA X (default parameters; Kumar et al., 2018).

Please note that the *F. rufa* sequences from workers sampled in the North York Moors were initially assumed to be *F. lugubris* (Procter, 2016), however, our genomic analysis (Chapter 3) caused us to reevaluate this and after visual inspection we are confident they are actually *F. rufa* workers. This corrected morphological species ID is reflected in our haplotype phylogeny (Fig. 2.2) but not the haplotype and morphospecies map (Fig. 2.3; it is noted in the caption for clarity).

2.4 Results

2.4.1 Species identification and morphological data

Based on morphological inspection of 159 nests, we identified 53 as *F.aquilonia*, 81 as *F. lugubris*, and 29 as *F. rufa*. Note that hereafter we will refer to 32 samples of *F. rufa* because the three nest samples from Gaitbarrows, England, were inspected twice (independently and using different workers each time), and we elected to retain both sets of count data in our dataset. The nCH (unilateral hairs on head silhouette) and nPn (unilateral hairs on pronotum) of nests we included in the sequencing effort are shown in Figure 2.1, and the species means and count ranges are summarised in Table 2.1. All the "expected" hair count ranges described below are found in (Stockan et al., 2016) and were those we used in assigning morphological species. A refined set of values has since been published (Seifert, 2021), which we have considered briefly in our Discussion.

As anticipated based on the keys used, workers from *F. rufa* nests have a very low mean nCH and are morphologically distinct from the other two species (circle shaped points on Fig. 2.1) regardless of whether *F. lugubris* is present in sympatry. Both the species mean and all upper limit for nCH were well within the expected range for *F. rufa* (0 – 3.6), and in fact sat within the range for *F. rufa* × *F. polyctena* hybrids (0 – 1.2).

By contrast, only three of the 32 samples inspected fell within the expected nPn value range for *F. rufa* (12.5 – 45) and this included two samples from the same nest (one of the aforementioned duplicated Gaitbarrows nests; 16_03). Of the remaining 29 nests, 17 had an nPn 12.5 but within the expected range for *F. rufa* × *F. polyctena* hybrids (5.8 - 16) and the remaining 12 were lower still and instead within the range of *F. polyctena* (0.1 – 5.6).

	nCH μ	nCH σ	nCH range	nPn μ	nCH σ	nCH range	n
F. aquilonia	6.35	2.06	2.33–10.43	6.80	2.93	1.83–13.86	53
F. lugubris	17.84	4.73	4.57–28.50	20.72	7.31	4.00–38.67	81
F. rufa	0.14	0.20	0–0.67	7.51	4.20	0.50–16.50	32

Table 2.1: Species level summary of standing setae on the head silhouette (nCH) and pronotum (nPn). The data given are calculated from the nest level data from all sampled nests assigned to each species, including mean (μ), standard deviation (σ), and the upper and lower limits of the range. The number of nests (n) included for each species is also given. The n for *F. rufa* is 32 because three nests were inspected (using two independent sets of workers) twice and we retained the data here.

Distinguishing *F. aquilonia* and *F. lugubris* from each other was more challenging and their morphological similarity resulted in an overlap in the numerical "hairiness" characters we used (Fig. 2.1). All *F. aquilonia* nests (cross shapes in Fig. 2.1) form a cluster within the expected values for nCH (1.3 – 12.3) and all but one nest adhered to the <13 value for "weakly haired" specimens. The only nest that fell above this nPn threshold (nPn = 13.86) was amongst the nests difficult enough to identify to be noted as potential intermediates, of which 15 were included in mtDNA sequencing (indicated by grey circles on in Fig. 2.1; eleven were ultimately assigned an *F. aquilonia* species ID and four a *F. lugubris* ID). *Formica lugubris* (triangle shaped points) does not cluster strongly like *F. rufa* and *F. aquilonia*, with much higher intraspecific variation and an overlap with the *F. aquilonia* cluster. This is consistent with the expected nCH range for *F. lugubris* of

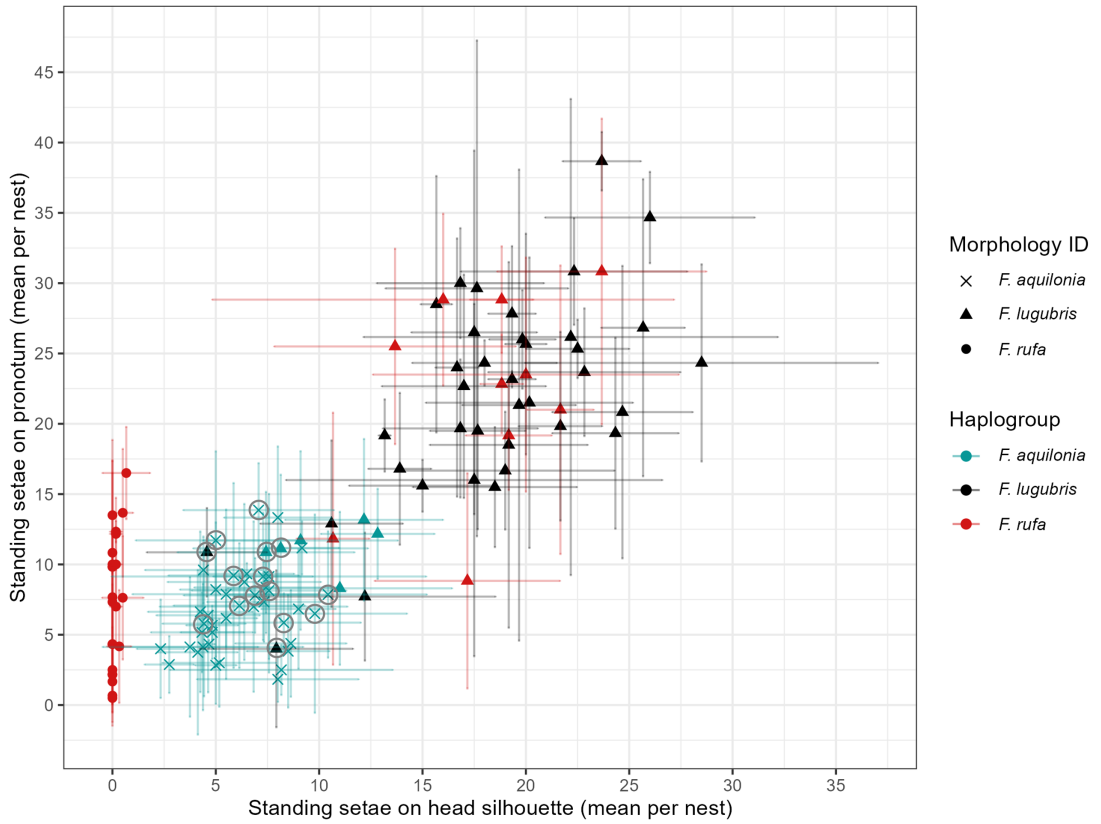

Figure 2.1: Scatterplot of the standing setae on the head silhouette (nCH; x axis) and standing setae on the pronotum (nPn; y axis) of workers from all British nests for which we generated both morphological and mtDNA sequence data. Each point represents a nest and shows the mean value of all workers inspected for that nest. Point shape is based on morphological species identification: cross for *F. aquilonia*, triangle for *F. lugubris*, and circle for *F. rufa*. Points are coloured based on the mitochondrial COI-II haplotype: teal for *F. aquilonia* group, black for *F. lugubris* group, red for *F. rufa* group. Lines around each point denote the standard deviation of the mean nCH (horizontal) or mean nPn (vertical). The grey circles indicate nests noted as potential intermediates based on morphology.

5 - 62, and all "overlapping" nests are within this range except one with nCH of 4.6. All nests with mean nCH >11 and/or mean nPn >15 were designated *F.lugubris* based on morphology, and the mitochondrial haplotypes of these nests is discussed below.

We do not exclude nest species misidentification as an explanation for some *F. lugubris* nests positioned in or near the *F.aquilonia* cluster on Fig. 2.1 and, indeed, for the

ID of some *F.aquilonia*. This is particularly the case in light of the 17 nests noted as potential intermediates due to their mixed morphology. Due to both the small number of characters used here (e.g. compared to the NUMOBAT morphometric method; see Seifert, 2009) and the overlap within the morphospace characterised in Fig. 2.1, we will describe Scottish nests as "*F. aquilonia*–like" or "*F. lugubris*–like" to ensure this uncertainty is clear.

2.4.2 Mitochondrial sequencing

The sequence diversity in our COI-II fragment was low. We identified 10 haplotypes across six species of *F. rufa* group from locations across Europe. We included *F. exsecta* (subgenus *Coptoformica*; sequenced by Procter, 2016) as an outgroup. Nine of the ten haplotypes were present across the three species extant in the British Isles (Fig. 2.2), including three unique to specific woodland locales (see below). Although we could not include all *F. rufa* group species currently present in Europe, the overall tree topology is consistent with that found previously using mitochondrial data (*cytochrome b*: Goropashnaya, Fedorov and Pamilo, 2004; Goropashnaya et al., 2012).

We observed very low sequence diversity within the "*F. lugubris* group": one geographically ubiquitous haplotype (A), and haplotype B with a single nucleotide polymorphism (SNP) present at a single woodland locale in England (haplotype B, Hebden Bridge) (see Fig. 2.3). Despite further low sequence diversity where single SNPs differentiate haplotypes, we identified four haplotypes within the "*F. aquilonia* group", including haplotype D unique to a single site in Cambus O'May (but represented in both a morphologically *F. aquilonia*-like and a *F. lugubris*-like nest). We observed four further haplotypes in the "*F. rufa* group" clade, only two of which included samples from British *F. rufa* nests (haplotypes I and J). Haplotype G of the "*F. rufa* group" was found in a single sample of unconfirmed species (probable hybrid between *F. aquilonia* and *F. polyctena*) from Finland (Beresford et al., 2017). Finally, haplotype H of the "*F. rufa* group" clade

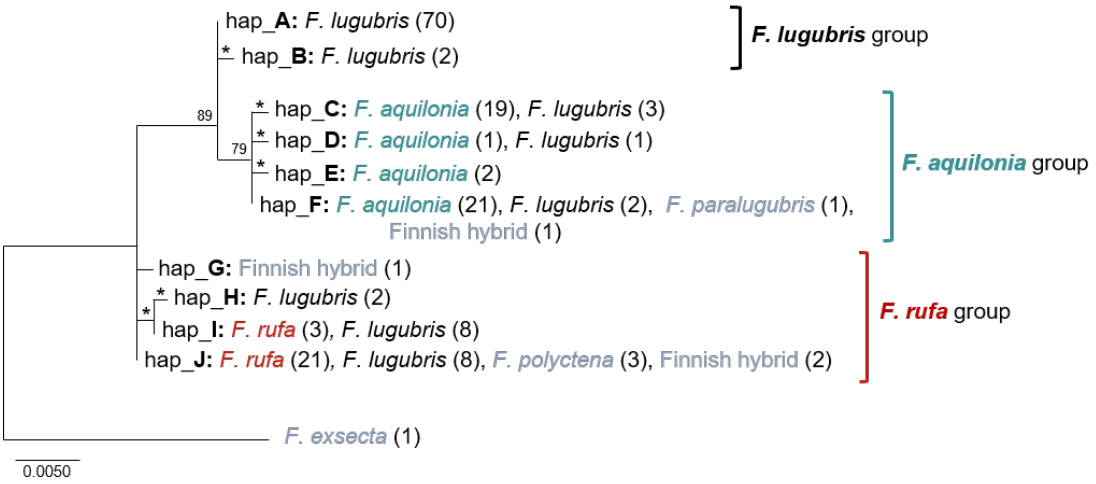

Figure 2.2: Simplified representation of the maximum likelihood tree of phylogenetic relationships between 10 haplotypes A-J, with *F. exsecta* as an outgroup. Branches are labelled with the haplotype letter (corresponding to those in Fig. 2.3 below) and the species present. Numbers in brackets denote the number of individuals of a species present in that position. The three major taxon groupings are colour coded based on the British species present: *F. lugubris* (black), *F. aquilonia* (blue), and *F. rufa* (red). Other species named are coloured grey to make our sample positions clearer. The numbers at certain nodes are bootstrap values, and stars represent a bootstrap value of 62–64 (removed for ease of viewing).

was only observed in two previously sampled *F. lugubris* nests from a woodland locale (Hazel Head) within the North York Moors (nest HH01 sequenced by us, and nest hh02 sequenced previously by Procter, 2016).

Figure 2.2 clearly indicates that *F. aquilonia*–like and *F. rufa* workers are restricted to their clades, however, there are workers from nests we identified morphologically as *F. lugubris* present in almost every haplotype on the tree. If our morphological species assignments are largely correct, this pattern suggests unidirectional mitochondrial introgression from the other two species into *F. lugubris*. However, whilst the *F. rufa* versus *F. lugubris* identifications are robust due to their dissimilarity, and thus the inference of one-way introgression is likely to reflect nature, this cannot be said of the *F. aquilonia* versus *F. lugubris* identifications and this appearance of mtDNA gene flow in only one

direction may be a result of methodological limitations (see further discussion in next section).

The geographical distribution of haplotypes in British samples are shown alongside their morphospecies in Figure 2.3. Nest morphospecies (outer pie segments) conform to expected species distributions based on the NBN data (NBN Atlas, 2018) used in sampling planning with the exception of a single nest in Scotland (nest 45_01; see following section).

2.4.3 Morphology and haplotype conflict

We found 26 nests morphologically identified as *F. lugubris* either here $n = 16$ or previously ($n = 10$; Procter, 2016) were positioned in either the *F. rufa* $n = 20$ or *F. aquilonia* $n = 6$ clades on the ML phylogeny (Fig. 2.2). Moreover, haplotype H in the *F. rufa* clade is only found in *F. lugubris* nests from the North York Moors with Figure 2.3 showing all three *F. rufa* haplotypes are present in North York Moors and nowhere else. This is also visible in Figure 2.1 where morphologically *F. lugubris* or *F. lugubris*–like nest data-points (triangles) are coloured either red or blue, denoting a *F. rufa* or *F. aquilonia* haplotype, respectively. The red *F. rufa* haplotypes can be seen dispersed amongst nests that are very clearly not *F. rufa* morphologically. Conversely, the conflicting *F.lugubris*–like morphology and *F. aquilonia* haplotype nests (blue triangles) are all found in or close to the *F. aquilonia* morphospace cluster. There are also black triangles (*F. lugubris* morphology and haplotype) within the *F. aquilonia* cluster. All of these nests are from Scotland, where *F. lugubris* and *F. aquilonia* co-occur, and as such are either (i) "good" *F. lugubris* that are at the lower end of the "hairiness" spectrum using our ID method, or (ii) are misidentified *F. aquilonia* with an introgressed *F. lugubris* haplotype. The presence of a red triangle (*F. lugubris* morphology with a *F. rufa* haplotype) within the *F. aquilonia* morphology cluster does show there are "good" *F. lugubris* within that hair count range. Interestingly, only two of the 16 nests with

mismatched morphology (identified here) and haplotype were originally noted as potential intermediates (nests 62_05 and 62_11, both from Loch Insh, Scotland). These latter observations may suggest the ID method to be quite robust, however, at least one nest has been noted as a likely misidentification : Nest 45_01 from Loch Achall in Scotland was identified morphologically as (*F.lugubris* despite this site being beyond the maximum northeasterly range of the species, and thus this ID is likely incorrect (a conclusion supported by all four other nests from this site being assigned to *F. aquilonia* and all five possessing *F. aquilonia* haplotypes). We retained the original species ID for nest 45_01 throughout this thesis.

Figure 2.3 shows no geographic structure in *F. rufa* haplotypes (inner pie segments). Mitochondrial introgression from *F. rufa* into *F.lugubris* has been detected wherever the species occur together, including in one woodland locale where only *F.lugubris* was found during our sampling (Coed y Cefn, Central Wales) though *F. rufa* may be present. Similar to *F. lugubris* haplotype A, the *F. aquilonia* haplotype F is found across its geographic range. The remaining *F. aquilonia* haplotype diversity is largely limited to the woodland locales in the Cambus O'May region, where all four haplotypes of the *F. aquilonia* clade are found.

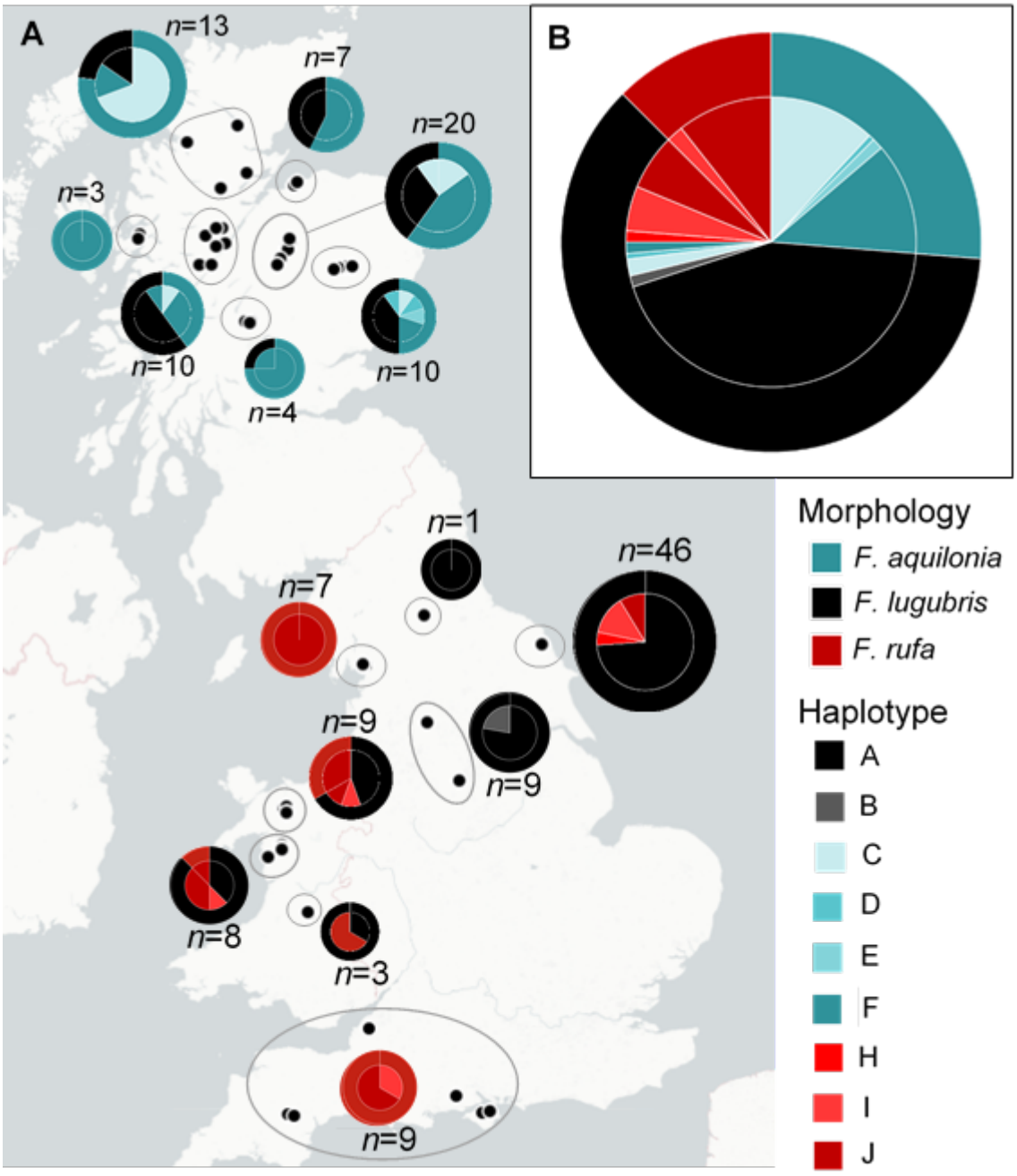

Figure 2.3: Sampling locations of *F. rufa* group nests across the British Isles (black circles; $n = 114$). Nested pie diagrams denoting relative compositions of nest morphospecies (outer pie) and COI-II haplotype of a worker from each nest (inner pie). The worker haplotypes are positioned to correspond to the correct morphospecies of the outer pie *i.e.* each separate nest is aligned despite not being represented as an individual segment. The three colour groups of the morphospecies and haplotypes match: *F. lugubris* (grayscale), *F. aquilonia* (blues), and *F. rufa* (reds). Map A shows the geographic pattern of the haplotypes and corresponding species, with sites grouped so as to be visually informative. Plot B summarises all British Isles nests ($n = 159$). Note: ddRAD analysis (Chapter 3) caused us to reevaluate the morphological ID of two of the *F. lugubris* nests from Procter (2016), and after visual inspection we are confident they are actually *F. rufa* workers. This is not reflected on the map.

2.5 Discussion

Our data provide clear evidence of hybridisation among *F. rufa* group species in Britain. Furthermore, they suggest the mtDNA haplotype diversity previously found in the NYM populations (Procter, 2016) is a result of gene flow between *F. lugubris* and *F. rufa*, with haplotypes introgressed from the latter into the former. At first glance all the introgression appears to be unidirectional, with both *F. aquilonia* and *F. rufa* mitochondrial COI-II haplotypes present in *F. lugubris* nests. However, whilst this is a robust inference for easily-distinguished *F. lugubris* and *F. rufa*, when considering *F. aquilonia* and *F. lugubris* this may be better reflect limitations in our morphological identification method than what is occurring in nature. Our findings suggest neither the morphological measures nor mitochondrial sequencing we used are sufficient to correctly infer nest species if used alone, with the exception the distinct morphology of *F. rufa* and *F.lugubris* if sampled in England or Wales.

The presence of multiple *F. rufa* clade haplotypes in *F. lugubris* workers from across Wales and the North York Moors is indicative of gene flow between these species (Fig. 2.2). The morphological dissilimilarity of these species makes misidentification unlikely, and the findings are concordant with the *F. lugubris* × *F. rufa* hybrid found previously in the Peak District, England (based on morphometric analysis; Seifert (2021)). In some Welsh locations (particularly in Snowdonia, northern Wales) this suggests current or at least recent gene flow between species as the two species are within dispersal distance of each other. This is not necessarily the case in the North York Moors, where small populations of *F. rufa* persisted until at least 2013 (Procter, 2016) but may have recently become extinct (E. J. H. Robinson, pers. comm., 2022), though the prevalence of *F. rufa* haplotypes suggests gene flow historically if indeed only one species is now present.

The *F. lugubris*–like morphs positioned in the *F. aquilonia* mtDNA clade (Fig. 2.2) may be the result of gene flow between species or may indicate the morphological characters

used here are insufficient to reliably distinguish species in Scotland. The presence of nests identified as *F. lugubris* by both morphology and mtDNA haplogroup (black triangles) in the *F. aquilonia* morphospace cluster on Figure 2.1 could be interpreted as supporting the hypothesis that the species identities may be correct and the haplotypes are introgressed from *F. aquilonia*. However, conversely it may be that these nests have been misidentified as *F. lugubris* and in fact suggest introgression from *F. lugubris* to *F. aquilonia*. The presence of a Welsh *F. lugubris* (with a *F. rufa* haplotype) within the *F. aquilonia* morphospace shows there are robustly identified *F. lugubris* at the lower end of the expected nCH and nPn measures for the species (Stockan et al., 2016). It is important not to over-interpret any single nest, however, and it is clear more data are needed for any robust conclusions to be drawn regarding Scottish populations. When selecting our morphological identification method we had to balance accuracy with practical feasibility for a project of this scope, where more fully developed approaches such as morphometric analysis (e.g. NUMOBAT; see Seifert, 2009) were not possible. Whilst still reliable in many cases, simpler methods such as that outlined here are more prone to identification error and caution must be used when interpreting results.

It must be noted that overall our nCH and nPn values align well with those from the large morphometric analysis in Seifert's (2021) taxonomic revison of the *F. rufa* group. Exceptions to this include the maximum values for nPn and nCH in *F. lugubris* to *F. aquilonia*. Interestingly, the values for British *F. rufa* match well with the data for *F. rufa* × *F. polyctena* hybrids, which supports the hypothesis they are of hybrid origin (Seifert et al., 2010; Seifert, 2021).

The apparent unidirectionality of the *F. aquilonia* mtDNA haplotype introgression into *F. lugubris* is interesting, particularly in light of the converse configuration elsewhere e.g. a morphological *F. aquilonia* with a *F. lugubris* mtDNA haplotype in southern Siberia (Goropashnaya, Fedorov and Pamilo, 2004). As introgression in the opposite direction is

clearly possible, we are left with two main explanations for this seeming unidirectionality: (i) Some *F. aquilonia* and/or *F. lugubris* were incorrectly identified using our slightly "rough and ready" approach, and gene flow is bidirectional; (ii) the gene flow is indeed one-way and may be a result of the slightly different population level effects and/or social behaviours of British red wood ant populations (see section 1.4.4). Recently, whole-genome analysis revealed asymmetric gene flow between *F. aquilonia* and *F. polyctena* across Europe, with introgression from *F. aquilonia* into *F. polyctena* (Portinha et al., 2022). This may be due to stronger prezygotic isolation mechanisms in *F. aquilonia* than *F. polyctena*, higher dispersal of *F. aquilonia* than *F. polyctena*, or difficulty in finding conspecific mates caused by smaller population sizes in *F. polyctena* (Portinha et al., 2022). Male-biased nuclear gene flow in fragmented *F. lugubris* populations in the Peak District (England) has previously been inferred when mtDNA showed strong geographic structure but nuclear microsatellite data did not (Gyllenstrand and Seppä, 2003). Our results may suggest a similar pattern but across species, where *F. lugubris* males may be better fliers and/or have different mating behaviours or timings and mate with heterospecifics rather than *F. lugubris* gynes, whilst *F. aquilonia* or *F. rufa* males are more restricted to their conspecific gynes. *F. aquilonia* and *F. lugubris* exhibit P-type (polygynous and polydomous) behaviours in Britain and *F. rufa* is intermediate between P and M-type in Britain (Stockan and Robinson, 2016). Thus we would expect largely short-distance dispersal by nest budding in the former two species, but *F. rufa* founds colonies by flight followed by social parasitism and is able to travel further distances. Our nest counts and anecdotal observations during our sampling trips suggest *F. lugubris* is the most common species in most locales where multiple species are present, perhaps excluding some sites with *F. aquilonia* in Scotland, and we think it is unlikely that *F. lugubris* males struggle to find conspecifics in the majority of cases.

A further alternative explanation of this one-way introgression could be physiological or behavioural differences due to levels of polygyny (see Seifert et al., 2010, and cita-

tions therein). In reference to *F. polyctena* × *F. rufa* hybrids of central Europe, Seifert et al. (2010) discuss trends of higher worker acceptance of invading mated queens in polygynous nests and an increased gyne size and strength in monogynous colonies (because these queens must displace the existing host queen when parasitising a nest during colony founding) (Seifert, 1991). The increased risk of failure in founding a colony in monogynous queens may also contribute to the introgression pattern as, unlike queens in polygynous colonies, they are not insulated from deleterious effects if their hybrid worker offspring have reduced fitness or increased mortality (see Feldhaar et al., 2008). This would mean that the chance of a successful colony foundation was decreased, despite the monogynous queen initially parasitising a nest and using the *Serviformica* workers to raise her brood. Whilst unlikely to affect *F. aquilonia* and *F. lugubris* interaction as they exhibit very similar levels of polygyny in Britain, these factors could contribute to the entry of mated (more monogynous) *F. rufa* queens into (highly polygynous) *F. lugubris* nests at our sampling sites, whether or not she had mated with a *F. lugubris* male. Additionally, increased frequency of intranidal (within-nest) matings within polygynous nests (Seifert, 1991) would contribute to back-crossing with *F. lugubris* and maintaining morphological distinction. Limited data on social behaviours in British red wood ants leave us unable to draw any robust conclusions as to the impact of social behaviours on hybridisation between species, and further investigation in this field may help shed light on the introgression patterns we found.

Given the fact that mitochondrial genomes are small and focused on coding energy-related cellular processes, we would not necessarily expect the mitochondrial genome of an individual to have an effect on the morphological characters we use to identify species. As such, the conflict we have found between mitochondrial and morphological species designation strongly suggests gene flow but does not allow us to make any further inferences about population structure or levels of hybridisation. A much more detailed portrait may be painted using nuclear DNA sequence data alongside the data presented

in this chapter, and an approach such as restriction associated DNA sequencing (RAD) library preparation would be a cost-effective means of generating large nuclear datasets for the number of nests we sampled. Such a set of RAD libraries would allow us to further investigate the species or population histories in British red wood ants. It would also allow us to detect signals of gene flow between species and make robust inferences about current or historical hybridisation amongst the British populations.

2.6 Acknowledgements

This work was funded by a NERC & Forest Research iCASE PhD studentship to EJHR and KD, supporting JM. We are grateful to A.C. Everett, M. Holgate, D. Lamin, K.A. Monaghan and J. Podesta for assistance with field sampling. Additional specimens were kindly provided by D. D. Burns (Longshaw Estate) and W. Dekoninck (Belgium).

Chapter 3

Assessing the nuclear DNA evidence for hybridisation in British wood ants

3.1 Abstract

Hybridisation between species is common in nature. Recent advances in genome sequencing technology have helped reveal its prevalence across diverse taxa. Mound-building red wood ant (*Formica rufa* group) species play a significant role in forest ecosystems across the Palearctic, and morphological and genetic data have shown extensive hybridisation between these species across Europe. We previously found considerable mismatch between morphospecies identity and mitochondrial haplogroup, and evidence of unidirectional introgression of mitochondrial haplotypes in the three species extant in the British Isles. To further explore signals of gene flow, we generated double-digest restriction associated DNA sequencing (ddRAD) libraries for 123 nests from across the British distributions of *F. rufa*, *F. lugubris* and *F.aquilonia*. The range of *F. lugubris* overlaps with that of both other species and there are forest locales where two species occur in sympatry. To explore the possibility of current gene flow between species we included

samples from both single-species and two-species woodlands. We generated a total of 135 ddRAD libraries for population structure and gene flow analyses. Principle component analyses (PCA) on 7591 SNPs shows three relatively tight genetic clusters that mainly correspond to the morphospecies, with few intermediates. Estimated coancestry analysis supports the species-based clustering of the PCA, however, it also suggests some gene flow between species including a possible F1 hybrid. This gene flow is particularly evident between *F. lugubris* and *F. rufa* across regions of northern Wales. The coancestry results also show various levels of within-species population structure based on geography. Overall our data do not support a hypothesis of hybrid origin for British *F. rufa*, and hint at possible descent from multiple continental "pure" *F. rufa* and hybrid *F. rufa* × *F. polyctena* lineages.

3.2 Introduction

Hybridisation between related species is prevalent in nature. It is now known to occur across at least 10% of animal and 25% of plant species (Mallet, 2005). Improvements to sequencing technology and reduction in the cost of generating large sequence data sets have allowed the role of hybridisation in a variety of evolutionary processes in diverse taxa to be explored with renewed vigour (Mallet et al., 2016).

Mound-building red wood ant (*Formica rufa* group) species are known to hybridise across their ranges in Eurasia (see Thesis Introduction for details), in some cases to the point of the species boundary collapsing and existing species nomenclature only remaining in place in the name of functional pragmatism (e.g. separation of *F. rufa* and *F. polyctena*; Seifert, 2021). Three species of red wood ant occur in the British Isles: *F. rufa* in England and Wales, *F. lugubris* in northern England, Wales and Scotland, and *F. aquilonia* in Scotland. In Chapter 2 of this thesis, we presented evidence of hybridisation between *Formica lugubris* and both the other species where their distributions overlap (or did

so until recently). We sequenced a 793bp fragment of the mitochondrial (mtDNA) gene COI-II and compared the results to morphological species identification data and found considerable mismatch between the two, suggesting gene flow between species.

We found unidirectional introgression of *F. rufa* mtDNA haplotypes into *F. lugubris* nests from England and Wales. Furthermore, our data showed mtDNA introgression from *F. aquilonia* into *F. lugubris* in Scotland that at first glance appeared unidirectional, however, we concluded more data were required to confirm if this reflects nature or limitations in our "rough and ready" morphological identification method. We found some nests difficult to identify (with possibly intermediate morphology) but otherwise little detectable signal of hybridisation occurring in Britain in the morphological data alone, though again this picture might be different if more precise morphometric analyses were applied to our samples (e.g. those described in Seifert, 2009). Our morphological data for British *F. rufa* aligned more closely to those for *F. rufa* × *F. polyctena* hybrids that "pure" *F. rufa* from a recent in-depth morphometric analysis (Seifert, 2021), supporting the hypothesis of a hybrid origin for British *F. rufa* populations(Seifert et al., 2010; Seifert, 2021).

The potentially one-way introgression is interesting in its own right, and may be indicative of the effect of social behaviour on interspecies mating, but also shows neither the morphological measures nor mitochondrial sequencing we used are sufficient to correctly infer nest species (in Scotland) if used alone. These patterns act as a reminder of the limitations of mtDNA if used as the only genetic marker; the mitochondrial genome is matrilineal and does not recombine, which makes it an excellent marker for tracing maternal ancestry but we must be careful not to over-interpret these data. Introgressed mtDNA haplotypes cannot tell us anything about the nuclear genome, which may largely or entirely match that of one parent species (e.g. Seifert and Goropashnaya, 2004). Data from the nuclear genome would allows us to explore population structure and gene flow

further.

Analysis of nuclear sequence data in European *F. rufa* group species suggests varying degrees of gene flow depending on species and geographical location (see Thesis Introduction for summary). Whole-genome analysis of 20 *F. aquilonia* and *F. polyctena* samples provided both further evidence of contemporary gene flow (adding to the extensive body thus far) and supported a hypothesis of species divergence with asymmetric gene flow the species (Portinha et al., 2022). This study also highlighted the importance of understanding the geographical context of study systems when selecting sampling populations, as hetrogeneity in rates of processes such as gene flow will affect demographic modelling (Portinha et al., 2022). A genotyping-by-sequencing approach was used to characterise a hybrid zone between *Formica selysi* and *F. cinerea* (non-*F. rufa* group species), the results of which suggested asymmetry in reproductive barriers to hybridisation (Purcell et al., 2016). A recently published genome assembly will aid future genomics research in the *F. rufa* group (*F. aquilonia* $\times$ *F. polyctena* haploid male genome; Nouhaud et al., 2022).

Reduced-representation genome sequencing is a cost effective means of recovering thousands of single nucleotide polymorphism (SNP) loci from large numbers of samples at relatively low cost. Approaches such as restriction associated DNA sequencing (RAD; Baird et al. (2008)) and related methods such as double-digest RAD (ddRAD; Peterson et al., 2012) and genotyping-by-sequencing (GBS; Narum et al., 2013) quickly became amongst the most widely used for high-throughput SNP discovery and genotyping in nonmodel organisms (Andrews et al., 2016). ddRAD has been used in a wide variety of ecological and evolutionary contexts from analysing connectivity patterns (Escoda et al., 2019) and estimating divergence times (Balmori-de la Puente et al., 2022) in small mammals to investigating population structure and selection in brown trout species (Magris et al., 2022) to inferring genetic diversity in important crop plant Mediterranean sesame

(Basak et al., 2019). Gene flow following secondary contact was found in the Grant's gazelles species complex using RAD data (Garcia-Erill et al., 2021), showing its efficacy in the kind of questions we would like to investigate. Using such an approach would allow us to generate a large, affordable nuclear data set with which to explore population structure and gene flow our *F. rufa* group samples.

We will generate a large set of ddRAD libraries from *F. rufa* group samples across the British Isles. Through analysis of these data we aim to shed light on both intraspecific population structure and gene flow between species the three *F. rufa* group species extant in the British Isles. We will also include samples from continental Europe to help clarify the status of British populations of *F. rufa*, which is thought originate from *F. rufa* × *F. polyctena* hybrids.

3.3 Methods

3.3.1 Sampling

To explore population structure and look for evidence the gene flow between species reflecting that previously observed in mitochondrial sequence data, we initially selected 189 workers from 150 nests sampled across the British Isles and several populations across Europe based on nest species (morphologically identified), mitochondrial haplotype, and the species composition of the sampling sites. For the final analysis (excluding low coverage libraries and technical repeats, see below) our dataset comprised 123 nests from British woodland sites: 20 *Formica rufa*, 65 *F. lugubris*, and 38 *F. aquilonia* based on morphology (Fig.3.1). We included samples from forests with a single species present and those where two species are within dispersal distance of each other thus allowing potential for current hybridisation (Fig. 3.1). We found distinguishing *F. lugubris* from *F. aquilonia* morphologically to be challenging, and thus do not treat the species identities of our Scottish samples as robust. This does not apply to the *F. lugubris* nests

from England or Wales, however, as *F. aquilonia* is not present and *F. lugubris* and *F. rufa* are easy to distinguish morphologically. As such we treat English and Welsh nest species with confidence and Scottish nest species identity with caution. See Chapter 2 for information on morphological species identification.

It has been suggested British *F. rufa* are *F. rufa* × *F. polyctena* hybrids (Seifert et al., 2010; Seifert, 2021). To shed light on this we included two *F. rufa* and four *F. polyctena* samples from Belgium. Finally, we included six hybrid *F. aquilonia* × *F. polyctena* workers from Finland (Beresford et al., 2017) to both act as comparison points for potential hybridisation within British samples, and to help shed light on the position of British *Formica rufa*.

To ensure reliable inferences from our library preparation and sequencing we included two kinds of technical repeat: one at the worker level (including the same worker DNA extraction twice; $n = 17$) and a second at the nest level (including a second worker from the same nest; $n = 20$.) These repeats gave a total of 189 libraries prepared before exclusion of technical repeats and those with poor coverage.

3.3.2 DNA extraction and ddRAD library preparation

We extracted total genomic DNA from 177 whole workers using the QIAGEN DNeasy® Blood & Tissue Kit following the manufacturer's protocol for insects, with an initial homogenisation step using microtube pestles and overnight proteinase K digestion at 56°C. To maximise DNA yield we ran the final elution step twice, running the eluent through the spin columns twice giving an elution volume of 200 µL. We also included DNA from 6 nests from Finland (extracted previously; Beresford et al. 2017) in the library preparation. DNA integrity was checked using agarose gel electrophoresis and quantified using a Qubit or Quantifluor fluorimeter. Where required, we increased DNA concentration by evaporation (in vacuum heater) and resuspension in 50 µL of double-

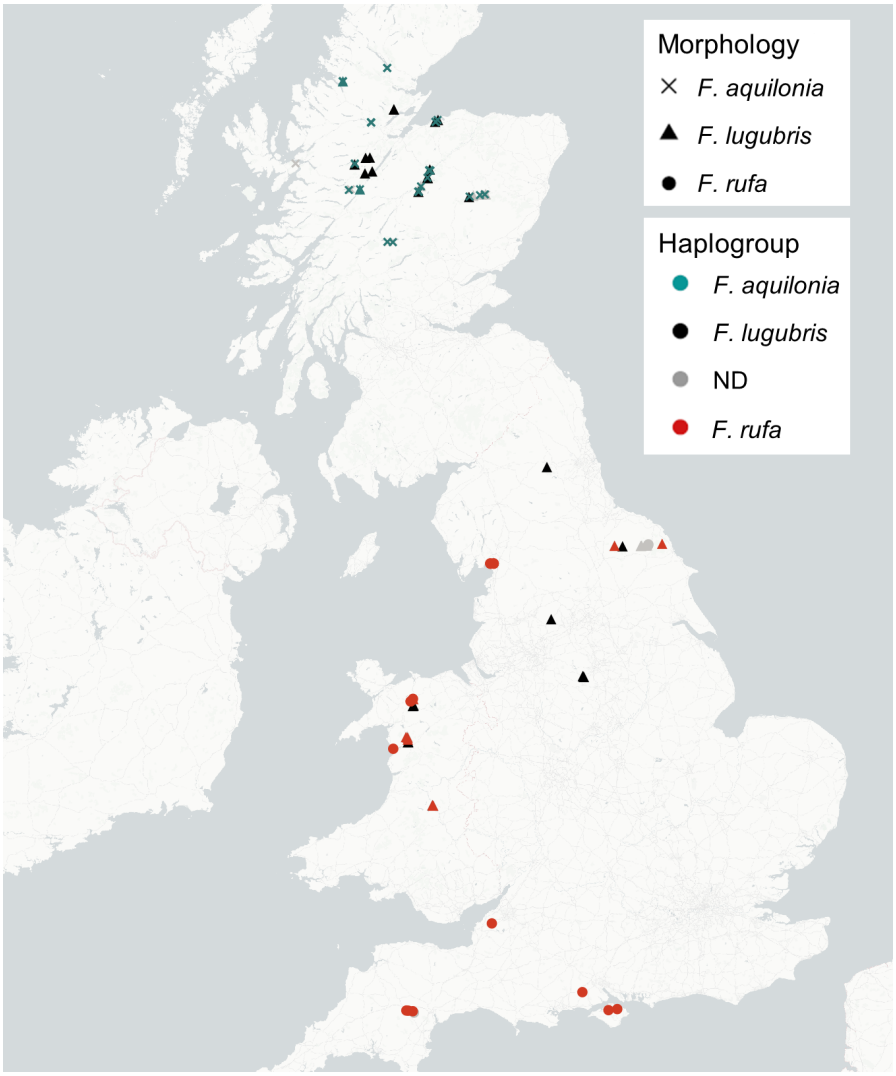

Figure 3.1: Locations of *F. rufa* group nests sampled across the British Isles included in this study ($n = 123$). Point shape is based on morphological species identification: cross for *F. aquilonia*, triangle for *F. lugubris*, and circle for *F. rufa*. Points colour is based on the mitochondrial COI-II haplotypes we recovered: teal for *F. aquilonia* group, black for *F. lugubris* group, red for *F. rufa* group, and grey for unsequenced nests. As a nest could include workers with different mitochondrial lineages, here we only included haplotype assignments if they apply to the specific worker used in ddRAD library preparation, thus "ND" includes 14 nests that have been sequenced but using a different worker. Not every nest from the New Forest or North York Moors is indicated individually due to missing geolocation data.

distilled water (ddH_2O).

Double digest RAD (ddRAD) libraries were prepared based on the protocol of DaCosta and Sorenson (2014). We double digested 300-1000ng sample DNA with *EcoRI* and *PstI* restriction enzymes (both New England Biolabs) using a reaction mix of 10 µL sample, 2 µL (40 U) of each enzyme, 2.5 µL 10X NEBuffer 2 (New England Biolabs), and 8.5 µL ddH_2O. The digests were incubated at 37°C for 18 hours followed by 20 minutes at 65°C to deactivate the enzymes. Each library was barcoded with a unique combination of P2 (paired-end compatible plate-level) and P1 (sample-level) adapters ligated using rATP (Promega), 10X NEBuffer 2 and T4 ligase (New England Biolabs), incubated at 24°C for 30 minutes followed by 20 minutes at 65°C to deactivate the ligase. Samples were size-selected to 300-450bp with agarose gel electrophoresis using custom internal size standards, and extracted from the gel using a QIAGEN MinElute® Gel Extraction Kit following manufacturer's protocol. We amplified the size-selected DNA using a PCR reaction mix of 10 µL size-selected template DNA, 3 µL each of the forward and reverse RAD primers (10 µM), 30 µL Phusion High-Fidelity PCR Master Mix (Finnzymes), and 14 µL ddH_2O with the following cycling conditions: 30 seconds at 98°C, 30 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 40 seconds at 72°C, followed by 5 minutes at 72°C. The PCR products were purified using AMPure XP beads (Agencourt) then quantified first using a Quantifluor fluorimeter and subsequently with qPCR (KAPA Biosystems) after pooling each row of the PCR plate to equalise the concentrations. We then pooled the rows to a final concentration of 5nM and submitted to the Liverpool Centre for Genomic Research (UK) for sequencing on one lane of the Illumina NovaSeq 6000 using paired-end 150bp sequencing.

3.3.3 SNP identification and filtering

We began processing the raw reads by trimming and removing Illumina adapters using Trimmomatic v0.39 (with parameters `SLIDINGWINDOW:4:30 MINLEN:80`; Bolger

et al., 2014). We removed PCR duplicates using the `clone_filter` module from the STACKS pipeline (Catchen et al., 2013), removed the P2 adapters (DBR region) with CUTADAPT (Martin, 2011), and de-multiplexed using STACKS `process_radtags` (Catchen et al., 2013). We aligned each library to a *F. aquilonia* × *F. polyctena* hybrid male reference genome (Nouhaud et al., 2021) using BWA-MEM (default parameters; Li (2013)) alongside SAMTOOLS (Danecek et al., 2021). At this stage we removed all libraries with fewer than 100k reads or those from which less than 75% reads mapped to the reference genome, as these would not have had sufficient coverage for accurate genotyping and are indicative of poor library construction. We also removed technical repeat libraries (see Sampling section above) after first checking each library from a nest clustered together in a preliminary coancestry analysis (see below for details). We found all libraries from each nests did group together so it was appropriate to remove the technical repeats, retaining only a single library from each nest (in this case the library with the highest number of mapped reads) except in a single case: one Finnish nest (RUFAG3) for which we had genomic DNA from two workers found to have different mitochondrial haplotypes (see Beresford et al., 2017), and as such both libraries generated from this nest were informative.

We identified single nucleotide polymorphisms (SNPs) and genotyped individuals using the STACKS `gstacks` (Catchen et al., 2013) resulting in 17392 RAD loci. In preparation for filtering we indexed the reference genome and sorted the SNPs using PICARD (*Picard toolkit*, 2019). SNPs were filtered using `bcftools` (Danecek et al., 2021) to select SNPs with a minimum: two alleles, coverage depth of five, genotype quality (Phred score) of 30, and two copies of the minor allele. These criteria removed poor quality SNP calls as well as singleton SNPs, which are more likely to be genotyping errors, leaving 7591 SNPs across 135 libraries for downstream analysis.

3.3.4 Population structure

Nucleotide diversity (π), observed and expected heterozygosity (H_O and H_E), and inbreeding coefficients (F_{IS}) were calculated at the population level using STACKS `populations` (we included parameter `-r 0.65` i.e retaining only loci present in 65% of the population; Catchen et al., 2013 for use in modelling the effects of habitat fragmentation on genetic diversity (Chapter 4).

We carried out a principal component (PC) analysis on the filtered SNP data using PLINK 2.0 (Purcell and Chang, n.d.; Chang et al., 2015) and visualised the results in R (R Core Team, 2022). An initial plot of PC1 and PC2 showed strong clustering between species. To explore any further structure we ran separate PC analyses on samples within the *F. aquilonia*, *F. lugubris* and *F. rufa* clusters (n of 44, 59 and 28 respectively) using the same parameters as above, and with allele frequencies calculated across all 135 samples (using the `--freq` function). Four nests from Finland were not included in the cluster subsets due to their intermediate positions on the initial PC analysis plot.

To infer common ancestry between nests we used `fineRADstructure` (Malinsky et al., 2018), which is relatively robust to the nonrandom missing data associated with ddRAD library preparation (Malinsky et al., 2018). We ran the first coancestry analysis on all 135 samples. This method comprised first calculating an initial coancestry matrix from the filtered genotype data (default parameters), then assigning individuals to populations using an MCMC method (parameters `-x 100000 -y 100000 -z 1000 -X -Y`), and finally generating a tree (parameters `-x 10000 -X -Y`). We plotted the results using a modified version of the R scripts included in the `fineRADstructure` package. To investigate whether British *F. rufa* is of hybrid origin we ran a second analysis (using the same parameters) on only the *F. rufa* ($n = 22$) and *F. polyctena* ($n = 4$) samples. We also included the two samples from Finland that clustered with the *F. rufa* group in our PC analyses (see Fig. 3.2) .

3.4 Results

3.4.1 ddRAD sequencing and SNP calling

We generated ddRAD libraries with sufficient coverage for 135 *F. rufa* group nests from the British Isles, Belgium, and Finland (this number excludes technical repeats, see section 2.2.). A total of 17392 loci were inferred using `gstacks`, of which 7591 biallelic SNPs were retained after quality filtering.

3.4.2 Principal component analyses

Principal component (PC) analysis of the filtered SNP data for 135 nests showed separation of morphology-based species identifications along both PC1 and PC2 (explaining 10.30% and 9.27% variation respectively; Fig. 3.2), with particularly strong clustering over PC2. Only morphologically *F. lugubris* or *F. lugubris*–like nests (triangle points) are found in all three species clusters, otherwise the morphospecies separate strongly. Belgian *F. polyctena* nests (open circles) are distinct from the *F. rufa* grouping, in which the two Belgian *F. rufa* nests sit. The known hybrid samples from Finland (open squares) are positioned either with the *F. rufa* cluster, or between it and the *F. aquilonia* cluster, which is consistent with their *F. aquilonia* × *F. polyctena* hybrid status characterised in Beresford et al. (2017). The presence of *F. rufa* mitochondrial haplotypes in the *F. lugubris* morphospecies cluster supports our previous inference of gene flow causing mitochondrial introgression into *F. lugubris* (see Chapter 2). No *F. aquilonia* haplotypes are positioned in the *F. lugubris* cluster, however, including in the sample we consider a potential hybrid (60_12, see below). Overall, the PCA does not indicate the presence of extensive gene flow between the British *F. rufa* group species.

There is no clear substructure within groups in Fig. 3.2, other than some geographical pattern along PC1 within the *F. lugubris* cluster with Scottish nests generally towards the right (*i.e.* closer to the *F. aquilonia* cluster.) However, the PC analyses on the

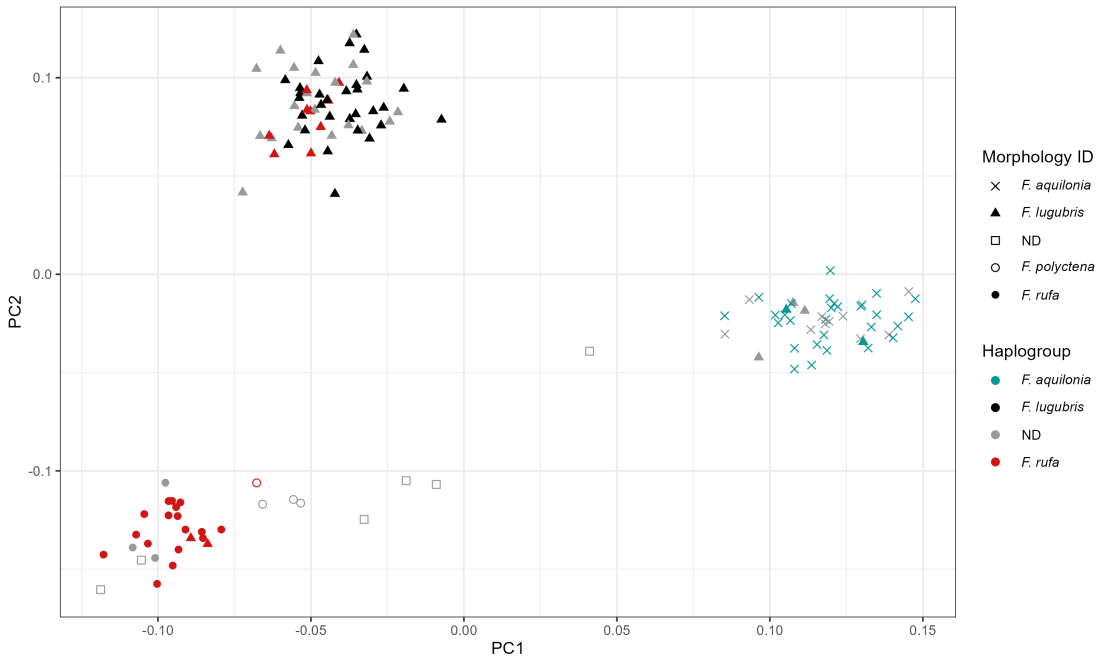

Figure 3.2: Prinicipal component (PC) analysis of filtered ddRAD SNP data for 135 wood ant nests. PC1 explains 10.30% variation, and PC2 explains 9.27%. Point shape is based on morphological identification: cross for *F. aquilonia*, triangle for *F. lugubris*, filled circle for *F. rufa*, open circle for *F. polyctena*, and open square for unidentified nests (Finnish nests only). Points are coloured based on the mitochondrial COI-II haplotypes we recovered (teal for *F. aquilonia* group, black for *F. lugubris* group, red for *F. rufa* group, and grey for unsequenced nests). As a nest could include workers with different mitochondrial lineages, here we only included haplotype assignments if they apply to the specific worker used in ddRAD library preparation, thus "ND" includes 14 nests that have been sequenced but using a different worker.

species-cluster subsets revealed some further structure within species groups (Fig. 3.3), particularly the *F. lugubris* and *F. rufa* subsets (plots B and C, respectively.) The *F. aquilonia* subset (Fig. 3.3, plot A) shows very little geographical substructure, other than the three outliers along the PC2 axis. All three outlier samples come from the same sampling site of Loch Achall, positioned towards the very edge of the northwestern species range and beyond the range of *F. lugubris*. One of the three was originally identified as morphologically as *F. lugubris* (the blue triangle), however, we think this erroneous (and an exemplar of the difficulty in distinguishing the two species, see Chapter

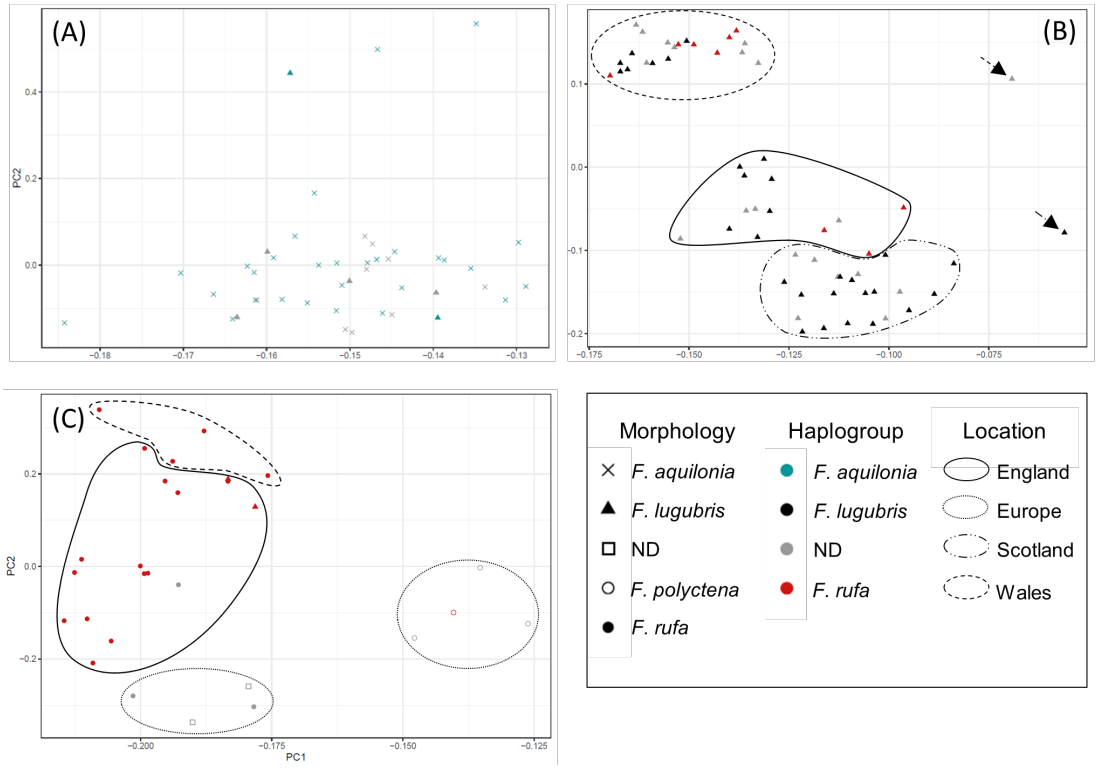

Figure 3.3: Principal component (PC) analysis of filtered SNP data for each morphospecies-based cluster from whole dataset analysis. Point shape is based on morphological identification: cross for *F. aquilonia*, triangle for *F. lugubris*, filled circle for *F. rufa*, open circle for *F. polyctena*, and open square for unidentified nests (Finnish nests only). Points are coloured based on the mitochondrial COI-II haplotype we sequenced: teal for *F. aquilonia* group, black for *F. lugubris* group, red for *F. rufa* group, and grey for unsequenced nests. (A) The *F. aquilonia* cluster (n=44) where PC1 and PC2 explain 16.32% and 3.38% variation, respectively. (B) The *F. lugubris* subset (n=59) where PC1 and PC2 are represent 13.3% and 4.97% of variation, respectively. The polygons indicate sample geography by country (excluding the two outliers), which are largely separated along PC2. (C) The *F. rufa* subset (n=28) subset where PC1 and PC2 explain 44.12% and 5.64% variation, respectively.)

2) and all the samples from this site are *F. aquilonia*. A fourth nest from this locale is situated within the main cluster, and all four are positioned within the main cluster on Fig. 3.2.

The *F. lugubris* subset (Fig. 3.3 B) reveals some interesting population structure with Welsh samples clustering tightly in the upper left (with one exception), English samples

loosely central and Scottish samples restricted to below -0.1 along PC2 (also with one exception.) There is no further substructure within the Welsh samples, irrespective of haplogroup (*F. lugubris* or *F. rufa*), other than a single sample (nest 31_11) far from the main cluster. Based on the coancestry analysis (see section below) this nest is a potential recent hybrid. Perhaps contrary to expectations based on geography, the English samples are positioned closer to the Scottish than Welsh *F. lugubris*, with one sample from the NYM below the -0.1 line on PC2. This structure is not seen in Fig. 3.2, where there in only very limited geographic structuring of *F. lugubris* samples. Within English nests, the North York Moors (NYM) samples show wide variation along the PC1 axis but there is little discernible structure along PC2. There is no substructure amongst the Scottish *F. lugubris* other than a single outlier (nest 60_12), which we also hypothesise to be of hybrid descent based on the estimated coancestry analysis below. Both the potential hybrid samples are discussed further in below. The samples from the NYM occupy a broader space along the PC1 axis than the samples from the entirety of either Wales or Scotland.

Finally, plot C shows the *F. rufa* subset, including six Belgian and two Finnish samples. There are two species in this analysis, clearly separated along PC1 with the four Belgian *F. polyctena* samples occupying the right half of the plot. The British and Belgian *F. rufa* and morphologically undetermined Finnish samples (*F. aquilonia* × *F. polyctena* hybrids that the authors were not able to distinguish morphologically; Beresford et al., 2017) cluster weakly according to geography. The British samples split largely by latitude with the Welsh and some northern English samples clusters towards the top, the southern English samples in the middle, and the Belgian and Finnish *F. rufa* at the bottom. There are two samples from Gaitbarrows (West Lancashire, northern England) within the southern England cluster, but otherwise no deviation from the pattern outlined. PC1 explaining 44% of the variation accounts for the difference between *F. rufa* (all samples) and *F. polyctena* (Belgium), and PC2 mainly showing variation within British

F. rufa samples. This PCA analysis may suggest that British *F. rufa* is not a *F. rufa* × *F. polyctena* hybrid, however, the sample numbers (particularly the *F. polyctena*) are very low and robust inference is not possible based on these data alone.

3.4.3 Estimated coancestry

The `fineRADstructure` analysis outputs a coancestry matrix displaying a summary of the haplotype relationships between pairs of individuals alongside a tree intended to illustrate the relationships between populations (Malinsky et al., 2018). The estimated coancestry analysis on our data (Fig.3.4) supports the species based clustering of the PCA (Fig, 3.2) as is clear in the heatmap matrix and the high support values (all 1.0) for the higher-level branching structure in the tree. The analysis also suggests some additional structuring within species based on geography alongside some signs of gene flow between species. We will outline the within-species structure first in the following section, but save relationships between species (including gene flow and introgression) for the Discussion.

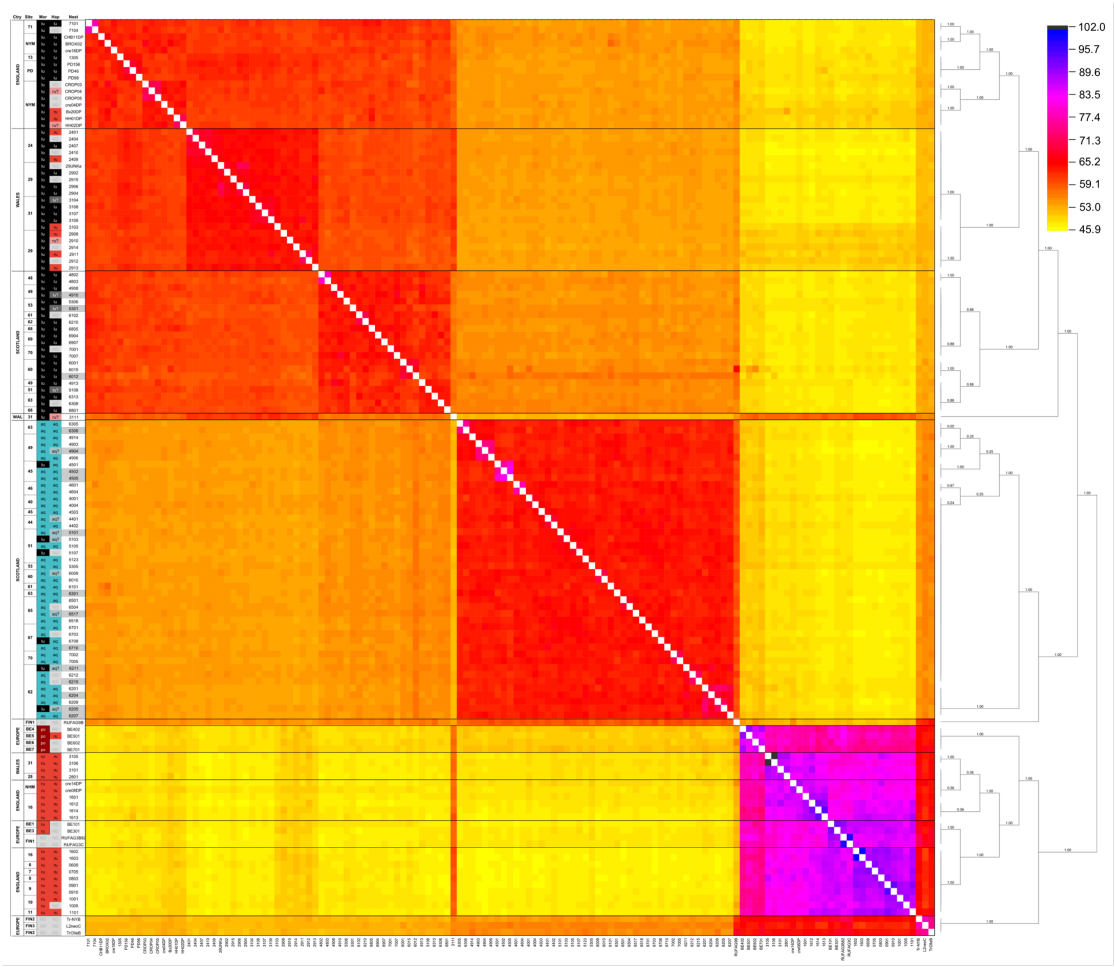

Figure 3.4: Clustered coancestry matrix of all good coverage ddRAD libraries ($n = 135$) from the British Isles, Belgium and Finland. Some sample details are included per row, from left to right: "Ctry" is country of origin, "Site" indicates woodland locale, "Mor" is the morphological species ID, "Hap" is the COI-II haplogroup we sequenced*, and "Nest" is the nest ID of that sample. Both morphology and haplogroups are colour-coded, with blue for *F. aquilonia*, black for *F. lugubris*, red for *F. rufa*, darker red for *F. polyctena*, and grey where no data is available. The nests noted as potential morphological intermediates during have their Nest IDs overlaid with grey. The tree on the right shows the relationships between populations, with posterior assignment probabilities values showing support for each branch. The matrix squares are coloured based on estimated coancestry (scale at top right). Black lines across the matrix separate sample site country within species, for clarity. *As nests could include workers with different mitochondrial lineages, we added question a mark after the "Hap" haplotypes for ddRAD libraries where a different worker from that nest underwent mitochondrial sequencing. For a high resolution version of this plot online click here.

3.4.3.1 Intraspecific structure

The *F. aquilonia* samples cluster together and show quite strong coancestry, but with the lowest level of within-species structure and poor support for the internal branches of the tree (posterior probabilities for a number of the branches are $\leqslant 0.25$). There some woodland sites where samples show increased coancestry such as (in order on the matrix) Boat of Garten (West Cairngorms), Culbin Sands West and Loch Achall (sites 63, 49 and 45, respectively), possibly indicating that multiple nests sampled were from the same extended polydomous colony formed by budding; however, *F. aquilonia* nests from both sites 63 and 45 are also positioned elsewhere in the species cluster structure, making it unlikely this reflects the whole population structure. A large proportion of *F. aquilonia* samples are within a single cluster, supporting the PCA analysis in indicating a lack of intraspecific population structure. There are no nests in this cluster with a non-*F. aquilonia* mitochondrial haplotype, although there are six nests morphologically identified as *F. lugubris*. There is a Finnish nest situated within the *F. aquilonia* group, which is in concord with its mitochondrial haplotype in (Beresford et al., 2017).

In contrast, *F. lugubris* samples show much more within species population structure, perhaps to be expected as it has the broadest geographical dispersal across the British Isles, including a substantial geographical distance separating the English/Welsh populations from those in Scotland. The tree branches are much better supported than those in *F. aquilonia*, with all branches in English and Welsh populations having a posterior probability of 1. Interestingly, the support is slightly lower in Scottish samples, where support values are either 0.88 or 1. Excluding one clearly separated sample (31_11; see following section), the nests from Wales cluster quite strongly together with little further structure. That said, there is some separation of samples based on whether their haplotype conflicts with morphology where nests from Coed y Brenin (site 29) and Gwydir (site 31) with *F. rufa* haplotypes cluster separately. Similarly, samples from

Scotland group together with little geographic structure evident with the exception of all *F. lugubris* from Aviemore (West Cairngorms, site 60) in a cluster with at least one nest hinting at coancestry with *F. aquilonia* (see below.) There is less clear overall grouping of all samples from England, and instead clusters are based on more local geography and, in the case of some NYM nests, the mitochondrial haplotype.

Samples within the *F. rufa* and *F. polyctena* group show much higher estimated coancestry, both within and between the two species (see Fig. 3.5 for our analysis of the latter.) This increased coancestry applies across all *F. rufa*, whether from Belgium, Finland or Britain. The substructure in this species group appears as a series of overlapping geographically unintuitive clusters, starting with two main groups: Most of the northern English samples with Welsh nests, and the Belgian and Finnish *F. rufa* with the southern English samples (plus two from Gait Barrows, site 16A, in the north of England.) Three Arnside Knott (site 16B) nests from the former group also share more ancestry with the British samples in the latter group than the other northern English and Welsh nests. The branching structure of the tree is well supported at higher levels, but less so for the clustering of Welsh and northern English (sites 31, 28, NHM and 16) where some branches are poorly supported (posterior probability of 0.36).

There are three known hybrids from Finland at the bottom of the matrix whose coancestry with the other species is as expected based on the species composition of their populations of origin, and they provide a good indication of what a hybrid signal amongst British nests would look like on this plot. We do not discuss these nests further as their status is already described (Beresford et al., 2017).

F. rufa and *F. polyctena* cluster together very strongly, suggesting coancestry between the two species across our sampling. To test whether this is due to the two species being closely related (as in previous phylogenies; Goropashnaya, Fedorov and Pamilo, 2004; Goropashnaya et al., 2012) or consistent with the hypothesis that British *F. rufa* are

descended from *F. rufa* × *F. polyctena* hybrids (Seifert et al., 2010) we carried out a second estimated coancestry analysis of the nests within this cluster only.

3.4.3.2 *Formica rufa* and *F. polyctena* coancestry analysis

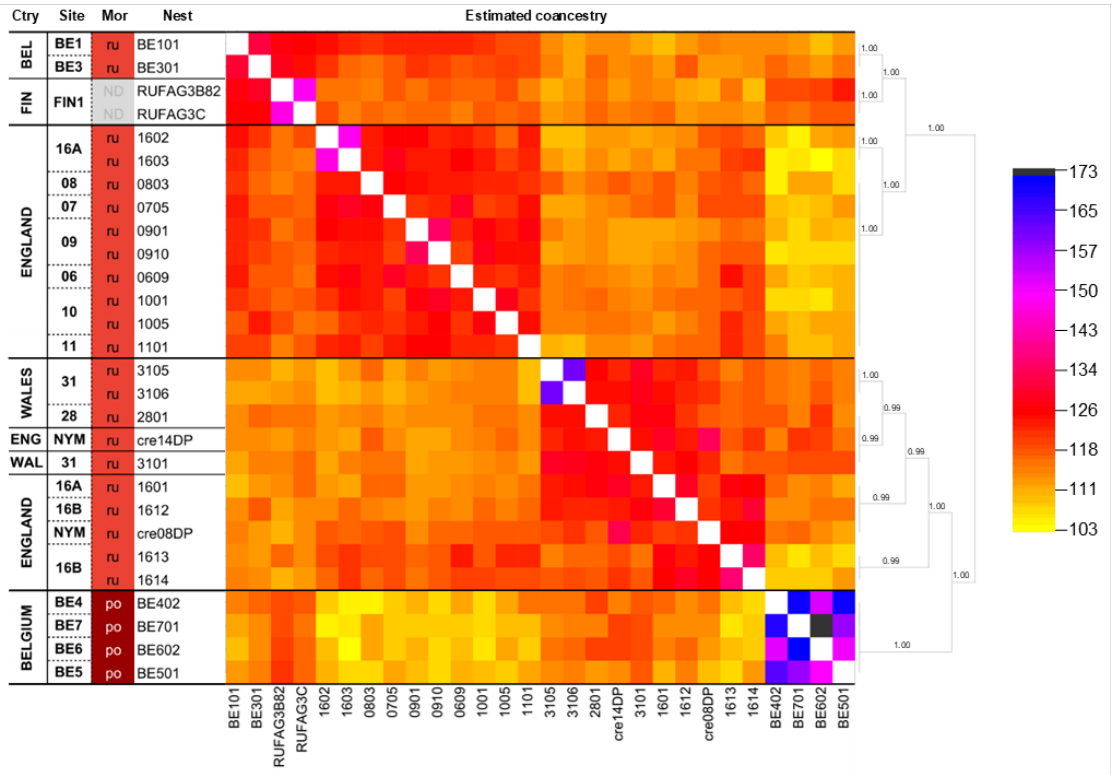

Figure 3.5: Clustered coancestry matrix of the *F. rufa* ($n = 24$) and *F. polyctena* ($n = 4$) samples from the British Isles, Belgium and Finland. Some sample details are included per row, from left to right: "Ctry" is country of origin, "Site" indicates woodland locale of a sample, "Mor" is the morphological species ID, and "Nest" is the nest ID of that sample. Morphology is colour-coded, with red for *F. rufa* and darker red for *F. polyctena*, and grey where no data is available. The tree on the right shows the relationships between populations, with posterior assignment probabilities values showing support for each branch. The matrix squares are coloured based on estimated coancestry (scale on the right).

The *F. rufa* and *F. polyctena* subset analysis (Fig. 3.5) shows a difference between the two species. Here *F. polyctena* shows very high within-species coancestry, and much less with the *F. rufa* nests. This difference does vary across samples, however, with the *F.*

polyctena nests exhibiting greater coancestry with the Finnish, and to a lesser extent Belgian, *F. rufa*. The tree on the right shows the samples split into two main clusters: the top group including Finnish and Belgian *F. rufa*, all the southern English *F. rufa* (with no substructure in the clustering) and two nests from northern England (Gaitbarrows; nests 16_02 and 16_03); the lower group including all Welsh and the remaining northern English *F. rufa* nests with the *F. polyctena* from Belgium. The posterior probability values supporting the branching are either 0.99 or 1, indicating excellent support for these clusters. This strongly suggests some British *F. rufa* are descended from *F. rufa* × *F. polyctena* hybrids and other lineages are not. Furthermore, this is borne out by the level of estimated coancestry indicated by the matrix colouration. Using the level of coancestry seen between Belgian *F. rufa* and *F. polyctena* as a "yardstick", we can see two levels of coancestry shared between British *F. rufa* and *F. polyctena*. They follow largely the same pattern as the tree, firstly with all the southern English (from sites 06 to 11) and two northern English nests show less coancestry with *F. polyctena* than the Belgian *F. rufa*, suggesting they are not of hybrid descent. A second group (the Welsh and remaining northern English nests) appears to share equal or higher coancestry estimates than the Belgian species, indicating a higher rate of gene flow between these British *F. rufa* nests and *F. polyctena* and a possible hybrid origin. These data suggest some British *F. rufa* are of hybrid origin and others are not, with some limited geographic structure to the pattern of dispersal.

3.5 Discussion

Our ddRAD data clearly demonstrate hybridisation amongst the red wood ant species inhabiting the British Isles but, in contrast with our mitochondrial (mtDNA) data, support a picture of sporadic hybridisation events. These rare points of gene flow are followed by repeated backcrossing with one parental species leaving little signal in the

nuclear genome. Overall our data support the hypothesis that British *F. rufa* are descended from multiple colonising lineages, including some of hybrid origin *F. rufa* × *F. polyctena*.

The principal component analysis (PCA; Fig. 3.2) of SNP data from all samples shows three relatively tight genetic species clusters with few intermediates and very little substructure, despite our mtDNA and morphological mismatch data suggesting gene flow. The clearly visible intermediate nests along PC1 are known or suspected hybrids from Finland (Beresford et al., 2017). The species cluster subset PCAs (Fig. 3.3) reveals some largely geography-based structure within the species. The estimated coancestry analysis supports the overall species-based clustering of the PCAs whilst adding some further within species structure, particularly for *F. lugubris* (Fig.3.4).

3.5.1 Species relationships and gene flow

F. lugubris and *F. aquilonia* show greater coancestry with each other than with *F. rufa*, which is concordant with our findings based on mtDNA COI-II sequence data and previous mtDNA phylogenies (Goropashnaya et al., 2012). This coancestry pattern supports the order of species divergence previously inferred from mtDNA, wherein *F. aquilonia* and *F. lugubris* share a common ancestor more recently than they do with *F. rufa* (Goropashnaya, Fedorov and Pamilo, 2004). *F. aquilonia* shares slightly more coancestry with *F. lugubris* from Scotland, which may suggest ongoing low-level hybridisation throughout the range as the species often come into contact. One *F. lugubris* nest (60_12) from Aviemore in the West Cairngorms (site 60) where both species occur in sympatry has higher coancestry with *F. aquilonia* than other *F. lugubris*, which suggests a hybrid descent and may support the latter premise (it is also an outlier on the *F. lugubris* subset PCA, Fig. 3.3 B). Interestingly, this nest was noted as an "potential intermediate" during morphological examination, along with 16 other nests (all of which have been indicated by grey markers on Fig. 3.4) perhaps suggesting this identifica-

tion difficulty was indeed down to hybridisation. However, none of the nests have any particularly obvious signals of mixed ancestry so this is difficult to argue conclusively. The pattern of mitochondrial consistency with estimated coancestry (*i.e.* there are no *F. aquilonia* mitochondrial haplotypes in the *F. lugubris* species cluster and vice versa) does not support regular hybridisation and instead lends credence to the former hypothesis of early gene flow only. This conclusion is further corroborated by the lack of any pattern of increased *F. lugubris* coancestry with in nests that were morphologically *F. lugubris*–like but genetically (both mtDNA and ddRAD) *F. aquilonia*.

A similar pattern can be seen in one of the two sub-clusters of *F. lugubris* from Wales showing a higher amount of coancestry with *F. rufa* compared to their conspecifics from England or Scotland. A high proportion of this cluster also possess *F. rufa* COI-II haplotypes. The introgression of both nuclear and mitochondrial genetic material allows us to infer these nests are descended from *F. lugubris* × *F. rufa* hybrids, which have since back-crossed with *F. lugubris*. Nest 31_11 from Gwydir in north Wales (site 31) has clustered with *F. lugubris* but demonstrates a much higher rate of coancestry with *F. rufa* than the backcrossed nests described. Similar to Scottish nest 60_12 above, this nest is clearly separate from the others on the *F. lugubris* subset PCA (Fig. 3.3 B). The relatively equal level of estimated coancestry it shows for both species suggest it is a very recent and perhaps even F1 hybrid nest.

There are samples in the *F. lugubris* cluster (both English and Welsh) that show mitonuclear discordance, where the a *F. rufa* haplotype conflicts with both the morphospecies and the nuclear genome data. In Chapter 2 we discussed a number of possible explanations for the observed unidirectionality of mtDNA introgression between *F. rufa* and *F. lugubris*, such as behavioural and/or physiological species differences due to colony characteristics (e.g. polygyny), and these may also be relevant here (see Chapter 2 Discussion). Another contributing factor may be genomic processes such as differential

selection acting on the nuclear versus mitochondrial genome. While there is some limited nuclear gene flow among the three British wood ant species where they overlap in their distributions, much greater levels of gene flow are detected using mitochondrial markers (Chapter 2). This suggests that while the species hybridise in nature, nuclear gene flow is perhaps limited by selection acting on incompatible allelic combinations. A similar pattern of non-random mito-nuclear mismatch was found in Finnish hybrid populations (Kulmuni and Pamilo, 2014; Beresford et al., 2017), one suggested explanation for which is selection favouring nuclear introgression in females *i.e.* heterospecific combination of nuclear and cytoplasmic material (see Beresford et al., 2017, for discussion of alternatives).

The tightness of the species clusters in our PCA results (Fig. 3.2) contrast strongly with the virtual continuum of nests shown by PC analysis of a large nuclear dataset from 96 nests of *F. aquilonia*, *F. polyctena* and *F. aquilonia* × *F. polyctena* hybrids in Finland (nine microsatellite markers from 647 workers across the 96 nests; Beresford et al., 2017). In combination with mtDNA sequencing these data revealed a picture of gene flow easily justifying the term "hybrid swarm" (Beresford et al., 2017). Whole-genome analysis revealed asymmetric gene flow between *F. aquilonia* and *F. polyctena* across Europe, with introgression from *F. aquilonia* into *F. polyctena* as the species diverged (Portinha et al., 2022). One hypothesis to explain this was the struggle of *F. polyctena* sexuals to find conspecific mates due to smaller population size (Portinha et al., 2022), which may explain why the hybrid swarm populations of Finland provide a fascinating juxtaposition to the picture of limited gene flow amongst British species. It may be that the small population size and resulting obligate hybridising of Finnish *F. polyctena* means it acts as a "hybrid bridge", contributing to much higher pattern of hybridsation and gene flow found there compared to Britain (J. Kulmuni, pers. comm., 29 July 2022). *F. polyctena* also hybridises with *F. rufa* where the species come into contact, including in parts of Finland (Seifert et al., 2010; Seifert, 2021), which may

support this hypothesis.

3.5.2 British *F. rufa* status

The PCA including all samples (Fig. 3.2) shows *F. polyctena* nests positioned relatively close to the *F. rufa* cluster, and a number of British *F. rufa* are closer these *F. polyctena* than are the Belgian *F. rufa*. In contrast, the subset PCA suggests British *F. rufa* is quite distinct from *F. polyctena*, with the species groups separated strongly on PC1 (which explains 44% total variation; Fig. 3.3 C). However, there are very few samples of *F. polyctena* and continental *F. rufa* present (4 and 2, respectively) and this makes robust drawing robust conclusions using the PC analyses alone impossible.

Comparing the levels of estimated coancestry shared between Belgian *F. rufa* and *F. polyctena* with the levels of coancestry shared between various British *F. rufa* and *F. polyctena* helps us to better infer whether the British nests are likely of hybrid descent.The estimated coancestry analysis (Fig. 3.5) shows *F. polyctena* has much higher coancestry with its conspecifics than any other samples. That said, the cluster structure of the analysis very clearly demarcates two main groups (with strong support of the branching structure) where some British *F. rufa* nests cluster with *F. polyctena* and others with continental *F. rufa*. The two main clusters are: (1) made up of all nests from southern England and two from Gaitbarrows in the north of England, which clusters with Belgian *F. rufa*, and; (2) comprising the remaining northern English nests and all those from Wales, which cluster with *F. polyctena*. This structure is also quite clear in the colour-coding in the pairwise matrix. Cluster 1 is more coherent in shows very little substructure within these British nests, whereas cluster 2 shows more varying levels of coancestry between nests (which is also reflected in the higher number of clusters in the tree). Some of the nests in cluster 2 (28_01 and 31_01 from Wales and cre14DP from northern England) seem to have levels of coancestry with *F. polyctena* higher than Belgian *F. rufa* does. This striking topology of the cluster tree alongside the visual

"yardstick" comparison approach suggest British lineages are descended from different colonising populations from the continent, including some from hybrid populations of *F. rufa* × *F. polyctena*.

These results support the hypothesis that a considerable proportion of British *F. rufa* are of hybrid origin (Seifert et al., 2010; Seifert, 2021). Our morphological data (Chapter 2) are in accordance with this inference, and align better with the *F. rufa* × *F. polyctena* than *F. rufa* data in the in-depth morphometric analysis of Seifert's (2021) taxonomic revision of the *F. rufa* group. Further investigation using a wider array of samples from continental Europe would help shed light on glacial refugia and colonisation routes into Britain.

3.5.3 Conclusions

Our data support a picture of sporadic gene flow between British *F. rufa* group species, and provide a sound footing for further investigation. We aimed to shed light on population structure and gene flow between species within British wood ants, and more in-depth analysis of our genomic data would allow further exploration of patterns of introgression and gene flow. For example, we would expect the level of introgression to be highest at sites where two species occur in sympatry and then decrease as distance from such locales increases, which we could test using approaches such as: ABBA-BABA statistics using Patterson's D (Green et al., 2010; Durand et al., 2011) to further test for the net effect of gene flow across our data, and; $f4$ tests of introgression that are more sensitive to ongoing gene flow among sympatric taxa (Patterson et al., 2012). Further inference of population structure using STRUCTURE (Pritchard et al., 2000) followed by estimating the proportions of ancestry from contributing populations using ADMIXTURE (Alexander et al., 2009) would allow us to make robust inferences about the hybridisation between British wood ants. This approach, especially if aided by the addition of further *F. rufa* and *F. polyctena* from continental populations, may also help clarify the origins

of British *F. rufa*. If British *F. rufa* are descended from different continental lineages (hybrid and otherwise) as our coancestry results suggest, we would see this reflected in different levels of admixture between populations.

We need more data on the ecology and social behaviours of British *F. rufa* group species in order to better understand the patterns of hybridisation indicated by the data we present in Chapters 2 and 3. If the mating and colony foundation behaviours previously discussed (e.g. queen recruitment, levels of polygyny) have contributed to the patterns of introgression we observed between P-type *F. lugubris* much less polygynous *F. rufa*, we would expect to see evidence of such behaviours in the field. Whilst in-depth field study across a broad geographical scale may be unfeasible as the time and expertise required are prohibitive, perhaps a more localised study of one or more of the locales in which we detected gene flow is possible. For example, woodland sites in Gwydir (Wales) would be useful for exploring hybridisation between *F. lugubris* and *F. rufa*. Field studies to investigate characteristics such as levels of polygyny in both species, the comparative size of queens, the timing of nuptial flights, or the presence of *F. rufa* queens in *F. lugubris* nests would help us make much more informed inferences. Sex-biased processes affect selection in Finnish *F. aquilonia* × *F. polyctena* (Kulmuni et al., 2010) and, though the levels of hybridisation we observed are much lower than those found in Finland, it may be that such processes also contribute to patterns of gene flow in British populations. Comparison of sequence data (such as microsatellites) from females and males would allow exploration of the effects (if any) of such processes.

Alongside more in-depth morphometric analysis of our nest samples, these suggested data would allow comparison of the data we have collected to the well–studied red wood ant populations on the continent This would mean the social ecology, hybridisation and morphology of British populations could be described in a much broader geographical context, and the scope for drawing more general conclusions would be improved.

3.6 Acknowledgements

This work was funded by a NERC iCASE PhD studentship to EJHR, supporting JM. The ddRAD library preparations and sequencing was funded by the NERC Environmental Omics Facility (project code: NBAF1206). We are grateful to A.C. Everett, M. Holgate, D. Lamin, K.A. Monaghan and J. Podesta for assistance with field sampling. Additional specimens were kindly provided by D. D. Burns (Longshaw Estate) and W. Dekoninck (Belgium).

Chapter 4

The genetic diversity of wood ants in a fragmented landscape

4.1 Abstract

High genetic diversity within populations or species gives them the potential to adapt to environmental changes. Habitat fragmentation causes isolation and a reduction in gene flow between populations, which can decrease the adaptive potential of a population and increase the risk of local extinction. Wood ants are poor dispersers and woodland specialists, which makes them more susceptible to the isolating effects of fragmented habitats. Here we investigate the relationship between habitat fragmentation and measures of genetic diversity in three wood ant species distributed across Britain. We generated 123 double-digest restriction associated DNA sequencing (ddRAD) libraries with sufficient coverage, from which we obtained 7591 SNPs. All our genetic diversity measures differed between species, with *Formica lugubris* showing the highest diversity. Contrary to expectation sites with multiple species present did not show higher genetic diversity within a given species than single-species sites, although our sample sizes for multispecies sites

were very low. There were species-specific effects of latitude for H_O and π, in both cases with *F. lugubris* increasing in diversity with latitude and *F. aquilona* and *F. rufa* both decreasing. Our measures of fragmentation did not significantly predict genetic diversity in most cases, although we did detect lower H_E in smaller fragments and in areas surrounded by less woodland cover. Overall, we find little evidence for impact of habitat fragmentation on genetic diversity within British wood ants. Further work with a greater range of fragment sizes and a wider range of species combinations would help address this question more fully.

4.2 Introduction

Maintaining genetic diversity is vital for populations to survive long-term. Genetic diversity increases adaptability, and thus resilience to environmental change, reducing extinction risk. Genetic diversity is influenced by effective population size, gene flow, and population structure. Social insects are haplodiploid and have small effective population sizes, often exhibit spatially structured populations, and reduced genetic variation can easily lead to inbreeding depression (Pamilo and Crozier, 1997). This can make them especially susceptible to environmental changes that disrupt gene flow and population dispersal.

Habitat fragmentation can result in populations becoming isolated and locally reduced in size, increasing risks of local extinction. When habitat patch size and connectivity is decreased, gene flow between populations decreases, and this in the long-term leads to decreases in effective population size and reduced genetic diversity (Fischer and Lindenmayer, 2007). Such habitat loss is likely to affect all organisms dependent on these ecosystems, however, it can be especially problematic for species with poor or short-range dispersal because habitat fragmentation is more likely to present a barrier to gene flow in these species. These barriers present a challenge to conservation of habitat special-

ists as anthropogenic land use change causes widespread fragmentation and connectivity loss.

Forest cover in the United Kingdom is a prime example of human-mediated habitat change. Around 13% of land is currently classified as woodland (Forestry Commission, 2022), up from its lowest point of 5% around 1900 (Mason, 2007). Most of this woodland is non-native plantations with ancient broadleaved woodland currently covering only around 2% of land, mainly comprised of smaller fragments (The Woodland Trust, 2017). Anthropogenic fragmentation of ancient, native UK forest has disrupted habitat continuity for woodland specialists, and the effects of continued land use change on such species are not well understood (Synes et al., 2020). In England, for example, woodland patches less than 2 hectares in size comprise just 6.8% of total forest area but make up 75% of total number of patches (Watts, 2006). Characterising the impact of such dramatic habitat fragmentation on genetic diversity, particularly in keystone species of the habitat, is of high conservation value (Vanhala et al., 2014; Watts et al., 2016).

Red wood ants (the *Formica rufa* group) are woodland specialists with an important role in forest ecosystems, including nutrient cycling, habitat modification, and seed dispersal (Stockan and Robinson, 2016). In Britain, *F. lugubris* and *F. aquilonia* generate new nests by budding, and British *F. rufa* also frequently use this mode of dispersal (Stockan and Robinson, 2016). This means habitat fragmentation is likely to present a barrier to gene flow, especially so for social insects (Gyllenstrand and Seppä, 2003). Despite extreme habitat fragmentation, there is little evidence of damaging reduction in genetic diversity or inbreeding depressions in British red wood ants (see Thesis Introduction). This may be due to polygynous colony organisation allowing maintenance of a higher effective population size in these populations (Gyllenstrand and Seppä, 2003), a view supported by the inbreeding depression found in monogynous *F. lugubris* from Ireland (Mäki-Petäys and Breen, 2007). Alternatively, demographic and genetic evidence

from populations of *F. lugubris* in the North York Moors (Procter, 2016) and Scottish *F. aquilonia* (Vanhala et al., 2014) suggests afforestation with non-native conifers may reconnect native, ancient woodland fragments and provide corridors through which population contact (and thus gene flow) can be re-established. However, we should not make broad assumptions about resilience to habitat fragmentation in British red wood ants based on these data. This suggests further investigation is needed to help provide a sound footing for forest management strategies to increase and/or maintain genetic diversity in red wood ants and other specialist species.

The keystone status of these species within woodlands, their poor dispersal abilities and their obligate reliance on woodland habitat, together make them attractive study organisms for investigating the impact of habitat fragmentation on genetic diversity and its conservation implications. Here we explore whether the extreme fragmentation of woodland habitat has affected genetic diversity in British red wood ant populations. To characterise this diversity we analysed restriction associated DNA (RAD) sequencing libraries for 123 *F. rufa* group nests across Britain, each assigned to populations per habitat fragment based on species and location. We expect populations in highly fragmented habitats may exhibit reduced genetic diversity and increased inbreeding within populations as gene flow between populations populations is reduced and small populations lose diversity due to genetic drift. To evaluate this we modelled whether habitat fragmentation predicted four measures of diversity: the observed and expected heterozygosity, nucleotide diversity, and the inbreeding coefficient. Observed heterozygosity (H_{O}) is the proportion of genotypes in a sample that are heterozygous, and can indicate the genetic variability of a population e.g. very low heterozygosity means little genetic variability and may be the result of events such as genetic bottlenecks. Comparing the observed heterozygosity to that which is expected under Hardy-Weinberg equilibrium (HWE), known as the expected heterozygosity (H_{E}), allows inference of processes such as population isolation (and consequent inbreeding) or, conversely, the mixing of formerly

isolated populations. A similar measure to expected heterozygosity, nucleotide diversity (π) is the mean pairwise difference between each between all possible pairs of samples in a given population (Nei and Li, 1979), and is described as equivalent to expected heterozygosity in the method used to estimate it hereHohenlohe et al. (2010); Catchen et al. (2013). The inbreeding coefficient F_{IS} measures the difference between observed and expected heterozygosities of an individual with respect to sub-population (Wright, 1931, 1978) and a high value suggests a considerable degree of inbreeding. These measures show related but slightly different aspects of genetic diversity, and as such we would predict a fragmented, isolated set of populations to have (i) reduced H_{O}, H_{E}, and π, and (ii) higher F_{IS}. Thus, we would expect our habitat fragmentation measures to be able to predict all four measures in our models (with H_{E} and π acting in part as controls for each other).

4.3 Methods

4.3.1 Wood ant nest samples

For the following analyses we included a single, good quality double-digest restriction associated DNA (RAD) sequencing library (*i.e.* one worker) from each British nest ($n = 123$) we sequenced in Chapter 3, resulting in the same dataset discussed in Sections 3.3 and 3.4 less the Finnish and Belgian samples (Fig. 4.1). Double-digest RAD (ddRAD) libraries were prepared based on the protocol of DaCosta and Sorenson (2014) using genomic DNA extracted from whole worker ants (see Section 3.2 for sampling and ddRAD library preparation methodology).

4.3.2 Woodland and population characteristics

In order to characterise the degree of fragmentation of the woodland sites that we sampled we located every ddRAD library (*i.e.* nest sample) based on the their woodland

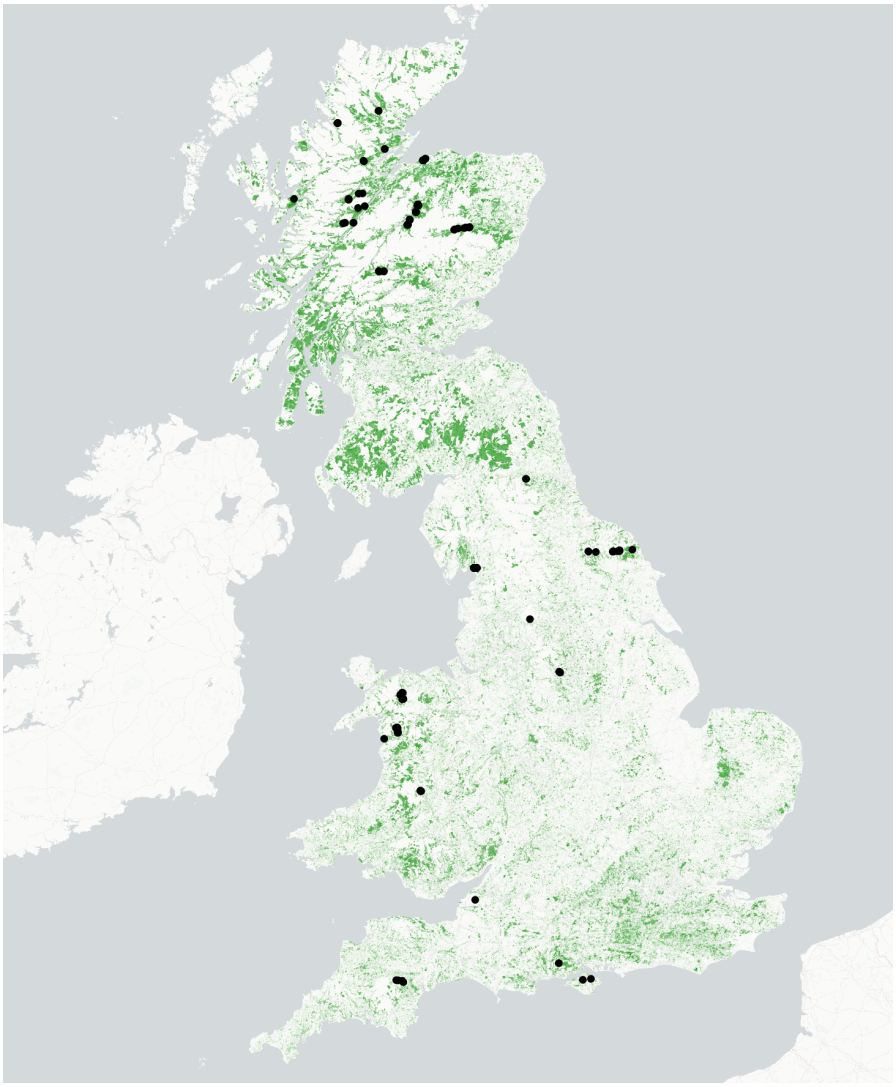

Figure 4.1: Locations of *Formica rufa* group nests (black circles) sampled across the British Isles between 2018 and 2021 included in this study (excluding technical repeats and libraries with poor coverage; $n = 123$). The green polygons show the woodland cover data from the National Forest Inventory 2019 survey (Forestry Commission, 2019). Not every nest from the New Forest or North York Moors is indicated individually due to missing precise geolocation data.

locale using the National Forestry Inventory (NFI) 2019 survey data shapefile (Forestry Commission, 2019). The NFI data woodland map covers "all forest and woodland area over 0.5 hectare with a minimum of 20% canopy cover, or the potential to achieve it, and

a minimum width of 20 metres" (where "potential to achieve" cover includes land areas of new planting or restocking after felling, etc.;Forestry Commission, 2019). To begin, we imported the NFI shapefile into QGIS 3.26.3 (QGIS Development Team, 2022) and used the dissolve function to produce two sets of woodland polygons for each sampling site, one a more conservative estimate of what constitutes a fragment, and one taking a less strict approach to fragment delineation (termed "broad" fragments). First for the "conservative" woodland fragments we included only the NFI polygons strictly adjacent to one another (starting from a nest position). Secondly, for each nest we also defined a "broad" woodland fragment that also included polygons separated from others by a narrow linear feature such as a track, minor road or minor waterway. The "conservative" fragments were selected to represent a very restrictive estimate of the woodland habitat that the sampled population of red wood ants may occupy, and the "broad" fragments were selected to include all woodland is connected without a substantial barrier to wood ant dispersal. Our broad fragmentation categorisation sometimes united multiple nearby conservative fragments into a single larger combined fragment, therefore our analysis at the broad fragmentation level contains fewer fragments. Where two wood ant species were present in a fragment population, we separated into two species-based populations for that fragment using the results of the ddRAD co-ancestry analysis (see Section 3.4.2) to assign nest species.

We recorded the area in hectares (ha) for each woodland fragment as an initial habitat fragmentation measure, because larger woodland fragments are more likely to be representative of an area of continuous woodland than a highly fragmented area of small woodland patches.

As an additional way of capturing landscape-level fragmentation, we also calculated the percentage woodland cover in a "buffer zone" radius 1856m around each nest using the Simple Features package (Pebesma, 2018) in R (R Core Team, 2022). This radius

was chosen based on the average conservative fragment area. We calculated percentage woodland cover in a larger buffer zone radius of 2625m (based on 2 × average conservative fragment area) to see how the two differed, however, did not include this in the modelling (see discussion of predictor variable correlation below).

Nests were assigned population IDs corresponding to the broad and conservative fragments in which they were located, and then for each population we recorded : (i) number of nest samples per population, (ii) the species of the samples (based on RAD coancestry), (iii) the possible wood ant species present in this woodland, (iv) the area of the woodland fragment (in hectares), (v) the average % of woodland cover within the buffer zone across all nests in that population and (vi) the average latitude of samples in the population. Latitude was included as a covariate in the models because the 3 species included have restricted latitudinal ranges, and including this as a covariate helps distinguish species and latitudinal effects on diversity.

4.3.3 Genetic diversity statistics

To characterise genetic diversity within each broad and conservative woodland fragment, we calculated four population level diversity statistics using `populations` program in STACKS (Catchen et al., 2013): nucleotide diversity, the observed and expected heterozygosity, and the inbreeding coefficient. STACKS `populations` calculates these statistics for each input population at each variable SNP at a locus as described in Catchen et al. (2013) Supplementary Materials. The observed heterozygosity (H_{O}) is the proportion of genotypes that are heterozygotes, and expected heterozygosity under HWE (H_{E}) is $2pq$. Due to the nature of RAD data the nucleotide diversity (π) is estimated following the method of Hohenlohe et al. (2010) (see Equations 2 and 3) and is described as equivalent to expected heterozygosity. We included both π and H_{E} to operate as controls for each other as we would expect them to have very similar results. Finally, STACKS `populations` calculates the inbreeding coefficient (F_{IS}) as $(\pi - H)/\pi$ (derived from

Hartl and Clark (2007) equation 6.4 p. 264; see STACKS Manual).

Sample composition for the two types of fragment definition (broad or conservative) differ (see Table 4.1) so we calculated the statistics separately for each and we included parameter `-r 0.65` *i.e.* retaining only loci present in 65% of the population across both sets of analyses (Catchen et al., 2013).

4.3.4 Modelling

4.3.4.1 Model variables

We modelled how well species, latitude, fragment size and woodland cover in buffer zone predict the level of genetic diversity across populations of *Formica* species from across the British Isles. These predictor variables were included as: the species of samples within a population (categorical variable with three levels), the average latitude of samples within a populations, the area of the woodland fragments within which population samples were taken, and the mean woodland cover in the buffer zones around nests within a population (the latter three are numerical variables). The species distribution across the British Isles is strongly influenced by latitude, and as such we included the interaction between population species and average latitude in the model.

4.3.4.2 Correlation amongst predictor variables

We generated the woodland cover around nest values for two sizes of buffer (1856m and 2625m) and initially intended to include both values in our modelling. To test whether to model both we first plotted the two together to check for visible patterns correlation, then calculated the Spearman rank correlation coefficient (Spearman's ρ; Spearman, 1904). We found the variables to be highly correlated ($\rho = 0.835$) and decided to include only the first buffer size based on the conservative woodland fragment area (c. 1856m radius around nests).

We also tested for collinearity between the buffer zone woodland cover and the woodland fragment area (broad and conservative) to establish whether it was more appropriate to model them separately. We found the buffer zone woodland cover was correlated with both the broad and conservative fragments (ρ of 0.606 and 0.580, respectively). We then calculated the Variance Inflation Factors (VIF; John and Georges, 1992) for weighted linear models using H_O as the response variable, with example model (weighted by number of samples in a population): $H_O \sim$ `Species + Woodland fragment area + Buffer zone woodland cover + Population latitude`. VIF values across all predictors and both woodland fragment area characterisations were below 1.6, which is much lower than the 'Rule of Thumb' for correcting models due to collinearity (usually 10, or 4 if being stringent). As such, we included both predictor variables in the same models.

4.3.4.3 Multiple regression modelling

We modelled H_O, H_E, π and F_{IS} separately for our broad and conservative woodland fragments using general linear models (Faraway, 2004). They were modelled separately as the sample composition for the different type of fragments were slightly different, and this affected both the diversity statistics and the values of the predictor variables. The number of nests included in each population varied between 1 and 8 in conservative fragments and 1 and 11 in broad fragments, as such we weighted the models by population sample size. We assessed the conditions of applicability and model fit using the standard diagnostic plots *i.e.* residuals vs fitted, normal Q-Q, and scale-location plots. We then employed an AIC-based (Faraway, 2016) approach to select the variables to include in our final model, removing variables from the model in-turn to find the model with the lowest AIC value. Where applicable, we excluded interactions between variables (*i.e.* `species:average latitude`) before the variables themselves. If multiple models had AIC values that differed by <2 we chose the simplest model. We used F-tests to

assess the significance of explanatory variables in the final model (using cutoff $p \leqslant 0.05$). Finally, we tested differences between focal groups using contrasts, adjusting the p values for multiple comparisons using Holm's method (Holm, 1979).

4.4 Results

As there were more woodland fragments when we used our "conservative" characterisation than when we used our "broad" characterisation, the number of nests per fragment and the number of fragments per species presence category differ between the two characterisations of fragmentation (Table 4.1). Our five species presence categories comprise: three single-species designations for *F. aquilonia*–only ("aq"), *F. lugubris*–only ("lu"), and *F. rufa*–only ("ru"), and; two multispecies designations where two species occurred in the same woodland fragment *i.e. F. aquilonia* & *F. lugubris* ("aq_lu") and *F. lugubris* & *F. rufa* ("lu_ru").

		aq	aq_lu	lu	lu_ru	ru	Total
Conservative	*F. aquilonia*	11	6	-	-	-	17
	F. lugubris	-	6	21	3	-	30
	F. rufa	-	-	-	3	9	12
	Total	11	12	21	6	9	59
Broad	*F. aquilonia*	10	6	-	-	-	16
	F. lugubris	-	6	16	2	-	24
	F. rufa	-	-	-	2	8	10
	Total	10	12	16	4	8	50

Table 4.1: Counts of population-level species and woodland fragment (*i.e.* location-level) species based on the conservative and broad characterisation of woodland fragments. See location-level species category designations in-text.

As outlined above, each population (as defined for genetic diversity measures) only included samples of a single species, which were assigned based on a nest's position in the ddRAD estimated coancestry analysis (see Fig. 3.4). Table 4.1 shows the number of populations per species included in our analyses, and how many of these populations fall into each species presence categories for the conservative and broader fragment characterisations.

Due to the very low counts in some categories (e.g. "lu_ru") we decided to only incorporate population species as a variable in the statistical models, and not whether the population is in a single-species or multispecies site, however, we still plot these data in full in Panel A of Figures 4.2 - 4.9.

4.4.1 Population genetic diversity summary

	Species	H_{E}	H_{O}	π	F_{IS}
Conservative	*F. aquilonia*	0.0703	0.0913	0.0974	0.0108
	F. lugubris	0.0762	0.118	0.118	0.000648
	F. rufa	0.0620	0.0986	0.101	0.00323
Broad	*F. aquilonia*	0.0735	0.0913	0.0981	0.0121
	F. lugubris	0.0830	0.121	0.122	0.00250
	F. rufa	0.0679	0.101	0.104	0.00584

Table 4.2: Mean population-level genetic diversity statistics by species of samples within a population, including both the conservative and broad characterisations of woodland fragments.

We found higher mean species observed heterozgosity (the proportion of heterozygous loci averaged across all loci; H_{O}) was higher than the mean species expected heterozygosity (the proportion of loci expected to be heterozygous under HWE; H_{E}) for all three

species irrespective of woodland fragment characterisation (Table 4.2). This suggests a divergence from the assumptions of Hardy-Weinberg Equilibrium (HWE), and the reasons for this are not clear. The inbreeding coefficient F_{IS} is low for all three species, but particularly so in *F. lugubris*. Due to the way STACKS `populations` estimates nucleotide diversity (π), we expected the results for π and H_E to be similar, however, they were much more comparable to H_O.

4.4.2 Response variable: Observed heterozygosity

The population-level H_O follows the same pattern in both sets of woodland characterisation (Figs 4.2 and 4.3) with *F. lugubris* the highest overall (grey boxes on boxplots), followed by *F. rufa* (red boxes) and then *F. aquilonia* (blue boxes). There is little discernible correlation between H_O and fragment area (Figs 4.2 and 4.3 C) or percentage woodland cover within buffer zones (Figs 4.2 and 4.3 D). Observed heterozygosity correlates with the average nest latitude within a woodland fragment, and the direction of this relationship differs between species (Figs 4.2 and 4.3 B).

4.4.2.1 Conservatively characterised woodland fragments

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0166	40.2353	<0.0001
Lat_avg	1	0.0003	1.3990	0.2422
Species:Lat_avg	2	0.0027	6.4522	0.0031
Residuals	53	0.0109	NA	NA

Table 4.3: ANOVA results for the final model for observed heterozygosity in conservative woodland fragments. The variables listed were included in the final model. Average latitude is included despite its non-significant contribution as a main effect because we retained variables contributing to a significant interaction

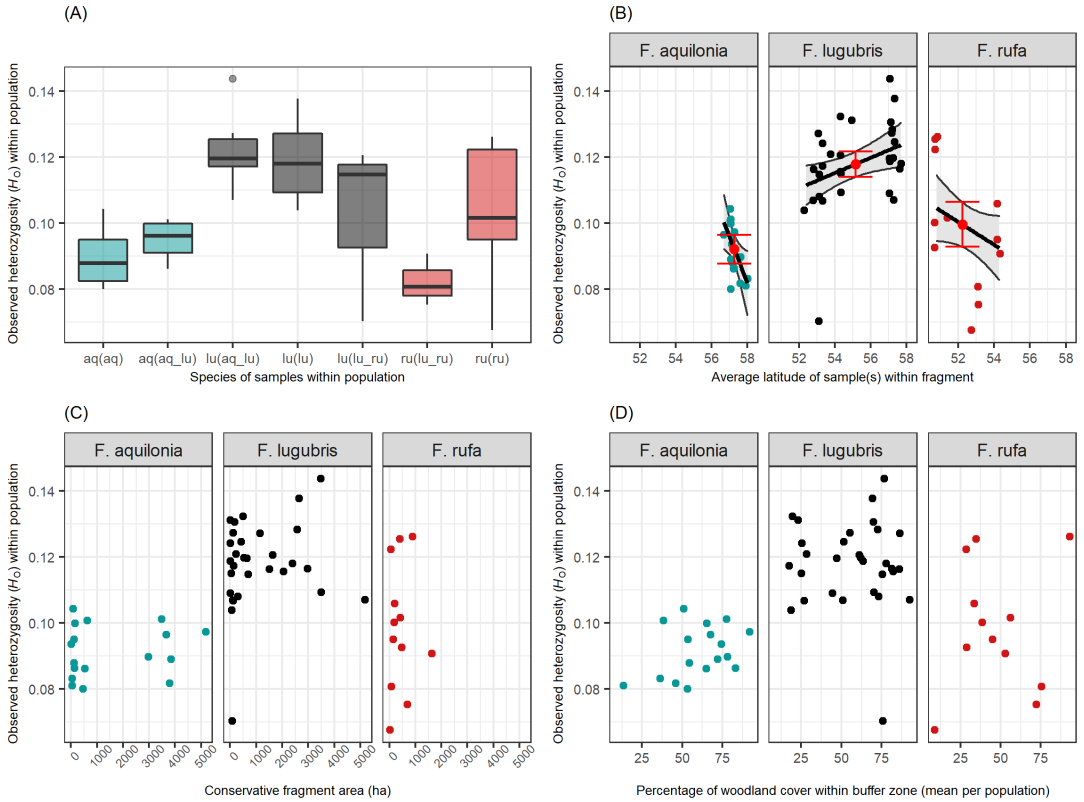

Figure 4.2: Observed heterozygosity (H_O; y axis) plotted against the four predictor variables of the model for conservatively characterised woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing H_O within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of H_O and the average latitude of samples within a woodland fragment, faceted by population species. The average latitude was included in the final model and fit is shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean H_O at the mean average latitude as predicted by the model. See in-text for values. (C) Scatterplot of H_O and the woodland fragment total area in hectares, faceted by population species. (D) Scatterplot of H_O and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

Population species ($p<0.0001$) and its interaction with the average latitude of a population ($p=0.0031$) both significantly predicted H_O (Table 4.3). The conservative fragment area and buffer zone woodland cover were removed during model selection. The effect

of average latitude on H_O was significantly different between *F. lugubris* and *F. rufa* (p=0.0291) and between *F. lugubris* and *F. aquilonia* (p=0.027). While the average H_O of *F. lugubris* increased by 0.0023 (95% CI: 0.0004 – 0.0042) with every degree of latitude, we saw a negative trend for *F. rufa* (-0.0035, 95% CI: -0.0076 – 0.0007). *Formica aquilonia* showed a similar negative pattern (-0.0142, 95% CI: -0.0263 – -0.0021), which did not significantly differ from *F. rufa* (p=0.0968).

To realistically compare the mean H_O of the species, we compared the mean values predicted for the average latitude of the samples within a population (marked with red points on Fig. 4.2 B). We see a similar pattern as above, where mean species H_O differed significantly in species pairs *F. lugubris* & *F. rufa* and *F. lugubris* & *F. aquilonia* (both with p<0.0001), but not between *F. aquilonia* and *F. rufa* (p=0.0671). *Formica lugubris* had the highest H_O of 0.118 (95% CI: 0.114 – 0.122) among all three species, followed by *F. rufa* (0.1, 95% CI: 0.093 – 0.106) and *F. aquilonia* (0.092, 95% CI: 0.088 – 0.096).

4.4.2.2 Broader woodland fragment characterisation

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0169	37.3518	<0.0001
Lat_avg	1	0.0002	1.0874	0.3027
Species:Lat_avg	2	0.0033	7.2943	0.0018
Residuals	44	0.0100	NA	NA

Table 4.4: ANOVA results for the final model for observed heterozygosity in broader woodland fragments. The variables listed were included in the final model. Average latitude is included despite its non-significant contribution as a main effect, because we retained variables contributing to a significant interaction.

The results for modelling H_O based on broader fragment characterisation follow the

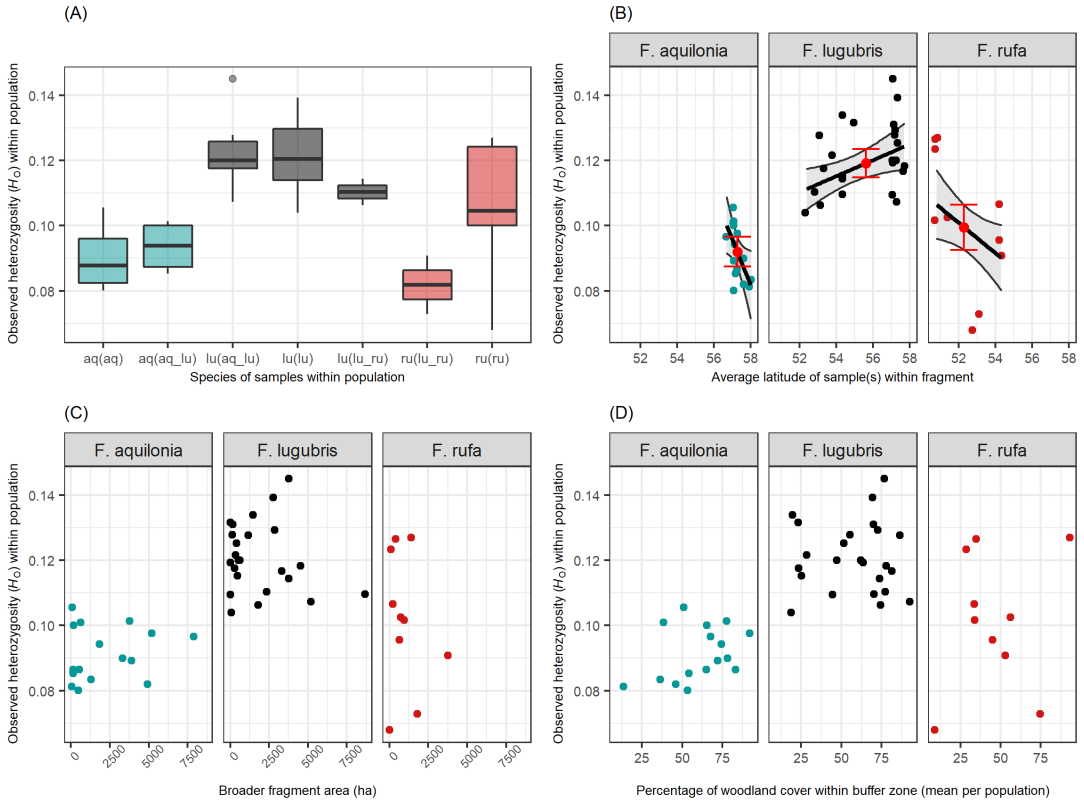

Figure 4.3: Observed heterozygosity (H_O; y axis) plotted against the four predictor variables of the model for broader characterisation of woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing H_O within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of H_O and the average latitude of samples within a woodland fragment, faceted by population species. The average latitude was included in the final model and fit is shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean H_O at the mean average latitude as predicted by the model. See in-text for values. ((C) Scatterplot of H_O and the woodland fragment total area in hectares, faceted by population species. (D) Scatterplot of H_O and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

same pattern as those that modelled using the conservative fragment populations. Again, fragment area and buffer woodland cover were excluded from the model, and both species and the species:average latitude interaction had a significant effect on H_O ($p<0.0001$

and p=0.0018, respectively) (Table 4.4). We saw a positive trend in the effect of average latitude on H_O in *F. lugubris* (0.0025 95% CI: 0.0005 – 0.0044, we saw negative trends in both *F. rufa* (-0.0046, 95% CI: -0.009 – -0.0003) and *F. aquilonia* (-0.014, 95% CI: -0.0267 – -0.0013). The effect was significantly different between *F. lugubris* and both *F. rufa* (p=0.014) and *F. aquilonia* (p=0.0269), but not between the two negative trends of *F. rufa* and *F. aquilonia* (p=0.1684).

F. lugubris had the highest mean species H_O predicated at the average species-wide latitude (red points on Fig. 4.3 B) at 0.119 (95% CI: 0.115 – 0.123), followed by *F. rufa* (0.099, 95% CI: 0.093 – 0.106) and *F. aquilonia* (0.092, 95% CI: 0.087 – 0.097). This once again followed the same patterns as for the conservative fragment modelling: the difference between *F. lugubris* and both other species had a $p<0.0001$, whereas the comparison between mean H_O of *F. aquilonia* and *F. rufa* was not significant (p=0.0749).

4.4.3 Response variable: Expected heterozygosity

Unlike H_O, the patterns in the population-level H_E differ between our two fragment characterisations.

For the conservatively characterised fragments, the difference between H_E per species (Fig. 4.4 A) is not as pronounced as that for H_O (though still has a significant effect in the model). There is a correlation between fragment area and H_E (Fig. 4.4 C), but none between either average latitude or percentage bufferzone woodland cover and H_E (Fig. 4.4 B and D, respectively).

The more broadly characterised fragment populations show a very different picture. Figure 4.5 A shows a clearer difference among the species. The complete opposite variables correlate: average latitude and percentage bufferzone woodland both correlate with H_E, and fragment area does not. In contrast to the latitude correlation in H_O, here we see a negative trend in all three species rather than just *F. aquilonia* and *F. rufa*.

4.4.3.1 Conservatively characterised woodland fragments

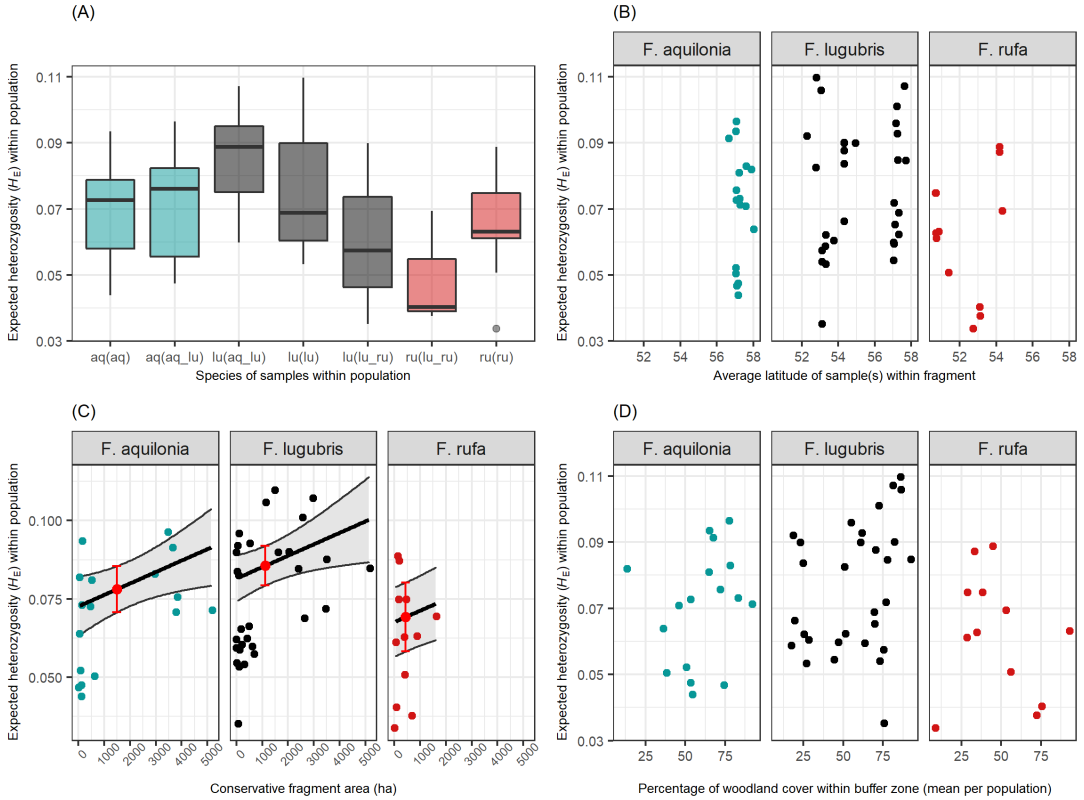

Figure 4.4: Expected heterozygosity (H_E; y axis) plotted against the four predictor variables of the model for conservatively characterised woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing H_O within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of H_E and the average latitude of samples within a woodland fragment, faceted by population species. (C) Scatterplot of H_E and the woodland fragment total area in hectares, faceted by population species. The fragment area was found to be significant in the model (p=0.0212), and the modelfits are shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean H_O at the mean fragment area as predicted by the model. See in-text for values. (D) Scatterplot of H_E and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0042	3.6351	0.0329
Conservative_frag_area_hec	1	0.0032	5.6271	0.0212
Residuals	55	0.0316	NA	NA

Table 4.5: ANOVA results for the final model for expected heterozygosity in conservative woodland fragments. The variables are named on the left, here both species and fragment area have a significant effect.

All predictors except population species and conservative fragment area were excluded from the model during selection. Population species (p=0.0329) and conservative fragment area (p=0.0212; see model fit line on Fig. 4.4 C) had a significant effect on H_E (Table 4.5).

We compared the mean H_E per species predicted by the model, using the value at the mean conservative fragment area across populations (red points on Fig. 4.4 C). Similarly to H_O, we found *F. lugubris* had the highest mean H_E with a value of 0.086 (95% CI: 0.079 – 0.092), which differed significantly (p=0.0375) from *F. rufa*, which had the lowest H_E at 0.069 (95% CI: 0.058 – 0.08). *Formica aquilonia* had an intermediate predicted mean H_E value with 0.078 (95% CI: 0.071 – 0.085), which did not significantly differ from either other species (both p=0.2533).

4.4.3.2 Broader woodland fragment characterisation

Only broader fragment area and the interaction between species and latitude were excluded from this model. The effect of the species of populations (p=0.005) and the average latitude (p=0.0383) on H_E were both significant, whilst that of mean buffer zone woodland cover did not quite meet the 5% significance threshold (p=0.081) (Table 4.6 and corresponding model fits on Fig. 4.5 B and D).

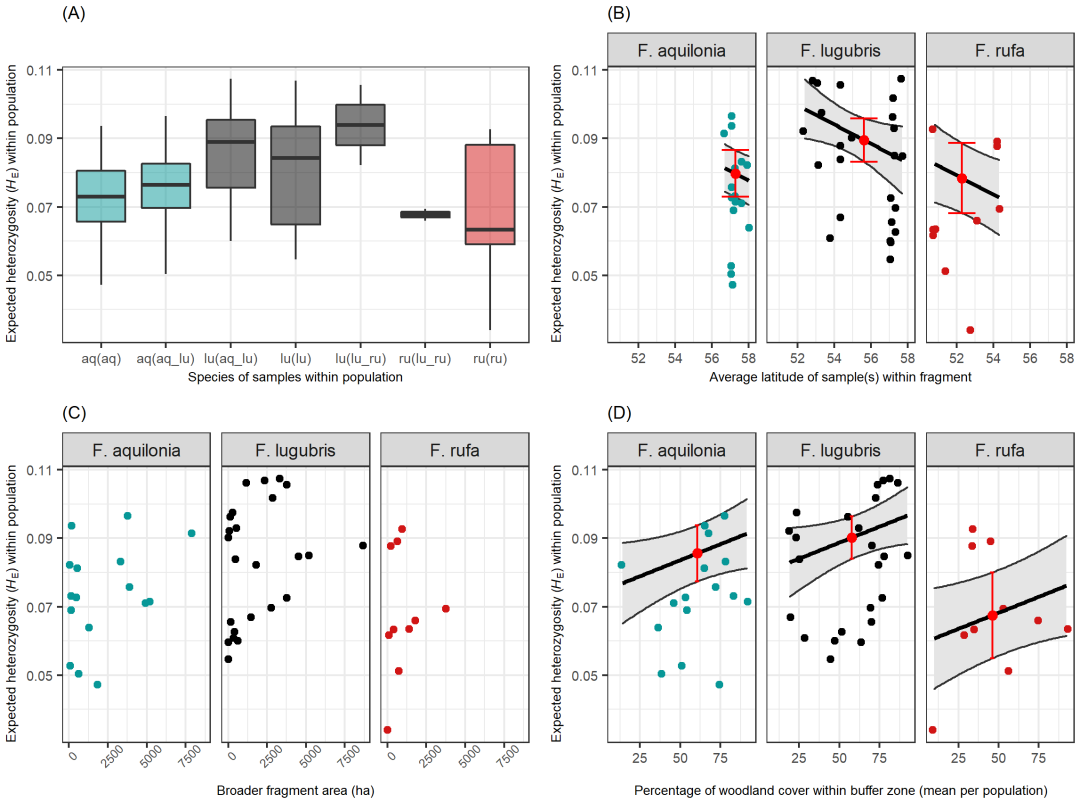

Figure 4.5: Expected heterozygosity (H_E; y axis) plotted against the four predictor variables of the model for broader characterisation of woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing H_O within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of H_E and the average latitude of samples within a woodland fragment, faceted by population species. The average latitude was found to be significant in the model (p=0.0383), and the modelfits are shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean H_O at the mean average latitude as predicted by the model. See in-text for values. (C) Scatterplot of H_E and the woodland fragment total area in hectares, faceted by population species. See in-text for values. (D) Scatterplot of H_E and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species. The mean percentage woodland cover was included in the final model, and the modelfits are shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean H_O at the mean fragment area as predicted by the model. See in-text for values.

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0059	5.9835	0.0050
Mean_percent_woodland	1	0.0016	3.1863	0.0810
Lat_avg	1	0.0022	4.5534	0.0383
Residuals	45	0.0221	NA	NA

Table 4.6: ANOVA results for the final model for expected heterozygosity in broader woodland fragments. The variables are named on the left, here both species and average latitude have a significant effect.

Firstly, we compared the mean H_E per species predicted by the model at the mean average population latitude (red points on Fig. 4.5 B). *F. lugubris* had the highest predicated mean H_E (0.09, 95% CI: 0.083 – 0.096), followed by *F. aquilonia* (0.08, 95% CI: 0.073 – 0.087) and *F. rufa* (0.078, 95% CI: 0.068 – 0.089). The difference between these predicted mean H_E values was not significant for any species pairs.

We also compared the predicted mean H_E per species at the mean per-population buffer-zone woodland cover (red points on Fig. 4.5 D). The results followed the same pattern as above in terms of species order, with *F. lugubris* highest (0.09, 95% CI: 0.084 – 0.096) followed quite closely by *F. aquilonia* (0.086, 95% CI: 0.077 – 0.094) and then *F. rufa* (0.068, 95% CI: 0.055 – 0.08). The difference between these predicted mean species H_E values was found to be significant between *F. lugubris* and *F. rufa* (p=0.0029). However, these differences were not significant between *F. lugubris* and *F. aquilonia* (p=0.4066) or between *F. aquilonia* and *F. rufa* (p=0.0802), though the latter is closer to the *a priori* chosen 5% threshold.

4.4.4 Response variable: Nucleotide diversity

The nucleotide diversity (π) follows a similar pattern of species and latitude effects to those of H_O (see Figs 4.6 and 4.7. In both woodland fragment characterisations *F. lugubris* shows the highest diversity, followed by *F. rufa* and *F. aquilonia*. Average latitude also correlates with π in both, whilst fragment area and percentage woodland cover within bufferzones do not.

4.4.4.1 Conservatively characterised woodland fragments

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0094	23.1152	<0.0001
Lat_avg	1	0.0008	3.7695	0.0575
Species:Lat_avg	2	0.0022	5.3987	0.0073
Residuals	53	0.0108	NA	NA

Table 4.7: ANOVA results for the final model for nucleotide diversity in conservative woodland fragments. The variables listed were included in the final model. Average latitude is included despite its non-significant contribution as a main effect, because we retained variables contributing to a significant interaction.

Population species ($p<0.0001$) and its interaction with the average latitude of a population ($p=0.0073$) both correlated significantly with nucleotide diversity (π; Table 4.7 and see Fig. 4.6 B for the plotted model fit). In this model, the significance of the effect of average latitude itself (*i.e.* not in interaction with species) on π was only slightly above the *a priori* chosen 5% value at $p=0.0575$. The conservative fragment area and buffer zone woodland cover were removed during model selection. The effect of average latitude on π was significantly different between *F. lugubris* and *F. rufa* ($p=0.0496$) and between *F. lugubris* and *F. aquilonia* ($p=0.049$). While the average π of *F. lugubris* increased by 0.0028 (95% CI: 0.0009 – 0.0047) with every degree of latitude, we saw a

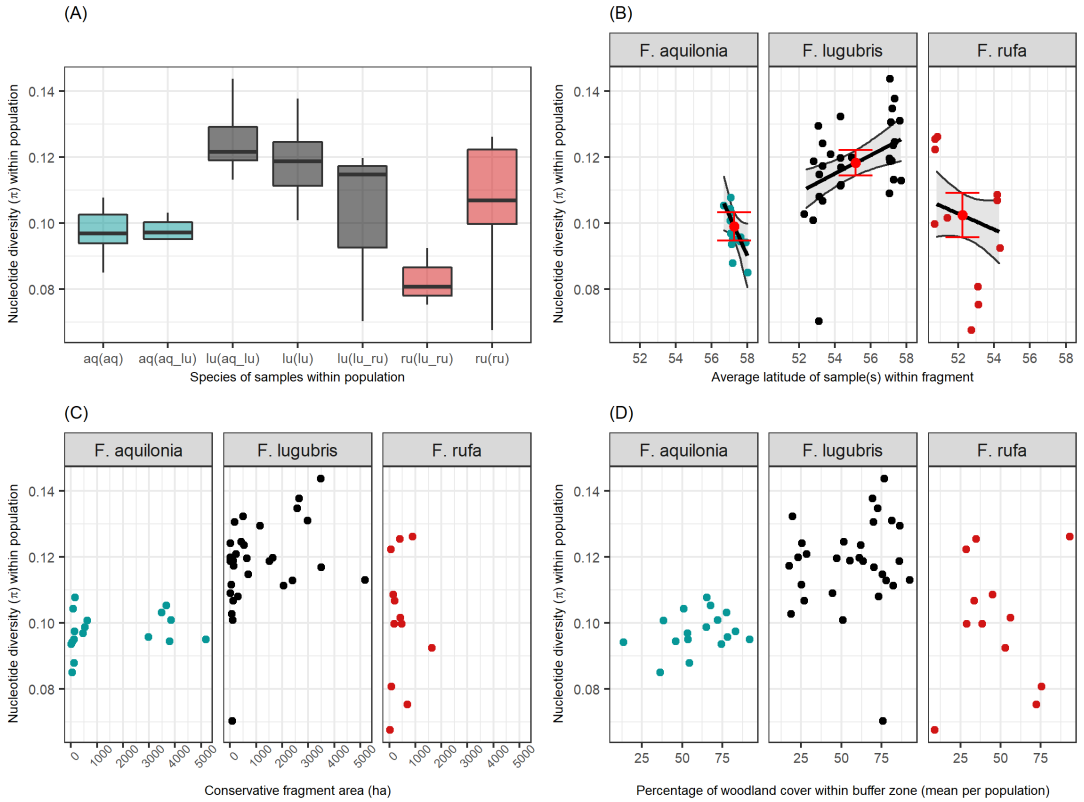

Figure 4.6: Nucleotide diversity (π; y axis) plotted against the four predictor variables of the model for conservatively characterised woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing π within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of π and the average latitude of samples within a woodland fragment, faceted by population species. The average latitude was included in the final model and fit is shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean π at the mean average latitude as predicted by the model. See in-text for values. (C) Scatterplot of π and the woodland fragment total area in hectares, faceted by population species. (D) Scatterplot of π and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

negative trend for *F. rufa* (-0.0024, 95% CI: -0.0065 – 0.0017). *F. aquilonia* showed a similar negative pattern (-0.0122, 95% CI: -0.0242 – -0.0002), which did not significantly differ from *F. rufa* (p=0.128).

The mean species π predicted at the mean average latitude (red points on Figure 4.6 B) differed significantly in species pairs *F. lugubris* and *F. rufa* (p<0.0001) and *F. lugubris* and *F. aquilonia* (p=0.0003), but not between *F. aquilonia* and *F. rufa* (p=0.394). *F. lugubris* had the highest π of 0.118 (95% CI: 0.114 – 0.122) among all three species, followed by *F. rufa* (0.102, 95% CI: 0.096 – 0.109) and *F. aquilonia* (0.099, 95% CI: 0.095 – 0.103).

4.4.4.2 Broader woodland fragment characterisation

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0103	27.4750	<0.0001
Lat_avg	1	0.0005	2.4013	0.1284
Species:Lat_avg	2	0.0026	7.0657	0.0022
Residuals	44	0.0082	NA	NA

Table 4.8: ANOVA results for the final model for nucleotide diversity in broader woodland fragments. The variables listed were included in the final model. Average latitude is included despite its non-significant contribution as a main effect, because we retained variables contributing to a significant interaction.

Population species (p<0.0001) and its interaction with the average latitude of a population (p=0.0022) both significantly predicted nucleotide diversity (π; Table 4.8 and see Fig. 4.7 B for the plotted model fit). In this model, the significance of the effect of average latitude itself (*i.e.* not in interaction with species) on π was not significant (p=0.1284). Similar to the previous π model, the broader fragment area and buffer zone woodland cover were removed during model selection. The effect of average latitude on π was significantly different between *F. lugubris* and *F. rufa* (p=0.016) and between *F. lugubris* and *F. aquilonia* (p=0.0296). While the average π of *F. lugubris* increased by 0.0026 (95% CI: 0.0008 – 0.0044) with every degree of latitude, we saw a negative trend

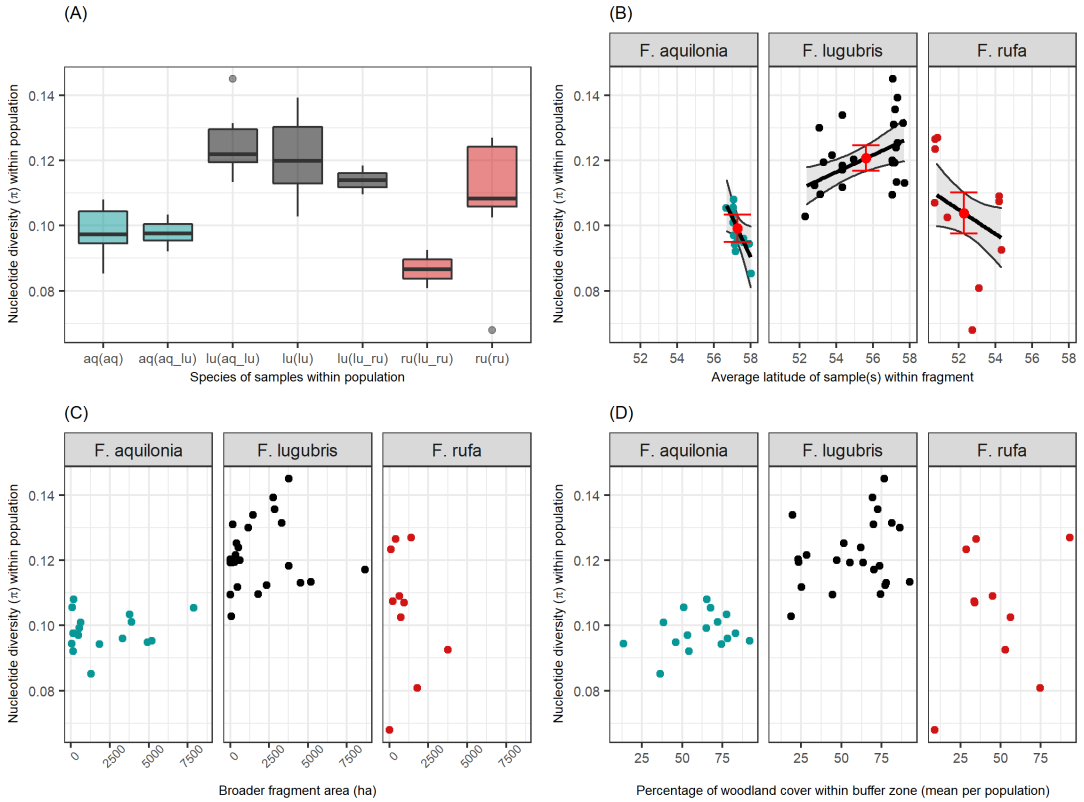

Figure 4.7: Nucleotide diversity (π; y axis) plotted against the four predictor variables of the model for broader characterisation of woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing π within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of π and the average latitude of samples within a woodland fragment, faceted by population species. The average latitude was included in the final model and fit is shown by the black lines. 95% CI shown by the light grey areas. The red point denotes the species mean π at the mean average latitude as predicted by the model. See in-text for values. (C) Scatterplot of π and the woodland fragment total area in hectares, faceted by population species. (D) Scatterplot of π and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

for *F. rufa* (-0.0037, 95% CI: -0.0077 – 0.0002). *F. aquilonia* showed a similar negative pattern (-0.0121, 95% CI: -0.0237 – -0.0006), which did not significantly differ from *F. rufa* (p=0.1741).

The mean species π predicted at the mean average population latitude (red points on Fig. 4.7 B) differed significantly in species pairs *F. lugubris* and *F. rufa* (p<0.0001) and *F. lugubris* and *F. aquilonia* (p<0.0001), but not between *F. aquilonia* and *F. rufa* (p=0.2212). *F. lugubris* had the highest π of 0.121 (95% CI: 0.117 – 0.125), followed by *F. rufa* (0.104, 95% CI: 0.097 – 0.11) and *F. aquilonia* (0.099, 95% CI: 0.095 – 0.103).

4.4.5 Response variable: Inbreeding coefficient

The inbreeding coefficient (F_{IS}) values across our populations are very low. As is visible in both the conservative and broader fragment data (Figs 4.8 and 4.9, respectively), a substantial number of F_{IS}values are below zero, indicating that observed heterozygosity is higher than expected heterozygosity under HWE and there is an excess of heterozygotes at these sites.

The only variable that had any effect was species, discussed in the sections below.

4.4.5.1 Conservatively characterised woodland fragments

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0033	9.2658	0.0003
Residuals	56	0.0101	NA	NA

Table 4.9: ANOVA results for the final model for the inbreeding coefficient in conservative woodland fragments. The variable is named on the left. Only species had a significant effect.

Population species was the only predictor variable that a significant effect (p=0.0003) on the inbreeding coefficient (F_{IS}) in our model (Table 4.9). All the other variables were excluded during model selection. As such, the only comparison we can make is mean F_{IS} per species (not plotted on Fig. 4.8. In contrast to other response variables,

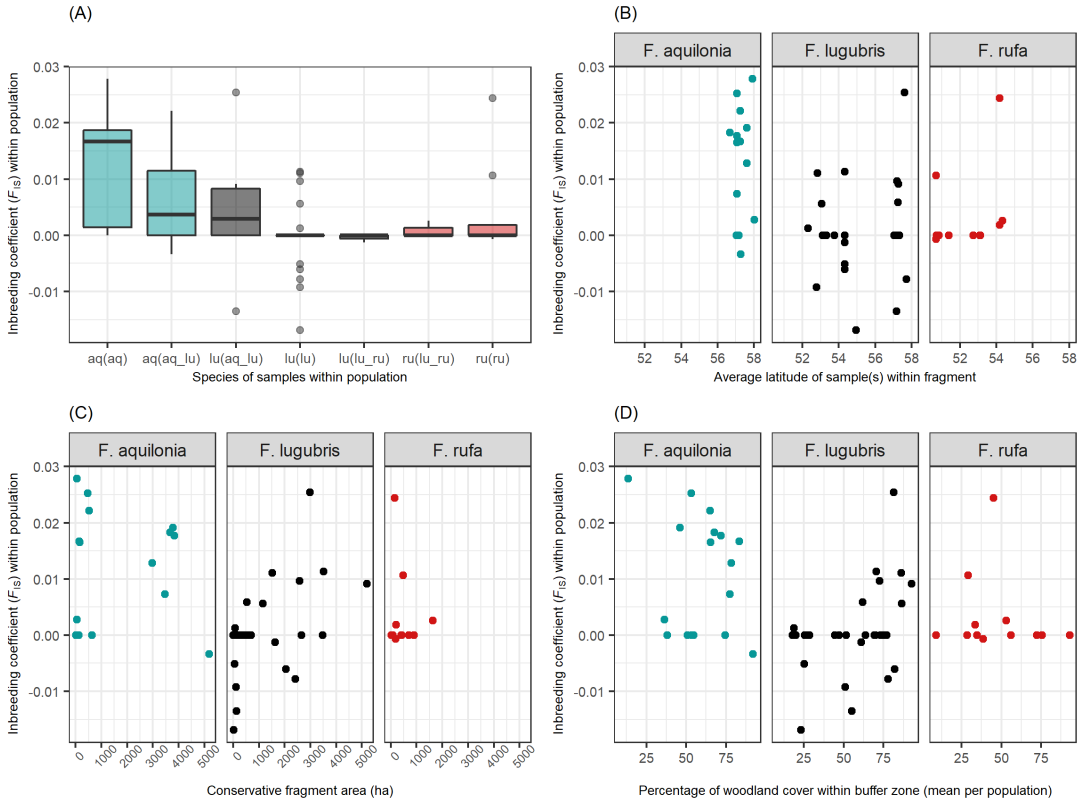

Figure 4.8: Inbreeding coefficient (F_{IS}; y axis) plotted against the four predictor variables of the model for conservatively characterised woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing F_{IS} within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of F_{IS} and the average latitude of samples within a woodland fragment, faceted by population species. (C) Scatterplot of F_{IS} and the woodland fragment total area in hectares, faceted by population species. (D) Scatterplot of F_{IS} and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

F. aquilonia has the highest predicted mean F_{IS} (0.013, 95% CI: 0.009 – 0.017), which differs significantly (p=0.0002) from that of *F. lugubris* (0.002, 95% CI: -0.002 – 0.005). *Formica rufa* has a slightly higher predicted F_{IS} (0.005, 95% CI: -0.001 – 0.012) than *F. lugubris*, but this difference is not significant (p=0.2977). The difference between

predicted mean F_{IS} in *F. rufa* and *F. aquilonia* is not significant either (p=0.0784), though is much closer to the *a priori* chosen 5% value.

4.4.5.2 Broader woodland fragment characterisation

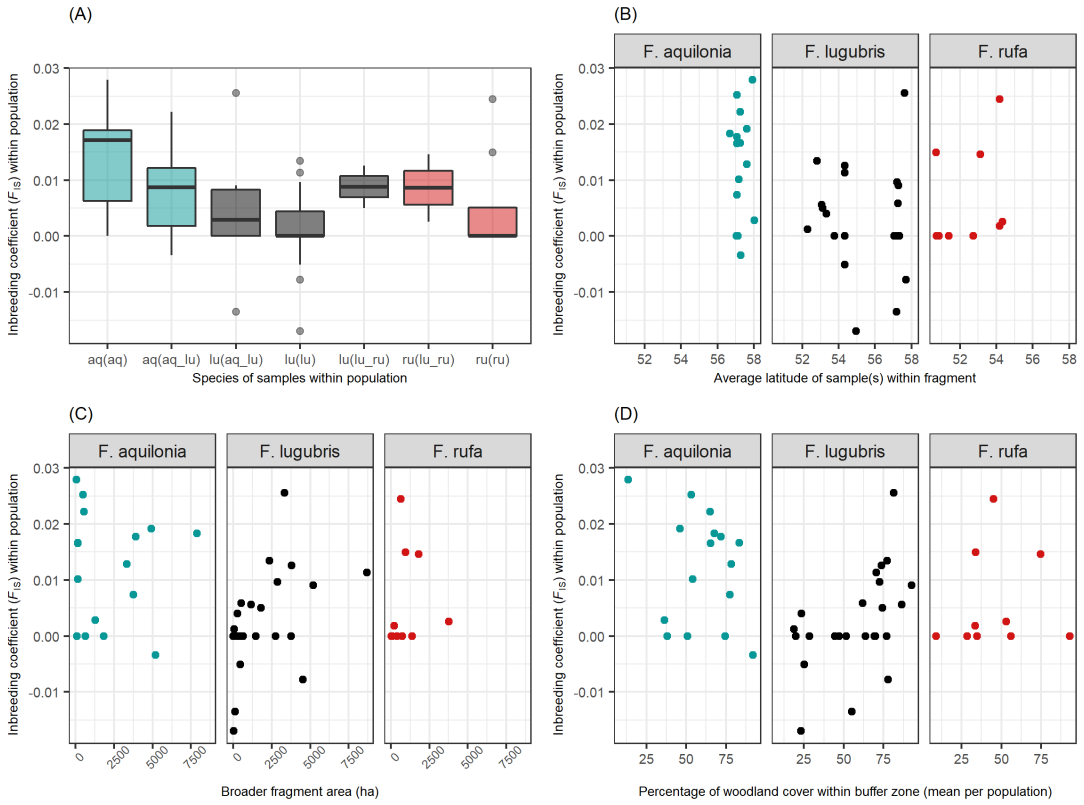

Figure 4.9: Inbreeding coefficient (F_{IS}; y axis) plotted against the four predictor variables of the model for the broader characterisation of woodland fragment populations. Population species is denoted by colour (blue for *F. aquilonia*, black/grey for *F. lugubris*, and red for *F. rufa*) and/or labelled facets. (A) Boxplot showing F_{IS} within each species, separated into location-level species categories (see category descriptions in-text). Labels follow the format [population species]([location-level species]). (B) Scatterplot of F_{IS} and the average latitude of samples within a woodland fragment, faceted by population species. (C) Scatterplot of F_{IS} and the woodland fragment total area in hectares, faceted by population species. (D) Scatterplot of F_{IS} and the mean percentage woodland cover within a 1855m buffer around the nests within each population, faceted by population species.

	Df	Sum Sq	F value	Pr(>F)
Species	2	0.0018	4.186	0.0212
Residuals	47	0.0102	NA	NA

Table 4.10: ANOVA results for the final model for the inbreeding coefficient in broader woodland fragments. The variable is named on the left. Only species had a significant effect.

Population species was the only predictor variable that a significant effect (p=0.0212) on the inbreeding coefficient (F_{IS}) in our model, echoing the results of the first model using the conservatively characterised woodland populations (Table 4.10). All other variables were excluded during model selection. As such, the only comparison we can make is mean F_{IS} per species (not plotted on Fig. 4.9). In contrast to other response variables, *F. aquilonia* has the highest predicted mean F_{IS} (0.014, 95% CI: 0.009 – 0.018), which differs significantly (p=0.0174) from that of *F. lugubris* (0.005, 95% CI: 0.001 – 0.009). *Formica rufa* has a slightly higher predicted F_{IS} (0.009, 95% CI: 0.003 – 0.016) than *F. lugubris*, but this difference is not significant (p=0.5544). The difference between predicted mean F_{IS} in *F. rufa* and *F. aquilonia* is not significant either (again, p=0.5544).

4.4.6 Summary of results

Overall, the results from our models were quite consistent in terms of the predictor variables we found to have a significant effect (Table 4.11). With the notable exception of H_E in conservatively characterised woodland fragments, our habitat fragmentation measures were not able to predict the genetic diversity in our samples.

	H_O	H_E	π	F_{IS}
Cons.	**Species**	**Species**	**Species**	**Species**
	Latitude	**Woodland frag. area**	Latitude	-
	Species:Latitude	-	**Species:Latitude**	-
Broad	**Species**	**Species**	**Species**	**Species**
	Latitude	**Latitude**	Latitude	-
	Species:Latitude	% Woodland cover	**Species:Latitude**	-

Table 4.11: Summary of the variables included in each final model. Predictors with a significant effect are in bold.

We would expect a concordance in the results for H_E and π as they are closely related measures of genetic diversity, however, their results differ. The data and modelling results for π more closely resemble that for H_O.

4.5 Discussion

As summarised in Table 4.11, we only found weak evidence of fragmentation affecting genetic diversity in our models. Our data show that the wood ant species differ in overall genetic diversity, but that potential for gene flow between species does not result in higher genetic diversity within our samples (*i.e.* Panel A in Figs 4.2 to 4.9). This latter result may be due to the very low sample sizes for the multi-species sites (see Table 4.1). Species was the only predictor of genetic diversity that was significant in every model, with a general pattern of higher diversity in *F. lugubris*.

We hypothesised that the highly fragmented woodland habitats of British *F. rufa* group wood ants would results in reduced observed heterozygosity (H_O), expected heterozygosity (H_E) and nucleotide diversity (π), and higher inbreeding coefficient (F_{IS}) values.

In our modelling, woodland fragment area (conservatively characterised) was able to predict the expected heterozygosity (H_{E}), however, no other combination of fragmentation or genetic measure yielded significant results. We discuss some ecological and behavioural reasons for these results below, but must also consider that our models may fail to fully capture the effects of habitat fragmentation because of low statistical power due to (i) nature of woodland composition in Britain and/or (ii) our sample sizes and distribution. In Britain, our available habitats are at the extremes of woodland fragmentation, especially when compared to continuous woodlands in central and northern Europe (Seifert et al., 2010). It may be that other measures of fragmentation would be more informative, such as incorporating number of habitat patches in a given area, or edge indices e.g. edge density, mean distance within-fragment to edge and others (see Riutta et al., 2014). It may be the connectivity between the fragments that is more important than the fragment sizes themselves in the British woodland landscape, and a functional connectivity modelling approach (e.g. Watts and Handley, 2010) may be of interest (if beyond the scope of this study). Fragmented forest landscapes facilitate hybridisation between wood ant species in Central Europe (Seifert et al., 2010). While we found no clear evidence for this effect in our dataset, unfortunately, we did not have sufficient power to clearly determine whether being at a multispecies site (*i.e.* with potential for gene flow between species where they co-occur) affects genetic diversity in our populations.

The dissimilarity in results for π and H_{E} was unexpected. The calculation used by STACKS `populations` to estimate π means this value should be "equivalent to expected heterozygosity" (Hohenlohe et al., 2010), however, we did not find this reflected in our results. The reasons for this are unclear and require a clearer understanding of how STACKS is calculating these statistics, and whether they may be affected differently by factors such as the amount of missing data typical of RAD sequencing data, and the small samples for some of the populations. Future work will involve a fuller investigation

of the properties of these population genetic parameters as calculated by STACKS.

The three species responded differently to latitude for two of our diversity measures, H_O and π. The general trend was the same (e.g Fig. 4.2 B) where H_O and π decreased with latitude in both *F. aquilonia* and *F. rufa*, but increased with latitude for *F. lugubris*. *Formica lugubris* differed significantly from the other two species in the effect of latitude, but *F. aquilonia* and *F. rufa* showed similar reductions in diversity with increased latitude. One explanation for this pattern may be their different ranges. Species often show reduced genetic diversity at the limits of their ranges and be under different selection pressures (Chuang and Peterson, 2016), and both *F. aquilonia* and *F. rufa* are at the northwestern limits of their range in our sampling area (see Fig. 1.1 for species distribution maps). In addition, both are limited to smaller latitudinal ranges than *F. lugubris*.

The mean F_{IS} was low for all three species, indicating no evidence of an inbreeding depression as a result of isolated populations. Overall, *F. lugubris* has higher genetic diversity and a lower inbreeding coefficient (F_{IS}) than the other species, followed by *F. rufa* and then *F. aquilonia* (see Table 4.2). This higher diversity in *F. lugubris* is somewhat counter-intuitive, as the species has P-type population organisation in Britain, meaning it has no mechanism for long-distance dispersal, whereas *F. rufa* in Britain is thought to produce new nests by social parasitism after flight. Interestingly, concordant with our study, a comparison of microsatellite data from fragmented, isolated *F. lugubris* populations in the Peak District with a fragmented but non-isolated Finnish population found similar genetic variation between the two and no evidence of an inbreeding depression (Gyllenstrand and Seppä, 2003). Together with this previous study, our data suggest that *F. lugubris* in Britain is remarkably genetically healthy for a species with such limited capacity for non-local gene flow. In contrast Mäki-Petäys and Breen (2007) found very little genetic variation, a largely monogynous colony structure, and evidence of an

inbreeding depression in an at-risk *F. lugubris* population in the Republic of Ireland. This highlights the effect that social behaviour can have on genetic variation and potential future viability of populations. Alternatively, the very low genetic diversity in the Irish populations may be a result of a founder effect followed by very small population sizes over time (Mäki-Petäys and Breen, 2007). Scottish *F. aquilonia* populations were found to be genetically diverse (based on 10 microsatellite loci) and show no evidence of inbreeding, despite extensive habitat fragmentation over time (Vanhala et al., 2014), again consistent with our findings. Our data alongside these findings from the literature suggest British wood ants are resilient to the effects of habitat fragmentation and subsequent population isolation. This could be due to polygynous colony structures resulting in high effective population sizes despite population isolation, particularly for the highly P-type (see section Thesis Introduction) *F. aquilonia* and *F. lugubris*. Limited genetic effects of extensive habitat loss were found in another habitat specialist insect, the woodland cricket (*Nemobius sylvestris*), indicating other taxa are resilient to fragmentation (Watts et al., 2016).

In conclusion, our data suggest that a better scale for addressing this question would be a European-wide survey of genetic diversity and fragmentation, allowing for a greater range of degrees of habitat fragmentation and a wider pool of wood ant species coexistence.

4.6 Acknowledgements

This work was funded by a NERC iCASE PhD studentship to EJHR, supporting JM. The ddRAD library preparations and sequencing was funded by the NERC Environmental Omics Facility (project code: NBAF1206). We are grateful to A.C. Everett, M. Holgate, D. Lamin, K.A. Monaghan and J. Podesta for assistance with field sampling. Additional specimens were kindly provided by D. D. Burns (Longshaw Estate).

Chapter 5

Discussion

5.1 Thesis overview

In this thesis I have combined morphological, mitochondrial (mtDNA) haplotype, and nuclear genome-wide double-digest restriction side associated DNA sequencing (ddRAD) data for three species of mound-building red wood ants (*Formica rufa* group) across their distributions in the British Isles in order to characterise population structure, look for signals of gene flow between the species, and model the effects of habitat fragmentation on their genetic diversity. I also use these data to explore whether the British populations of *F. rufa* are descended from hybrid *F. rufa* X *F. polyctena*. I included *F. rufa* group samples from Belgium and Finland to form a basis for comparison for inference of hybridisation amongst British species, and to shed light on the question of *F. rufa* species origins in the British Isles. In Chapter 2, I assigned morphospecies identities and mtDNA haplotypes to nests from Britain, showing considerable mismatch between the two. These results suggested *F. lugubris* hybridised with both other species where their British distributions overlapped. Furthermore, the low mtDNA sequence diversity, overlap between morphological character data, and the pattern of discordance

between the morphological and mitochondrial data suggested neither were sufficient to reliably assign species to nests amongst Scottish populations. In Chapter 3, I generated ddRAD libraries for nests across our British and European sampling effort. Nests clustered strongly by species across multiple analyses of these data, somewhat contradicting the results of Chapter 2. However, there were a small number of British nests that were clearly hybrids or backcrossed hybrids, displaying similar patterns of co-ancestry to their known hybrid cousins from Finland. When combined, the data from Chapters 2 and 3 provide compelling support for a picture of sporadic hybridisation between species, followed by backcrossing that leaves mtDNA introgression as the only signal of hybridisation. The data of Chapter 3 also hinted that some lineages of British *F. rufa* may be descended from hybrid *F. rufa* X *F. polyctena*, but not all. In Chapter 4, I modelled the effects of woodland habitat fragmentation on various measures of genetic diversity within the ddRAD data for British samples. The results of the modelling were mixed, but generally the results showed genetic diversity differed between species and there were species-specific effects of latitude. While our fragmentation measures did not significantly predict genetic diversity in most cases, the effect of woodland cover or woodland fragment area on expected heterozygosity (gene/allele diversity) was found to be significant, perhaps hinting that this is worth further investigation.

5.2 Species relationships in British red wood ants

Overall, our data suggest sporadic hybridisation events followed by backcrossing into parental species, usually *F. lugubris* based on the patterns of mtDNA introgression compared to nuclear genome. This indicates species boundaries are quite strong in Britain, especially when compared to the porosity found in some continental European populations (e.g. *F. rufa* with *F. polyctena*; Seifert, 2021). There is no evidence of the kind of widespread hybridisation found elsewhere, such as "hybrid zones" (Purcell et al.,

2016) or "hybrid swarms" in Finland (see Beresford et al., 2017). Counter-intuitively, we found no evidence that increased fragmentation is increasing hybridisation in reduced habitat landscapes where two species co-occur, although our sample sizes were small and may have been insufficient to capture any signal (Chapter 4).

All nests that were identified as *F. aquilonia* using the morphological characters described in Chapter 2 also had an *F. aquilonia* haplotype and were positioned with *F. aquilonia* in the `fineRADstructure` and PC analyses of the ddRAD data (Chapter 3)(see Fig. 5.1). The same pattern of consistency is seen with nests morphologically identified as *Formica rufa*. However, this is not the case for *F. lugubris*. Haplotypes from all three species groups on our phylogeny (Chapter 2) are represented in morphologically *F. lugubris* nests. This mismatch implied introgression from both species into *F. lugubris*, however, this was not borne out by the ddRAD data. Regarding the morph. *F. lugubris*/mtDNA *F. rufa*/RAD *F. lugubris*, I think we spelled out quite a convincing case for hybridisation and backcrossing in Chapters 2 and 3. However, for the *F. lugubris*/*F. aquilonia* mismatch nests it may appear, at first glance, this is purely a case of morphological misidentification in two difficult to distinguish species and that all the morphologically *F. lugubris* samples with *F. aquilonia* mtDNA and genomic material are misidentified *F. aquilonia* (see Limitations section below). While I accept this could be the case I think, for a number of reasons, the reality may be slightly more complicated: (i) The putative hybrid sample (60_12, see Fig. 3.4 that clusters with *F. lugubris* in the RAD data but shares higher coancestry with *F. aquilonia* than its conspecifics. This nest is morphologically *F. lugubris* and has a *F. lugubris* haplotype, which suggests a male *F. aquilonia* mated with a *F. lugubris* gyne and then backcrossed with *F. lugubris*. (ii) All but one (excluding the misidentified Achall nest 45_01) of the mismatched morph. *F. lugubris*/mtDNA *F. aquilonia* were from mixed species woodland and thus have the potential to be a product of hybridisation. (iii) There is an increase in coancestry between *F. aquilonia* and Scottish *F. lugubris* compared to English and Welsh *F. lugubris*,

suggesting some form of gene flow. (iv) Local experts were sure of relatively common "morphological intermediates" between the two species where they occur in sympatry, suggesting gene flow (Macdonald, M, pers. comm., 2018) If male *F. aquilonia* did mate with *F. lugubris* gynes we would not see a mitochondrial signal, and the P-type social structure of *F. lugubris* means the newly-mated queen would likely return to her colony where her offspring may backcross via intranidal mating. Further analysis of the ddRAD data alongside increased understanding of the other phenotypic, life history and biological characteristics of the species would help shed light on the question of *F. lugubris* and *F. aquilonia* hybridisation.

Irrespective of my theorising, *F. lugubris* and *F. aquilonia* are clearly robust species in their British ranges. Seifert (2021) questions whether we should maintain species names for entities that hybridise extensively and have little to no reproductive boundary (e.g. *F. rufa* with *F. polyctena*, or *F. polyctena* with *F. aquilonia*), but ultimately sides with pragmatism and decides to keep the current species names so as to not lose decades of references in the literature. Unlike the rampant gene-flow found between some species pairs on the continent, I don't think the data we presented give us any reason to start questioning currently accepted species taxonomy (or concepts).

5.3 Limitations and further work

One obvious limitation of the work presented was the nature of the morphological character data I was able to include in species identification (Chapter 2). I selected characters that are commonly used in Britain to distinguish different species (see Chapter 2 methods), however, did not have the time or expertise to apply the level of morphometric data collection and analysis required for more in-depth investigation. Given enough of either, applying an approach such as NUMOBAT (Seifert, 2009) during morphological examination and identification of our samples would improve the reliability of the assigned

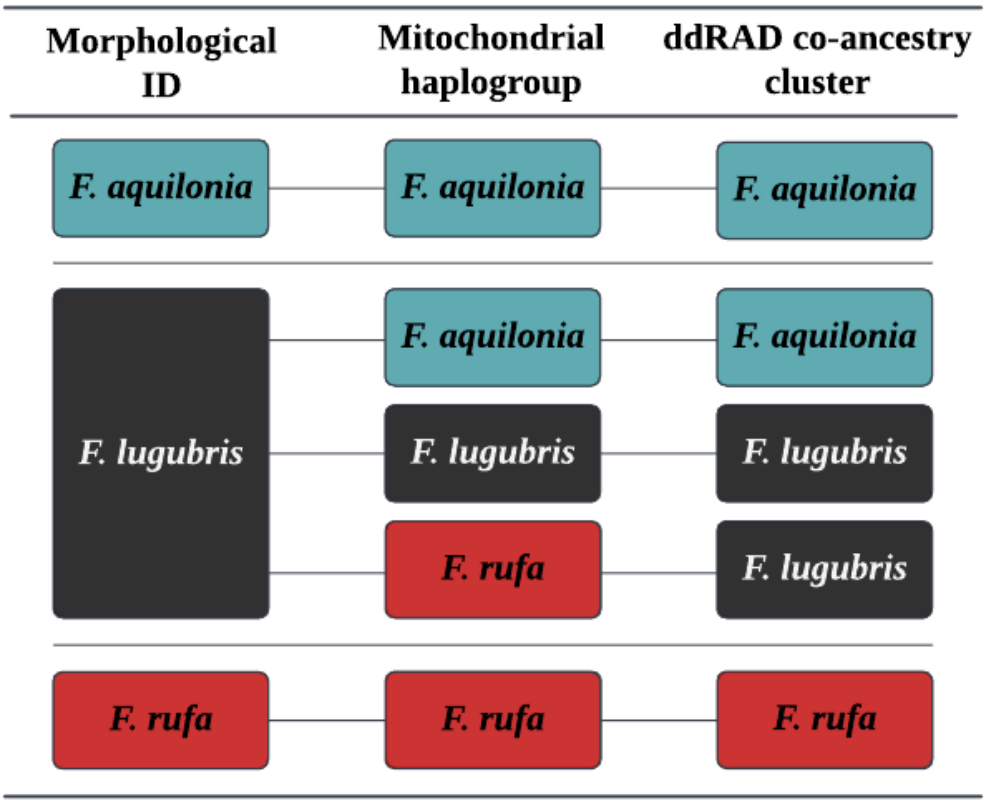

Figure 5.1: Visual summary of the possible morphology, mitochondrial haplogroup and ddRAD co-ancestry cluster analysis combinations. Morphological species identification is on the left, mitochondrial haplogroup in the middle and ddRAD result on the right. Both *F. aquilonia* and *F. rufa* show consistency across data types, whereas morphological *F. lugubris* shows multiple combinations. However, the simplicity of this summary should be viewed with caution, particularly in light of the uncertainty of some morphological identification.

species IDs. Furthermore, application of, for example, linear discriminant analyses (see Seifert (2021) for a comprehensive description and discussion of Eurasian red wood ant morphometrics) would have allowed a more compelling discussion of (i) how British populations align morphologically with those from the continent, and (ii) how hybridisation among species has affected morphological phenotypes of British wood ant species (if at

all). This may have changed the results and inferences regarding the Scottish nests in particular, however, such speculation is of little use. In-depth morphological or morphometric analysis for these nests is possible on stored samples and may help clarify the morphological status of Scottish species. (Romantically if irrelevantly, we have several nest samples from the Black Wood at Rannoch in Scotland, the site from which Yarrow first described *F. aquilonia*; Yarrow, 1955). I think this would be a considerable but worthwhile undertaking, especially considering the genetic data I have presented here (Chapters 2 and 3) alongside which any new morphological data could be used to provide fresh insights into species composition and boundaries in British *F. rufa* group species.

I used the COI-II mitochondrial gene fragment to allow direct comparison with North York Moors populations (Procter, 2016). However, this fragment is little used in the literature and our results could not be compared to other *F. rufa* group mtDNA data (e.g. Goropashnaya et al., 2012; Beresford et al., 2017). Unfortunately, this meant I could not include published data from a number of other studies of *F. rufa* group species across Europe that used either another cytochrome oxidase fragment (Holzer et al., 2009; Bernasconi et al., 2011; Vanhala et al., 2014) or a region of cytochrome b (e.g. Gyllenstrand and Seppä, 2003; Goropashnaya, Fedorov and Pamilo, 2004; Kulmuni et al., 2010; Beresford et al., 2017). Sequencing the other fragments in workers from our nest samples would allow direct phylogenetic analysis with all the published data. Alternatively, because mtDNA does not recombine just sequencing a subset of our samples to infer how the haplotypes I found relate to those in the other mtDNA regions would allow suitable comparison and a broader pan-European context for the data presented in Chapter 2.

In Chapter 3, I presented principal component and estimated coancestry analyses of the genomic ddRAD dataset, which showed us an interesting picture of population structure

and limited hybridisation within British wood ants. However, to infer more about demographic and population structure, current gene flow patterns and varying rates of gene flow within the genomes I needed to apply further tests. Use of ABBA-BABA and $f4$ statistics, and analysis using STRUCTURE and ADMIXTURE would have enabled further worthwhile analysis of our ddRAD dataset.

In Chapter 4, I modelled the effects of two habitat fragmentation measures on genetic diversity. I think these models would have been improved quite dramatically by (i) the use of individual–level rather than population–level statistics, and (ii) the inclusion of some form of woodland age–related statistic. Forest age has been shown to be an important habitat characteristic with regard to *F. lugubris* (Procter, 2016) and *F. polyctena* (Berberich et al., 2020) in English and German fragmented landscapes, respectively. The latter is an exhaustive landscape and wood ant population survey, which would be a huge undertaking for any study population and not all the variables they used are scaleable to a national study for example. However, including data from the Ancient Woodland Inventories of England, Wales and Scotland would be a useful means of testing whether characteristics such as woodland age or proximity to ancient/pre-modern forestry are able to predict genetic diversity. Human mediated destruction of ancient forestry is habitat fragmentation and so I would view these variables as an extension to what we attempted in Chapter 4.

As discussed above, there is more to be done with the ddRAD data. In addition to the statistics outlined, I would be very interested in several avenues of further investigation. First, using the data to infer demographic histories for the species in the UK and, with the help of some further sampling from continental wood ant populations, perhaps inferring routes of colonisation into the British Isles from glacial refugia. Applying the kind of modelling carried out by Portinha et al. (2022) (again alongside further samples from the continent) may allow us to further test any hypotheses that might arise from the

previous analysis. If time and money were no object, I would like to sample populations I could not visit during the 2018 sampling season (although not in a "throw more data at the problem" manner). In particular sites from further south in Wales as *F. rufa* are thought to be struggling there, even to the point of population extinction in some woodlands, and I would like to investigate whether there is a genetic diversity component to the poor population health.

Gathering additional data on the characteristics of nests and individuals would also aid robust interpretation of the genetic data. Two lines of questioning I have are as follows: (i) Are the *F. rufa* more polygynous than their continental conspecifics all the way across their British range, or have some populations shifted behaviour and what might this tell us about hybridisation with *F. lugubris*? (ii) Are there any biological differences between hybrid and parental species nests? For example, nest temperatures are different in hybrid colonies versus parental species in some regions of Finland (Kulmuni, J., pers. comm. 2022). Is the same true here? Could it tell us anything about the ecological effects/drivers of hybridisation?

5.4 Implications for conservation

The results presented in this thesis are broadly positive for practical conservation of British *F. rufa* group species. First, they are "real" species in that there are clearly species boundaries maintaining a relatively high level of reproductive isolation, and this removes the issue of whether rare hybrids should be protected (Balzani et al., 2022). Furthermore, I have characterised genetic diversity across a variety of populations thus updating species distributions (i.e. confirming species presence in some woodland) and improving our understanding of population health. Indeed, some of the data presented here conflict with the impressions of local experts on the ground (especially regarding *F. rufa* in Wales and the potential for hybridisation in Scotland) and hopefully will be

of use in clarifying local management and conservation strategies.

Further analysis of these data will only enhance the potential usefulness in a conservation context.

5.5 Conclusion

The data I have presented in this thesis has added to the knowledge of wood ant genetic diversity within the UK. The combination of morphological and genetic data have given a reasonably robust characterisation of species composition, population structure and limited gene flow between species. The habitat fragmentation modelling hints at that further investigation may be able to succeed where our statistical power was insufficient.

In addition, this work provides a large genomic data set that may serve as the foundation upon which to build an understanding of British wood ants that begins to echo that of continental populations.

Appendix A

Appendix I: Material for Chapter 2

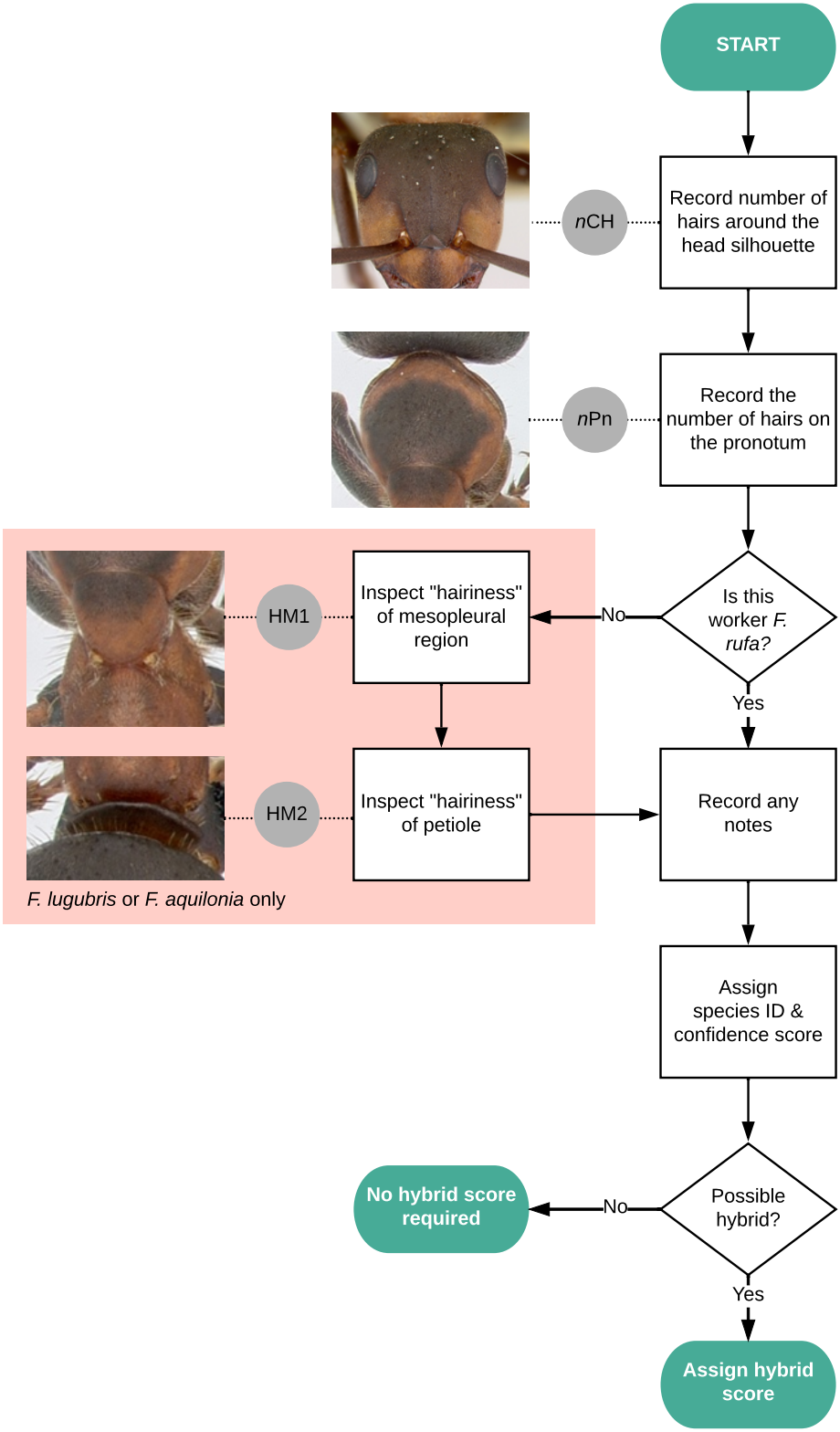

Figure A.1: Schematic representation of our morphological data collection process flow.

List of References

Abbott, R. J., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J. W., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C. D., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A. W., Parisod, C., Pfennig, K., Rice, A. M., Ritchie, M. G., Seifert, B., Smadja, C. M., Stelkens, R., Szymura, J. M., Väinölä, R., Wolf, J. B. W. and Zinner, D. (2013), 'Hybridization and speciation', *Journal of Evolutionary Biology* **26**(2), 229–246.

Alexander, D., Novembre, J. and Lange, K. (2009), 'Fast model-based estimation of ancestry in unrelated individuals', *Genome research* **19**, 1655–64.

Anderson, E. and Stebbins, G. (1954), 'Hybridization as an evolutionary stimulus', *Evolution* **8**(4), 378–88.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G. and Hohenlohe, P. A. (2016), 'Harnessing the power of RADseq for ecological and evolutionary genomics', *Nat. Rev. Genet.* **17**(2), 81–92.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. and Johnson, E. A. (2008), 'Rapid SNP discovery and genetic mapping using sequenced RAD markers', *PLoS One* **3**(10), e3376.

Balmori-de la Puente, A., Ventura, J., Miñarro, M., Somoano, A., Hey, J. and Castresana, J. (2022), 'Divergence time estimation using ddRAD data and an isolation-with-migration model applied to water vole populations of arvicola', *Sci. Rep.* **12**(1), 4065.

Balzani, P., Dekoninck, W., Feldhaar, H., Freitag, A., Frizzi, F., Frouz, J., Masoni, A., Robinson, E., Sorvari, J. and Santini, G. (2022), 'Challenges and a call to action for protecting european red wood ants', *Conservation Biology* .

Bargum, K., Boomsma, J. J. and Sundström, L. (2004), 'A genetic component to size in queens of the ant, formica truncorum', *Behav. Ecol. Sociobiol.* **57**(1), 9–16.

Barton, N. H. and Gale, K. S. (1993), Genetic analysis of hybrid zones.

Basak, M., Uzun, B. and Yol, E. (2019), 'Genetic diversity and population structure of the mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq', *PLoS One* **14**(10), e0223757.

Berberich, G. M., Berberich, M. B. and Gibhardt, M. (2020), 'Red wood ants prefer mature pine forests in variscan granite environments (formica rufa-group)', *bioRxiv* .
URL: *https://www.biorxiv.org/content/early/2020/11/27/2020.11.26.390112*

Beresford, J., Elias, M., Pluckrose, L., Sundström, L., Butlin, R. K., Pamilo, P. and Kulmuni, J. (2017), 'Widespread hybridization within mound-building wood ants in Southern Finland results in cytonuclear mismatches and potential for sex-specific hybrid breakdown', *Molecular Ecology* **26**(15), 4013–4026.

Bernasconi, C., Cherix, D., Seifert, B. and Pamilo, P. (2011), 'Molecular taxonomy of the Formica rufa group (red wood ants) (Hymenoptera: Formicidae): a new cryptic species in the Swiss Alps ?', *Myrmecological News* **14**(January), 37–47.

Bernasconi, C., Maeder, A., Cherix, D. and Pamilo, P. (2005), 'Diversity and genetic structure of the wood ant formica lugubris in unmanaged forests', *Ann. Zool. Fennici* **42**, 189–199.

Bolger, A. M., Lohse, M. and Usadel, B. (2014), 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics* **30**(15), 2114–2120.
URL: *https://doi.org/10.1093/bioinformatics/btu170*

Bourke, A. F. G. and Franks, N. R. (1995), *Social Evolution in Ants*, Princeton University Press.

Breen, J. (1979), 'Nest sites of formica lugubris (hymenoptera, formicidae) in irish plantation woods', *Journal of Life Sciences* **1**, 13–32.

Brelsford, A., Purcell, J., Avril, A., Tran Van, P., Zhang, J., Brütsch, T., Sundström, L., Helanterä, H. and Chapuisat, M. (2020), 'An ancient and eroded social supergene is widespread across formica ants', *Curr. Biol.* **30**(2), 304–311.e4.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. and Cresko, W. A. (2013), 'Stacks: an analysis tool set for population genomics', *Molecular Ecology* **22**(11), 3124–3140.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.12354*

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. and Lee, J. J. (2015), 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience* **4**(1). s13742-015-0047-8.
URL: *https://doi.org/10.1186/s13742-015-0047-8*

Chuang, A. and Peterson, C. R. (2016), 'Expanding population edges: theories, traits, and trade-offs', *Global Change Biology* **22**(2), 494–512.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13107*

Collingwood, C. A. (1979), 'The formicidae (hymenoptera) of fennoscandia and denmark.', *Fauna Entomologica Scandinavica* **8**(0), 1–174.

Commission, F. (2022), Chapter 1: Woodland area, planting and restocking, 2022 edn., Technical report.
URL: *https://cdn.forestresearch.gov.uk/2022/09/Ch1*$_W$*oodland*$_2$*022.pdf*

Consortium, T. H. G. (2012), 'Butterfly genome reveals promiscuous exchange of mimicry adaptations among species', *Nature* **487**, 94–98.

DaCosta, J. M. and Sorenson, M. D. (2014), 'Amplification biases and consistent recovery of loci in a double-digest rad-seq protocol', *PLOS ONE* **9**(9), 1–14.
URL: *https://doi.org/10.1371/journal.pone.0106713*

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. and Li, H. (2021), 'Twelve years of SAMtools and BCFtools', *Gigascience* **10**(2).

de la Filia, A. G., Bain, S. A. and Ross, L. (2015), 'Haplodiploidy and the reproductive ecology of arthropods', *Current Opinion in Insect Science* **9**, 36–43. Pests and resistance * Behavioural ecology.
URL: *https://www.sciencedirect.com/science/article/pii/S221457451500084X*

DEBOUT, G., SCHATZ, B., ELIAS, M. and MCKEY, D. (2007), 'Polydomy in ants: what we know, what we think we know, and what remains to be done', *Biological Journal of the Linnean Society* **90**(2), 319–348.
URL: *https://doi.org/10.1111/j.1095-8312.2007.00728.x*

Degnan, J. H. and Rosenberg, N. A. (2009), 'Gene tree discordance, phylogenetic inference and the multispecies coalescent', *Trends Ecol. Evol.* **24**(6), 332–340.

Dekoninck, W., Maebe, K., Breyne, P. and Hendrickx, F. (2014), 'Polygyny and strong genetic structuring within an isolated population of the wood ant formica rufa', *Journal of Hymenoptera* **41**, 95–111.

Domisch, T., Risch, A. and Robinson, E. (2016), Wood ant foraging and mutualism with aphids. Wood ant ecology and conservation, *in* J. A. Stockan and E. J. H. Robinson, eds, 'Wood Ant Ecology and Conservation'.

Donisthorpe, H. S. J. K. (1927), *British Ants, Their Life-History And Classification*, 2 edn, Routledge.

Durand, E. Y., Patterson, N., Reich, D. and Slatkin, M. (2011), 'Testing for ancient admixture between closely related populations', *Mol. Biol. Evol.* **28**(8), 2239–2252.

Ellegren, H. and Galtier, N. (2016), 'Determinants of genetic diversity', *Nat. Rev. Genet.* **17**(7), 422–433.

Ellis, S. and Robinson, E. J. H. (2015), 'The role of non-foraging nests in polydomous wood ant colonies', *PLoS ONE* **10**(10).

Escoda, L., Fernández-González, Á. and Castresana, J. (2019), 'Quantitative analysis of connectivity in populations of a semi-aquatic mammal using kinship categories and network assortativity', *Mol. Ecol. Resour.* **19**(2), 310–326.

FAO (2020), Global forest resources assessment 2020 – key findings, Technical report, Rome.

Faraway, J. J. (2004), *Linear Models with R*, 1 edn, Chapman and Hall/CRC.

Faraway, J. J. (2016), *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 2 edn, Chapman and Hall/CRC.

Feldhaar, H., Foitzik, S. and Heinze, J. (2008), 'Lifelong commitment to the wrong partner: hybridization in ants', *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1505), 2891–2899.
URL: *http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2008.0022*

Fischer, J. and Lindenmayer, D. B. (2007), 'Landscape modification and habitat fragmentation: a synthesis', *Global Ecology and Biogeography* **16**(3), 265–280.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1466-8238.2007.00287.x*

Forestry Commission (2019), 'National Forest Inventory'.
URL: *data-forestry.opendata.arcgis.com/datasets/67b8a420316940b593c835685388be01_0/explore?locati*
2.724655%2C6.12

Fournier, D., Battaille, G., Timmermans, I. and Aron, S. (2008), 'Genetic diversity, worker size polymorphism and division of labour in the polyandrous ant cataglyphis cursor', *Animal Behaviour* **75**(1), 151–158.
URL: *https://www.sciencedirect.com/science/article/pii/S0003347207003636*

Frankham, R., Ballou, J. and D.A., B. (2002), *Introduction to Conservation Genetics*, Cambridge University Press, Cambridge, UK.

Frouz, J., Jílková, V. and Sorvari, J. (2016), Contribution of wood ants to nutrient cycling and ecosystem function., *in* J. A. Stockan and E. J. H. Robinson, eds, 'Wood Ant Ecology and Conservation', pp. 207–220.

Garcia-Erill, G., Kjær, M. M., Albrechtsen, A., Siegismund, H. R. and Heller, R. (2021), 'Vicariance followed by secondary gene flow in a young gazelle species complex', *Molecular Ecology* **30**(2), 528–544.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15738*

Goropashnaya, A., Fedorov, V. and Pamilo, P. (2004), 'Recent speciation in the Formica rufa group ants (Hymenoptera, Formicidae): Inference from mitochondrial DNA phylogeny', *Molecular Phylogenetics and Evolution* **32**(1), 198–206.

Goropashnaya, A., Fedorov, V., Seifert, B. and Pamilo, P. (2012), 'Phylogenetic relationships of Palaearctic Formica species (hymenoptera, Formicidae) based on mitochondrial cytochrome b sequences', *PLoS ONE* **7**(7), 1–7.

Goropashnaya, A. V., Fedorov, V. B., Seifert, B. and Pamilo, P. (2004), 'Limited phylogeographical structure across Eurasia in two red wood ant species Formica pratensis and F. lugubris (Hymenoptera, Formicidae)', *Molecular Ecology* **13**(7), 1849–1858.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D. and Pääbo, S. (2010), 'A draft sequence of the neandertal genome', *Science* **328**(5979), 710–722.

Gyllenstrand, N. and Seppä, P. (2003), 'Conservation genetics of the wood ant, Formica lugubris, in a fragmented landscape', *Molecular Ecology* **12**(11), 2931–2940.

Gyllenstrand, N., Seppä, P. and Pamilo, P. (2004), 'Genetic differentiation in sympatric wood ants, Formica rufa and F. polyctena', *Insectes Sociaux* **51**(2), 139–145.

Gyllenstrand, N., Seppä, P. and Pamilo, P. (2005), 'Restricted gene flow between two social forms in the ant Formica truncorum', *Journal of Evolutionary Biology* **18**(4), 978–984.

Haatanen, M.-K. and Sorvari, J. (2013), 'Similarity of body size in queens of the wood ant Formica aquilonia from optimal and sub-optimal habitats indicates a strong heritable component', *Journal of Insect Science* **13**(1). 115.
URL: *https://doi.org/10.1673/031.013.11501*

Hakala, S., Seppä, P. and Helanterä, H. (2019), 'Evolution of dispersal in ants (hymenoptera: Formicidae): a review on the dispersal strategies of sessile superorganisms', *Myrmecological News* **29**, 35–55.

Hall, T. (1999), 'BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95 / 98 / NT', *Nucleic Acids Symposium Series* .

Hambler, C., Henderson, P. A. and Speight, M. R. (2011), 'Extinction rates, extinction-prone habitats, and indicator groups in britain and at larger scales', *Biological Conservation* **144**(2), 713–721.
URL: *https://www.sciencedirect.com/science/article/pii/S000632071000385X*

Härkönen, S. K. and Sorvari, J. (2014), 'Species richness of associates of ants in the nests of red wood antformica polyctena(hymenoptera, formicidae)', *Insect Conserv. Divers.* **7**(6), 485–495.

Hartl, D. L. and Clark, A. G. (2007), *Principles of population genetics*, 4 edn, Sinauer Associates.

Heinze, J., Foitzik, S., Fischer, B., Wanke, T. and Kipyatkov, V. (2003), 'The significance of latitudinal variation in body size in a holarctic ant, leptothorax acervorum', *Ecography* **26**, 349 – 355.

Helanterä, H., Strassmann, J. E., Carrillo, J. and Queller, D. C. (2009), 'Unicolonial ants: where do they come from, what are they and where are they going?', *Trends Ecol. Evol.* **24**(6), 341–349.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. and Cresko, W. A. (2010), 'Population genomics of parallel adaptation in threespine stickleback using sequenced rad tags', *PLOS Genetics* **6**(2), 1–23.
URL: *https://doi.org/10.1371/journal.pgen.1000862*

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics* **6**(2), 65–70.
URL: *http://www.jstor.org/stable/4615733*

Holzer, B., Chapuisat, M., Kremer, N., Finet, C. and Keller, L. (2006), 'Unicoloniality, recognition and genetic differentiation in a native formica ant', *J Evol Biol* **19**, 2031–9.

Holzer, B., Keller, L. and Chapuisat, M. (2009), 'Genetic clusters and sex-biased gene flow in a unicolonial Formica ant', *BMC Evolutionary Biology* **9**(1), 69.
URL: *http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-9-69*

Hughes, J. (2006), 'A review of wood ants (Hymenoptera: Formicidae) in Scotland', *Scottish Natural Heritage Commissioned Report No.178* **(ROAME No.**

John, F. and Georges, M. (1992), 'Generalized collinearity diagnostics', *Journal of the American Statistical Association* **87**(417), 178–183.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475190*

Juhász, O., Bátori, Z., Trigos-Peral, G., Lőrinczi, G., Módra, G., Bóni, I., Kiss, P. J., Aguilon, D. J., Tenyér, A. and Maák, I. (2020), 'Large- and small-scale environmental factors drive distributions of ant mound size across a latitudinal gradient', *Insects* **11**(6).
URL: *https://www.mdpi.com/2075-4450/11/6/350*

Keller, L. (1991), 'Queen number, mode of colony founding, and queen reproductive success in ants (hymenoptera formicidae)', *Ethology Ecology & Evolution* **3**, 307–316.

Kulmuni, J., Nouhaud, P., Pluckrose, L., Satokangas, I., Dhaygude, K. and Butlin, R. K. (2020), 'Instability of natural selection at candidate barrier loci underlying speciation in wood ants', *Molecular Ecology* **29**(20), 3988–3999.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15606*

Kulmuni, J. and Pamilo, P. (2014), 'Introgression in hybrid ants is favored in females but selected against in males', *Proceedings of the National Academy of Sciences* **111**(35), 12805–12810.
URL: *http://www.pnas.org/cgi/doi/10.1073/pnas.1323045111*

Kulmuni, J., Seifert, B. and Pamilo, P. (2010), 'Segregation distortion causes large-scale differences between male and female genomes in hybrid ants', *Proceedings of the National Academy of Sciences* **107**(16), 7371–7376.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018), 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution* .

Li, H. (2013), 'Aligning sequence reads, clone sequences and assembly contigs with bwa-mem'.
URL: *https://arxiv.org/abs/1303.3997*

Linksvayer, T. A. (2006), 'Direct, maternal, and sibsocial genetic effects on individual and colony traits in an ant', *Evolution* **60**(12), 2552–2561.

Magris, G., Marroni, F., D'Agaro, E., Vischi, M., Chiabà, C., Scaglione, D., Kijas, J., Messina, M., Tibaldi, E. and Morgante, M. (2022), 'ddRAD-seq reveals the genetic structure and detects signals of selection in italian brown trout', *Genet. Sel. Evol.* **54**(1), 8.

Mäki-Petäys, H. and Breen, J. (2007), 'Genetic vulnerability of a remnant ant population', *Conservation Genetics* **8**(2), 427–435.

Mäki-Petäys, H., Zakharov, A., Viljakainen, L., Corander, J. and Pamilo, P. (2005), 'Genetic changes associated to declining populations of formica ants in fragmented forest landscape', *Mol. Ecol.* **14**(3), 733–742.

Malinsky, M., Trucchi, E., Lawson, D. J. and Falush, D. (2018), 'RADpainter and fineRADstructure: Population inference from RADseq data', *Mol. Biol. Evol.* **35**(5), 1284–1290.

Mallet, J. (2005), 'Hybridization as an invasion of the genome', *Trends in Ecology and Evolution* **20**(5), 229–237.

Mallet, J. (2007), 'Hybrid speciation', *Nature* **446**(7133), 279–283.

Mallet, J., Besansky, N. and Hahn, M. W. (2016), 'How reticulated are species?', *BioEssays* **38**(2), 140–149.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201500149*

Martin, M. (2011), 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet J.* **17**(1), 10.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. and Jiggins, C. D. (2013), 'Genome-wide evidence for speciation with gene flow in heliconius butterflies', *Genome Res.* **23**(11), 1817–1828.

Mason, W. (2007), 'Changes in management of British forestry between 1948 and 2000 possible future trends', *Ibis* **149**(2), 42–52.

Mayr, E. (1963), *Animal Species and Evolution*, Belknap Press.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L. and Seehausen, O. (2017), 'Ancient hybridization fuels rapid cichlid fish adaptive radiations', *Nat. Commun.* **8**(1), 14363.

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y. and Schumer, M. (2021), 'The genomic consequences of hybridization', *eLife* **10**, e69016.
URL: *https://doi.org/10.7554/eLife.69016*

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R. and Hohenlohe, P. A. (2013), 'Genotyping-by-sequencing in ecological and conservation genomics', *Mol. Ecol.* **22**(11), 2841–2847.

NBN Atlas (2018), 'NBN Atlas occurrence download'.
URL: *http://nbnatlas.org*

Nei, M. and Li, W. H. (1979), 'Mathematical model for studying genetic variation in terms of restriction endonucleases.', *Proceedings of the National Academy of Sciences* **76**(10), 5269–5273.

Nouhaud, P., Beresford, J. and Kulmuni, J. (2021), 'Cost-effective long-read assembly of a hybrid formica aquilonia × formica polyctena wood ant genome from a single haploid individual', *bioRxiv* .
URL: *https://www.biorxiv.org/content/early/2021/03/10/2021.03.09.434597*

Nouhaud, P., Beresford, J. and Kulmuni, J. (2022), 'Assembly of a hybrid formica aquilonia × f. polyctena ant genome from a haploid male', *J. Hered.* **113**(3), 353–359.

Pamilo, P. and Crozier, R. H. (1997), 'Population biology of social insect conservation', *Memoirs of the Museum of Victoria* **52**(2), 411–419.

Parmentier, T., Dekoninck, W. and Wenseleers, T. (2014), 'A highly diverse microcosm in a hostile world: A review on the associates of red wood ants (Formica rufa group)', *Insectes Sociaux* **61**(3), 229–237.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012), 'Ancient admixture in human history', *Genetics* **192**(3), 1065–1093.

Pebesma, E. (2018), 'Simple Features for R: Standardized Support for Spatial Vector Data', *The R Journal* **10**(1), 439–446.
URL: *https://doi.org/10.32614/RJ-2018-009*

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. and Hoekstra, H. E. (2012), 'Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species', *PLoS One* **7**(5), e37135.

Picard toolkit (2019), `https://broadinstitute.github.io/picard/`.

Portinha, B., Avril, A., Bernasconi, C., Helanterä, H., Monaghan, J., Seifert, B., Sousa, V. C., Kulmuni, J. and Nouhaud, P. (2022), 'Whole-genome analysis of multiple wood ant population pairs supports similar speciation histories, but different degrees of gene flow, across their european ranges', *Molecular Ecology* **31**(12), 3416–3431.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16481*

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000), 'Inference of Population Structure Using Multilocus Genotype Data', *Genetics* **155**(2), 945–959.
URL: *https://doi.org/10.1093/genetics/155.2.945*

Procter, D. S. (2016), The effects of forest cover change and polydomous colony organisation on the wood ant Formica lugubris, PhD thesis.

Punttila, P. (1996), 'Succession, forest fragmentation, and the distribution of wood ants', *OIKOS* **75**, 291–298.

Punttila, P. and Kilpeläinen, J. (2009), 'Distribution of Mound-Building Ant Species (Formica spp., Hymenoptera) in Finland: Preliminary Results of a National Survey', *Annales Zoologici Fennici* **46**(1), 1 – 15.
URL: *https://doi.org/10.5735/086.046.0101*

Purcell, J., Zahnd, S., Athanasiades, A., Türler, R., Chapuisat, M. and Brelsford, A. (2016), 'Ants exhibit asymmetric hybridization in a mosaic hybrid zone', *Molecular Ecology* **25**(19), 4866–4874.

Purcell, S. and Chang, C. (n.d.), 'Plink 2.0', `www.cog-genomics.org/plink/2.0/`.

QGIS Development Team (2022), *QGIS Geographic Information System*, QGIS Association.
URL: *https://www.qgis.org*

R Core Team (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Riutta, T., Slade, E. M., Morecroft, M. D., Bebber, D. P. and Malhi, Y. (2014), 'Living on the edge: quantifying the structure of a fragmented forest landscape in england', *Landsc. Ecol.* **29**(6), 949–961.

Robinson, E., Stockan, J. and Iason, G. (2016), Wood ants and their interaction with other organisms., *in* J. A. Stockan and E. J. H. Robinson, eds, 'Wood Ant Ecology and Conservation', p. 177–206.

Robinson, N. A. and Robinson, E. J. H. (2008), 'The population of the red wood ant Formica rufa L. (Hymenoptera: Formicidae) at Gait Barrows National Nature Reserve, Lancashire, England over the 20 year period 1986–2006: nest longevity, reproduction and the effect of management.', *British Journal of Entomology and Natural History* **21**(4), 225–241.

Rosengren, R. and Pamilo, P. (1983), 'The evolution of polygyny and polydomy in mound-building Formica ants.', *Acta Entomol Fenn* p. 65–77.

Rosengren, R. and Sundström, L. (1991), *The interaction between red wood ants Cinara aphids and pines: - a ghost of a mutualism past?*, Oxford University Press, Britannia, pp. 80–92.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N. and Reich, D. (2014), 'The genomic landscape of neanderthal ancestry in present-day humans', *Nature* **507**(7492), 354–357.

Saunders, D. A., Hobbs, R. J. and Margules, C. R. (1991), 'Biological consequences of ecosystem fragmentation: A review', *Conserv. Biol.* **5**(1), 18–32.

Seehausen, O. (2004), 'Hybridization and adaptive radiation', *Trends Ecol. Evol.* **19**(4), 198–207.

Seifert, B. (1991), 'The phenotypes of the Formica rufa complex in East Germany', *Abh Ber Naturkundemus Görlitz* **65**(1), 1–27.

Seifert, B. (1996), 'Formica paralugubris nov. spec. – a sympatric sibling species of Formica lugubris from the western Alps (Insecta: Hymenoptera: Formicoidea: Formicidae)'.
URL: *https://www.antwiki.org/wiki/images/3/3c/Seifert*$_1$*996a.pdf*

Seifert, B. (2009), 'Cryptic species in ants (hymenoptera: Formicidae) revisited: we need a change in the alpha-taxonomic approach', *Myrmecological News* **12**, 149–166.

Seifert, B. (2019), 'Hybridization in the European carpenter ants Camponotus herculeanus and C. ligniperda (Hymenoptera: Formicidae)', *Insectes Sociaux* **0**(0), 0.
URL: *http://dx.doi.org/10.1007/s00040-019-00693-0*

Seifert, B. (2021), 'A taxonomic revision of the Palaearctic members of the Formica rufa group (Hymenoptera: Formicidae) – the famous mound-building red wood ants', *Myrmecological News* **31**, 133–179.

Seifert, B. and Goropashnaya, A. V. (2004), 'Ideal phenotypes and mismatching haplotypes - Errors of mtDNA treeing in ants (Hymenoptera: Formicidae) detected by standardized morphometry', *Organisms Diversity and Evolution* **4**(4), 295–305.

Seifert, B., Kulmuni, J. and Pamilo, P. (2010), 'Independent hybrid populations of Formica polyctena X rufa wood ants (Hymenoptera: Formicidae) abound under conditions of forest fragmentation', *Evolutionary Ecology* **24**(5), 1219–1237.

Seppä, P. and Pamilo, P. (1995), 'Gene flow and population viscosity in myrmica ants', *Heredity* **74**, 200–209.

Skinner, G. J. and Allen, G. W. (1996), *Ants (Naturalists' Handbooks 24)*, Pelagic Publishing.

Soltis, D. and Soltis, P. (2017), 'The dynamic nature of polyploid genomes', *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8089–8091.

Sorvari, J., Elo, R. and Härkönen, S. (2016), 'Forest-built nest mounds of red wood ant formica aquilonia are no good in clear fells', *Applied Soil Ecology* **101**, 101–106.
URL: *https://www.sciencedirect.com/science/article/pii/S0929139316300191*

Sorvari, J., Haatanen, M.-K. and Vesterlund, S.-R. (2011), 'Combined effects of overwintering temperature and habitat degradation on the survival of boreal wood ant', *J. Insect Conserv.* **15**(5), 727–731.

Sorvari, J. and Hakkarainen, H. (2004), 'Habitat-related aggressive behaviour between neighbouring colonies of the polydomous wood ant formica aquilonia', *Animal Behaviour* **67**(1), 151–153.
URL: *https://www.sciencedirect.com/science/article/pii/S0003347203003592*

Sorvari, J. and Hakkarainen, H. (2005), 'Deforestation reduces nest mound size and decreases the production of sexual offspring in the wood ant formica aquilonia', *Ann. Zool. Fennici* **42**, 259–267.

Sorvari, J. and Hakkarainen, H. (2007), 'The role of food and colony size in sexual offspring production in a social insect: An experiment', *Ecological Entomology* **32**, 11 – 14.

Sorvari, J. and Hakkarainen, H. (2009), 'Forest Clear-Cutting Causes Small Workers in the Polydomous Wood Ant Formica aquilonia', *Annales Zoologici Fennici* **46**(6), 431 – 438.
URL: *https://doi.org/10.5735/086.046.0604*

Spearman, C. (1904), 'The proof and measurement of association between two things', *Am. J. Psychol.* **15**(1), 72.

Stockan, J. A. and Robinson, E. J. H., eds (2016), *Wood Ant Ecology and Conservation*, Cambridge University Press, Cambridge.
URL: *http://ebooks.cambridge.org/ref/id/CBO9781107261402*

Stockan, J. A., Robinson, E. J. H., Yao, I. and Seifert, B. (2016), Introducing wood ants: Evolution, phylogeny, identification and distribution., *in* J. A. Stockan and E. J. H. Robinson, eds, 'Wood Ant Ecology and Conservation', pp. 1–36.

Sundström, L., Seppä, P. and Pamilo, P. (2005), 'Genetic population structure and dispersal patterns in Formica ants — a review', *Annales Zoologici Fennici* **42**, 163–177.
URL: *http://www.jstor.org/stable/23735905%0Ahttp://about.jstor.org/terms*

Sundström, L., Seppä, P. and Pamilo, P. (2004), 'Genetic population structure and dispersal patterns in formica ants - a review', *Ann. Zool. Fenn.* **42**.

Synes, N. W., Ponchon, A., Palmer, S. C., Osborne, P. E., Bocedi, G., Travis, J. M. and Watts, K. (2020), 'Prioritising conservation actions for biodiversity: Lessening the impact from habitat fragmentation and climate change', *Biological Conservation* **252**, 108819.
URL: *https://www.sciencedirect.com/science/article/pii/S0006320720308776*

The Woodland Trust (2017), 'Position Statement: Ancient Woodland'.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994), 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through ldots', *Nucleic Acids Research* .

Vanhala, T., Watts, K., A'Hara, S. and Cottrell, J. (2014), 'Population genetics of Formica aquilonia wood ants in Scotland: The effects of long-term forest fragmentation and recent reforestation', *Conservation Genetics* **15**(4), 853–868.

Wade, T. G., Riitters, K. H., Wickham, J. D. and Jones, K. B. (2003), 'Distribution and causes of global forest fragmentation', *Conservation Ecology* **7**(2).
URL: *http://www.jstor.org/stable/26271943*

Walker, J., Gaffney, V., Fitch, S., Muru, M., Fraser, A., Bates, M. and Bates, R. (2020), 'A great wave: the storegga tsunami and the end of doggerland?', *Antiquity* **94**(378), 1409–1425.

Watts, K. (2006), 'British forest landscapes: The legacy of woodland fragmentation', *Quarterly Journal of Forestry* **100**(4), 273–279.

Watts, K. and Handley, P. (2010), 'Developing a functional connectivity indicator to detect change in fragmented landscapes', *Ecological Indicators* **10**(2), 552–557.

Watts, K., Vanhala, T., Connolly, T. and Cottrell, J. (2016), 'Striking the right balance between site and landscape-scale conservation actions for a woodland insect within a highly fragmented landscape: A landscape genetics perspective', *Biological Conservation* **195**, 146–155.
URL: *https://www.sciencedirect.com/science/article/pii/S0006320715302184*

Welch, R. C. (1978), 'Changes in the distribution of the nests offormica rufa l. (hymenoptera: Formicidae) at blean woods national nature reserve, kent, during the decade following coppicing', *Insectes Sociaux* **25**, 173–186.

Wright, S. (1931), 'EVOLUTION IN MENDELIAN POPULATIONS', *Genetics* **16**(2), 97–159.
URL: *https://doi.org/10.1093/genetics/16.2.97*

Wright, S. (1978), *Evolution and the genetics of populations*, Vol. 4, University of Chicago Press.

Wu, C.-I. (2001), 'The genic view of the process of speciation', *J. Evol. Biol.* **14**(6), 851–865.

Yang, Z. (2010), 'A likelihood ratio test of speciation with gene flow using genomic sequence data', *Genome Biol. Evol.* **2**, 200–211.

Yarrow, I. H. H. (1955), 'The British Ants allied to Formica rufa L. (Hym., Formicidae)', *Transactions of the Society for British Entomology* **12**(1), 1–48.