2022

A Phenotypic and Genomic Investigation of the Identity and Variation in the European Dark Bee (Apis mellifera mellifera)

Victoria G Buswell

PHD

A Phenotypic and Genomic Investigation of the Identity and Variation in the European Dark Bee (Apis mellifera mellifera)

Buswell, Victoria G

Award date:
2022

Awarding institution:
University of Plymouth

UNIVERSITY OF PLYMOUTH

A Phenotypic and Genomic Investigation of the Identity and Variation in the European Dark Bee (*Apis mellifera mellifera*)

by

Victoria G. Buswell

A thesis submitted to the University of Plymouth in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Biological and Marine Sciences

December 2022

Acknowledgements

I would like to thank my supervisors Mairi Knight, Jonathan Ellis and Vanessa Huml for the opportunity to pursue a PhD. I would especially like to express my deep gratitude to my supervisors for their tremendous guidance, support and encouragement throughout this process that has help shaped me as a fledgling scientist. I am also grateful to Andrew Brown for helping guide me through the bee keeping community and providing access to the wealth of knowledge in the B4 group.

A special thanks to the wider bee keeping community who provided samples, particularly members of B4, the Bee Improvement and Bee Breeding Association, the British Bee Keeping Association and the Bee Improvement Programme for Cornwall.

Many thanks to collaborators Gabriele Soland-Reckeweg, Sonja Leidenberger and Mark Barnett.

I have had the pleasure of sharing this journey with Ciaran, Lucy, and Anna, who delved into dungeons, fought dragons, rolled dice, and drank beers with me during this process. A special thanks to Elīna for all her reassurance and listening to me when I needed it.

Finally, I would like to thank my family, Hannah, Tiz, Jeb, and especially Mum and Dad, whose never-ending support and belief in me kept me going.

Author's Declaration

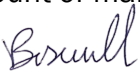
At no time during the registration for the research degree has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

No work submitted for a research degree at the University of Plymouth may form part of any other degree for the candidate either at the University of Plymouth or another institution.

This study was financed by the Natural Environment Research Council (NERC) as a Collaborative Awards in Science and Engineering (CASE) studentship. This study was carried out in collaboration with Bring Back Black Bees (B4) a community interest company that provided bee keeping knowledge and access to bee keepers in the *Apis mellifera mellifera* conservation community. Additional funding was provided by the Eva Crane Trust.

Conferences were attended at which the work was presented and chapters are in preparation to be published with co-authors.

Word count of main body of thesis: 42,047

Signed: 

Date:02/12/2022

Conferences and presentations

2022
Oral presentation:
January, Assessing introgression in *Apis mellifera mellifera* colonies
The 55th annual Population Genetics Group Meeting. Online event
Oral presentation
June; Genetic assessment of subspecies.
The INTERREG Bee research conference, Nordens Ark, Sweden.

2021
Oral presentation:
UK *A. m. mellifera* populations and introgression
November, SICAMM Conference, online event

2020
Oral Presentation:
February, Introgression in honey bees Sustainable Beekeeping: A Future without Imports 2020. Cornwall, England

2019
February, Bee Keeping Survey. The Eden Project, The Eden Projects 2nd Conference For Sustainable Beekeeping. Cornwall, England

2018
February, Bee Keeping Survey. The Eden Project, Sustainable Bee Keeping a Future without Imports. Cornwall England
Oral presentation:
Local Adaptation in UK populations of dark European honey bee (*Apis mellifera mellifera*)
July, The 75th anniversary conference of the Welsh bee keeping association. Aberystwyth University, Wales
Oral presentation:
July, Local Adaptation in UK populations of dark European honey bee (*Apis mellifera mellifera)*. Plymouth University Ecology and Evolution Research Group. Plymouth England.

2017
October, Community engagement conference
Poster presentation:
Investigating the Phenotypic and Genomic basis of Local Adaptation in the UK native honey bee (*Apis mellifera mellifera*). Plymouth, UK

Abstract

Victoria Buswell

A phenotypic and genomic investigation of the identity of *Apis mellifera mellifera*

Apis mellifera (the Western honey bee) is a predominantly human managed pollinator consisting of different evolutionary lineages each containing a number of subspecies. *Apis mellifera* is threatened by various factors for example, parasites, viruses and pesticide use. The native subspecies of the British Isles *Apis mellifera mellifera* is additionally threatened by introgression resulting from the importation of foreign subspecies perceived to have more desirable characteristics. This thesis investigates the integrity of *Apis mellifera mellifera* in the UK and how that can be assessed.

Initially, introgression levels over time in a conservation program were assessed. This revealed temporal decreases in introgression, indicating that some bee keeping management strategies can be effective in the conservation of this subspecies. Next, an assessment of different methods for measuring introgression in a social insects such as honey bees revealed that pooled colony approaches can be a powerful tool in the assessment of colony level introgression. Following this, an assessment of the status of *Apis mellifera mellifera* in the wider British Isles was performed using whole genome data. This revealed a largely introgressed population across the British Isles with the exception of a few key locations: Ireland, the South West of England, the Inner Hebrides, Northern Scotland, the Isle of Man, and Jersey. Finally, an examination of the 'purity' of subspecies in a phenotypic monitoring project and an assessment the genomic differences between subspecies highlights the importance of genetic assessment in scientific studies and indicates regions of high differentiation between subspecies. This work has implications for introgression

assessment in social insects and future directions for *A. m. mellifera* conservation in the British Isles and further afield.

Contents

List of Figures

List of Tables

1. General introduction

1.1 Background

Three topics that still evoke debate in evolution and ecology are introgression, local adaptation, and the concept of 'species'. Introgression is the transfer of alleles between genetically distinct lineages via hybridisation and repeated backcrossing (Anderson 1949). Hybridisation is the production of offspring via the interbreeding of genetically distinct lineages and backcrossing is mating back into either parental lineage or species (Anderson 1949; Futuyma and Kirkpatrick, 2017). Introgression is known to have important consequences for species and populations in terms of long term viability (Rhymer and Simberloff, 1996; Winger, 2017; Burgarella, *et al.,* 2019). It may promote adaptive potential through augmenting genetic diversity (Hedrick, 2013; Suarez-Gonzalez, *et al.,* 2018; Barbato, *et al.,* 2020), but, conversely, it can disrupt adaptation via the breaking up of co-adapted gene complexes (Rhymer and Simberloff, 1996; Lawson, *et al.,* 2017). Local adaptation can be simply defined as a population having a higher fitness at its native site than any other population of the same species introduced to that site (Williams, 1966). How local adaptation is measured and defined in different biological systems, under what circumstances it arises, and the interplay of the different evolutionary forces involved are widely debated in the current literature (Blanquart, *et al.,* 2013; Kawecki and Ebert, 2004; Tigano and Friesen, 2016). Finally, discussions regarding species concepts have been ongoing for centuries (Mayer 1982; Phillimore and Owens 2006). Specifically relevant to this work is how subspecies are defined and how that definition sits

alongside other terms such as evolutionary significant units (ESU), ecotypes or geographical races (Patten 2009; Ebach and Williams 2009).

These three topics are central to this thesis and are clearly inter-related. For example, at what level of introgression is local adaptation affected? When does introgression benefit populations and individuals through increasing genetic diversity? Do we define subspecies based on the occurrence of locally adaptive traits or are those ecotypes?

1.2 Introgression

Introgression has two contrasting consequences for the species or lineages involved. On the one hand, introgression could be damaging to a lineage or species as it may disrupt local adaptation via the breaking up of co-adapted gene complexes that have been built up over time through the action of natural selection (Mallet, 2005; Currat *et al.,* 2008). Additionally hybridisation and introgression can lead to genomic extinction, when hybrids or introgressed individuals replace both or one of the parental lineages and the intact parental genome no longer exists (Allendorf and Luikart., 2009; Epifanio and Philipp, 2000). Introgression and hybridisation also have the potential to lead to outbreeding depression (Edmands, 2007; Frankham *et al.,* 2011). Outbreeding depression occurs when genetically distinct lineages or species interbreed and the fitness of their offspring is lower than those of either parent (Freeland, *et al.,* 2011). On the other hand, introgression could increase the persistence of populations as it increases genetic diversity via the introduction of new alleles, and therefore boosts the adaptive potential of the population or individual (Hendrick 2013; Burgarella *et al.,* 2019).

1.2.1 Introgression as a disruptive force

The negative impact of introgression has been documented in several cases. Red-legged partridge, *Alectoris rufa*, are a good example as populations across the Iberian Peninsula and throughout France and Italy have experienced introgression from the previously imported Chukar partridge, *Alectoris chukar,* (Madge and McGowan 2002, Barbanera *et al.,* 2005; Barilani *et al.,* 2007). Hybrid partridges, (F1) individuals, are easily identified using morphological characteristics (Wilkinson, 1987; Wilkinson, 1991) and have been confirmed using molecular data (Barbanera, *et al.,* 2005). Introgressed individuals that are the result of backcrossing, were identified using molecular methods, specifically using random amplified polymorphic DNA (RAPD) markers (Barbanera *et al.,* 2005; Negro, et al., 2001). Non-introgressed red-legged partridge were no longer thought to exist in Italy and were thought to be at a high risk of extinction (Barbanera, *et al.,* 2013) until a remnant population was found in northwest Italy (Tizzani *et al.,* 2013). The loss of 'pure' individuals in the population means that the intact genome of this species in this location no longer exists. Even though the species may exist elsewhere, this population likely no longer contains any individuals with a genome that possess the full suite of co-adapted gene complexes developed under the local environmental conditions. This loss of local adapted gene suites jeopardises the survival of this local population and if there are no 'pure' individuals left, this is potentially an example of genomic extinction. Fitness, survival rate and breeding performance were compared in 'pure' *A. rufa* partridges and *A. rufa* partridges that show introgression with *A. chukar* (Casas, *et al.,* 2012). Introgressed partridges have lower overall survival rate than red-legged partridge (Casas, *et al.,* 2012). This lower adult survival rate is largely due to higher predation rates, though

the exact reasons for the increased vulnerability to predation remains unknown (Dávila, *et al.,* 2019). Conversely, despite this lower adult survival rate the introgressed population persists due to a constant influx of introgressed partridges that are released every year from game farms (Blanco-Aguiar, *et al.,* 2008) and potentially because introgressed individuals on average lay eggs in larger clutch sizes (Casas *et al.,* 2012).

Another documented case of disruptive introgression is that of the wildcat, *Felis silvestris*. *Felis silvestris*, the only remaining native felid in Great Britain, was once widespread but is now restricted to Scotland where it exists as an isolated population (Sainsbury *et al.,* 2019). These Scottish wild cats have introgressed with domestic cats, *Felis catus* (Macdonald *et al.,* 2004; Macdonald *et al.,* 2010). In the current wildcat population introgression is ubiquitous (Senn *et al.,* 2018) and no individuals without any hybridisation or introgression have been found (Sainsbury *et al.,* 2019; McCombe *et al.,* 2021). A continuum of genetic backgrounds shows that repeated hybridisation and backcrossing have taken place (McCombe, *et al.,* 2021). Although no studies have been able to assess the implications for the fitness of individuals, estimates predict that at the current rate of introgression, via the continued crossing with domestic cats, this species is now at risk of complete extinction (McCombe, *et al.,* 2021; Breitenmoser, *et al.,* 2019).

1.2.2 Introgression as a driver of diversity and adaptation

Adaptive introgression is when the transfer of alleles from one lineage or species to another results in increased fitness for the recipient and the transferred alleles are maintained by selection (Burgarella, *et al.,* 2019; Suarez-Gonzalez, *et al.,* 2018).

There are examples of adaptive introgression in the wild from naturally occurring hybrid zones: locations where genetically distinct populations of the same species or divergent lineages make contact and interbreed (Barton and Hewitt, 1989). For example, the saltmarsh sparrow (*Ammospiza caudacuta*) and Nelson's sparrow (*Ammospiza nelson*) both occur in tidal marshes along the north east coast of North America and Canada (Walsh *et al.,* 2018). The saltmarsh sparrow is a marsh specialist, while the Nelson's sparrow occupies a broader ecological niche that includes grasslands and brackish marshes as well as tidal marshes (Greenlaw, 1993). Walsh *et al.* (2018) examined both species in allopatric and sympatric populations using whole genome sequencing. Introgression was found in the sympatric sparrows, but not in the allopatric populations. Introgressed gene regions were linked to adaptive traits to tidal marsh environments such as osmotic regulation, responses to salt stress and responses to water deprivation (Walsh *et al.,* 2018). This study highlights the transfer of adaptive variation via introgression between two ecologically divergent species.

Another good example comes from *Heliconius* butterflies, a genus of neotropical pollinators that have been the focus of many studies because of their adaptive wing patterns and toxic chemical defences (e. g Turner, 1981; Jiggins, 2017). Sympatric species of these butterflies often possess similar wing patterns and are Müllerian mimics of one another; similar wing patterns have evolved as shared warning markings and individuals with similar patterns therefore benefit from greater predator protection. Adaptive introgression has played a crucial role in wing patterning in *Heliconius* butterflies, particularly between *Heliconious melpomene* and *Heliconious timareta*. The alleles transferred via introgression and hybridisation have

a fitness advantage in the form of protective wing colour patterning (Dasmahapatra, *et al.,* 2012; Pardo-Diaz, *et al.,* 2012).

1.3 Local adaptation

1.3.1 The occurrence of local adaptation

Local adaptation occurs when a population evolves genotypes that are selected for by the regional prevailing conditions, providing an advantage for those individuals with those genotypes in their specific habitat (Blanquart, *et al.,* 2013). Initially, for local adaptation to arise the existence of a spatially varied environment is required that creates a differing selective pressure, resulting in 'genotype × environment interactions' (Kawecki and Ebert, 2004). An important aspect of these interactions is that the same alleles have different effects on fitness in different environments and, as a consequence, no individual genotype is superior in all environments (Blanquart, *et al.,* 2013; Kawecki and Ebert, 2004; Yeaman and Otto, 2011). This adaptation to different environments leads to trade-offs. Effectively, natural selection exerted by the local habitat acting upon genotypes is the primary evolutionary force shaping local adaptation although there are other evolutionary forces that also have an impact as discussed further below. Importantly, natural selection affects allele frequencies, and consequently genetic variation, in different ways (Li, 1997). Stabilising selection, when natural selection favours the average phenotype, selects against phenotypic extremes and results in a decrease in genetic variation. Directional selection, when natural selection favours a particular genotype, simultaneously selects against other genotypes and reduces genetic variation. Diversifying selection is when two or more genotypes are favoured and this usually increases genetic variation (Freeland, Kirk

and Peterson, 2011; Futuyma and Kirkpatrick, 2017). Local adaptation can occur in different scenarios and as a consequence can have different effects on genetic diversity.

1.3.2 Local adaptation, gene flow and genetic drift

Gene flow affects the extent of local adaptation (Blanquart, *et al.,* 2013; Kawecki and Ebert, 2004) because restricted gene flow into the local population allows advantageous genotypes to be maintained (thus enabling local adaptation). In contrast, high gene flow can disrupt local adaptation, as external genotypes enter the local population and break up combinations of alleles favourable in the local habitat. Local adaptation can be maintained in the presence of gene flow but it is inevitably dependent on the relative strength of selection and extent of gene flow (Tigano and Friesen, 2016; Yeaman and Otto, 2011).

A good example of how gene flow can break down or interfere with local adaptation comes from farmed salmon (*Salmo salar*) (Karlsson *et al.,* 2016). Farmed salmon have been subject to strong artificial selection for traits that are commercially important, such as growth. Consequently, farmed salmon are both genetically and phenotypically different to wild salmon (Christie, *et al.,* 2016). When farmed salmon are released or escape they interbreed with the local population and there are now well-documented instances where this has caused a loss of local adaptation in local wild salmon populations (Bourret, *et al.,* 2011; Bolstad, *et al.,* 2017). For example, Bourret *et al* (2011), sampled wild and farmed salmon captured in the Magaguadavic River (New Brunswick, Canada) between 1980 and 2005. A comparison of genetic differences using F_{ST} genome scans revealed that the number of divergent loci reduced over time. There has been additional evidence of changes in size and age at

maturation in wild populations that have introgressed with farmed salmon (Bolstad, *et al.,* 2017).

Evidence for reduced local adaptation in the face of gene-flow has also been observed in Roesel's bush crickets (*Metrioptera roeselii*). Larger body size in ectothermic organisms confers a higher fitness in colder climates (Crushman, *et al.,* 1993; Van der Have, 1996). Populations in the Baltic Sea coastal area were examined using latitude-matched paired sites. Isolated populations were found to have larger body size in higher latitude, as predicted. However, those populations with high gene flow to lower latitude populations showed no such adaptation (Cassel-Lundhagen, *et al.,* 2011).

Equally, there are examples of adaptation occurring despite gene flow being present. A study examining selection pressure on the common frog (*Rana temporaria*) is a good illustration of how strong selective pressure can still result in local adaptation despite high gene flow (Muir, *et al.,* 2014). *R. temporaria* shows local adaptation to altitude in larval fitness traits the form of larval period (the number of days spent as a tadpole, from hatching of the egg to metamorphosis) and growth rate (Laugen, *et al.,* 2003; Palo, *et al.,* 2003). In Scotland *R. temporaria* is found from zero to over one thousand meters above sea level and has a continuous habitat and distribution. This provides a good study system to investigate adaptation with gene flow (Muir, *et al.,* 2014). The highest sites (≥900 m) had a significantly shorter larval period and higher growth rate, allowing individuals to grow more quickly without having to undergo metamorphosis at a smaller size. This larger metamorphic weight leads to increased adult survival (Altwegg and Reyer 2003). Differing summer and winter temperature parameters associated with altitude were observed to exert such a

strong selection pressure on the larval stage that adaptation occurs even in the face of gene flow (Muir, *et al.,* 2014).

An additional example of adaptation with gene flow is in deer mice (*Peromyscus maniculatus*) in Nebraska, USA (Pfeifer, *et al.,* 2018). In response to predation from avian species, mice in the Sand Hills area have evolved a light colour coat that matches the lighter soil in that habitat when compared to the surrounding areas (Linnen, *et al.,* 2013). Gene flow between the lighter and darker coated mice is high but the strong selection pressure of predation maintains this colour morph in the local Nebraska habitat (Pfeifer, *et al.,* 2018). This is an example of diversifying selection, where more than one genotype is being selected for and generally leads to an increase in genetic variation (across the area). In contrast, when local adaptation happens in isolation by directional selection, this would generally reduce diversity at loci.

Understanding gene flow in the study of local adaptation of *A. mellifera* is essential for two reasons. Firstly, in honey bees virgin queens mate with multiple drones from the surrounding area at drone congregation sites. Drones have been seen to fly an average of 3.75km to mate (Utaipanon, *et al.,* 2019), additionally, queens will also fly to mate with drones and it is thought that 7km from the colony would be the average distance of mating, if queens are assumed to have a mating flight distance similar to drones (Utaipanon, *et al.,* 2019). This creates gene flow between apiaries and more broadly across populations. Secondly, common UK bee keeping practice is to import and export honey bees, resulting in honey bees commonly being transported across and between countries as bee keepers have a preference for certain subspecies with perceived more desirable traits. This trade means colonies and queens from foreign subspecies are traded across large geographical scales,

creating gene flow between imported and native subspecies (De La Rua, *et al.,* 2009). An added complication is that bee keepers have different preferences for the characteristics of their bees, some place more emphasis on increased honey or pollen yield, others on disease resistance. These different requirements create different regimes of artificial selection even within the same locality, which interferes with natural selection and potentially also on local adaptation.

Another important factor affecting local adaptation is genetic drift (Blanquart, *et al.,* 2013; Yeaman and Otto, 2011). Genetic drift has different effects on local adaptation depending on the population size. Since drift can overcome selection when population size is very small, it can reduce the potential for the maintenance of local adaptation by reducing combinations of alleles with additive effects (Yeaman and Otto, 2011). As alleles with large effects on fitness are less likely to be lost by drift (Tigano and Friesen, 2016) loci with large effects on fitness should have a greater contribution to local adaptation (Kawecki and Ebert, 2004; Tigano and Friesen, 2016). In other words, in a small population, traits that are underpinned by many alleles of small effect are potentially more likely to be lost by drift when compared to a trait that is underpinned by a single allele of large effect.

Rather than selection acting on different genetic variants leading to different genotype frequencies in different locations (local adaptation), phenotypic plasticity can be another evolved response. Phenotypic plasticity is the ability of one genotype to produce different phenotypes under different environmental conditions. Both local adaptation and phenotypic plasticity can occur in combination and disentangling these effects has become ever more important to help us understand how species adapt to environmental change (Gao *et al.,* 2018). In common frogs, Muir *et al.* (2014) not only investigated local adaptation, but tested these population for

phenotypic plasticity by performing a common garden experiment. Eggs were collected from the populations at different altitudes and placed in a range of air temperatures. All populations showed phenotypic plasticity, however, high altitude individuals greater plasticity than the lower altitude individuals.

1.3.3 Local adaptation and genetic variation

While these evolutionary forces dictate the extent of local adaptation, adaptive potential – including the potential for local adaptation – requires available sources of genetic variation (Barrett and Schluter 2008; Hedrick, 2013). The original *de novo* source of all adaptive genetic variation is mutation; however, adaptive potential can also rest on standing genetic variation or adaptive introgression. Standing genetic variation simply refers to pre-existing alleles present in the population and reflects the population's previous effective population size. Local adaptation resulting from standing variation is predicted to be quicker as selection is acting upon alleles already present in the population due to the higher initial frequencies of the alleles (Hedrick, 2013; Hermisson & Pennings, 2005). Gene flow can also enhance standing genetic variation (Tigano and Friesen, 2016). New mutations are the original source genetic variation (Tigano and Friesen, 2016) and can be in the form of point mutations (single nucleotide polymorphisms, SNPs) or larger scale changes such as inversions or gene duplications (Stapley, *et al.,* 2010). Adaptive introgression (when the transfer of alleles from one lineage or species to another results in increased fitness for the recipient) is increasingly recognised as an important factor during evolutionary change, and is discussed in more detail below.

1.3.3.1. Adaptive introgression

Introgression, as discussed in section 1.1.2, is also a source of genetic variation that can drive adaptation. Adaptive introgression was first acknowledged in plants when Anderson (1949) recognised that the genetic variation resulting from hybridisation and subsequent introgression would exceed genetic variation created by new mutations. There are many examples of adaptive introgression from crop studies. For example, hybridisation and subsequent introgression between sunflower species *Helianthus annuus* and *Helianthus debilis* is thought to have enabled adaptation of herbivore resistance traits (Whitney *et al.,* 2006) and abiotic tolerance traits (Whitney *et al.,* 2010). Adaptive introgression has also been documented between iris species (*Iris fulva* and *Iris breviculis*) in traits relating to flood tolerance (Martin *et al.,* 2006). There is also evidence of adaptive introgression from cultivated crop to wild relatives, For example, alleles from maize (*Zea mays*) cultivars have transferred to a wild relative, annual teosintes, in Spain and France. Le Corre *et al* (2020) used genome wide SNPs to investigate the extent of admixture between cultivated maize and teosintes and tested whether local maize varieties have contributed to teosintes adaptation in Europe. They reported that introgression in the genomic region that contains the gene ZCN8, which is responsible for flowering time, had opened a new niche for the weeds in Europe and, as a result of adaptive introgression, caused a new agricultural weed to emerge (Le Corre, *et al.,* 2020). They also noted that hybridisation is still ongoing between the two as hybrid-like plants are still regularly observed in France and Spain.

Examples of adaptive introgression in plants and crops can provide a good analogy for honey bees. Plants have undergone the process of domestication where humans have selected for traits for their own benefit, converting wild species into crop plants

(Olsen *et al.,* 2013). Despite this domestication, these domesticated lines can still exchange genetic material with their wild counterparts. Honey bees are referred to as semi-domesticated or domesticated (Aizen and Harder 2009; Grupe and Quandt 2020) and undergo strong trait selection by beekeepers (Ibrahim 2007; Morfin *et al.,* 2020; Wragg *et al.,* 2016) while still being able to mate with feral or wild populations around them. An important difference, however, is that mating in honey bees is difficult to fully control compared to the situation with fully domesticated animals such as cattle where breeding is wholly managed.

1.4 Domestication

Domestication is not purely the realm of human activities and has evolved in the natural world more than once, for example the cultivation of fungal species by ants, ambrosia beetles and termites (Mueller *et al.,* 2005), but by far the most frequent domesticator is humans. Humans have domesticated animals and plants for food, materials, to perform tasks (e.g. guide dogs, cart horses or carrier pigeons) and for companionship (Meyer and Purugganan, 2013). But how does domestication affect adaptability, traits, and genetic diversity?

All species domesticated by humans are the result of selection, conscious or unconscious, from wild ancestors (El-Kassaby, 1992; Olsen and Wendel, 2013). Between fully domesticated species and their wild relatives there is a domestic classification referred to as landraces (LRs) (Frankel, 1970; Zeven, 1998). LRs are the result of primitive domestication where wild species have been selected for tolerance to the environment or pest and pathogen resistance, not for productivity. Varieties of plants can also be LRs. Varieties are locally adapted populations while cultivars are artificially bred. LRs provide a reservoir of genetic resources that can be

actively managed to improve the domestic traits that are desirable and adapt the cultivars as needed (Notter 1999; Toro *et al.,* 2009). Centuries of local selection in different environments has led to the evolution of an enormous diversity of gene complexes in LRs and was regarded as an outstanding characteristic of these varieties (Frankel and Soulé, 1981).

Not only are varietal populations and land races an important genetic resource for crop plants, but this is the case for animals, too. Despite this, the selection of highly productive crop cultivars and animal breeds can either present a threat to these historic genetic resources or the local varieties can be used to improve standard breeds. For example, Holstein-Friesian cattle are highly productive dairy cattle originally from Europe. Because of their high productivity, improved transport systems and easy distribution of genetic resources via artificial insemination, such breeds have replaced the use of local breeds (Groeneveld, *et al.,* 2010) and Holstein-Friesian cows are now globally distributed, including in East Africa (Food and Agriculture Organisation., 2007). When these cows were imported to Kenya they were unaccustomed to the heat, local forage and diseases (Bohmanova *et al.,* 2007). To address these issues Kenyan farmers bred the imported Holstein-Friesian cattle with cattle from the local area (Kim and Rothschild, 2014). The farmers used the locally adapted population to improve the cattle that have a desirable trait of high milk yield. This example perfectly highlights the importance of conserving locally adapted breeds and wild non-domesticated relatives.

Domestication generally reduces genetic diversity as traits targeted by breeders are selected for and the genetic regions determining those traits, thus become less diverse reducing variation across the genome as a whole (Flint-Garcia, 2013). The domesticated cultivar or breed becomes more popular as its productivity aspects

outperforms the LRs, so it begins to replace them (Frankel, 1970). This can result in the loss of some LRs (e.g. Wade 2012, McLean-Rodríguez, *et al.,* 2019) further reducing genetic diversity and in turn jeopardising the future adaptive potential of the domestic breed or cultivar.

There is a parallel here to honey bee subspecies since replacement of one honey bee subspecies with another due to more desirable or commercial traits is well documented (Nielsdatter, *et al.,* 2021). In two German provinces, for example, the local honey bee subspecies *Apis mellifera mellifera* was completely replaced by another subspecies *Apis mellifera carnica* due to the preferred traits of *Apis mellifera canica* (Kauhausen-Keller and Keller, 1994; Maul and Hähnle, 1994).

1.5 Classifications and definitions of evolutionary and conservation units

Central to the topic of the evolution of local varieties is the designation and classification of varieties, races, subspecies and species. Debate over the definitions of these has been, and remains, ongoing (Hey, 2001; Lowry, 2012). Starting at the species level, the species concept is a prime example of these debates with approximately 24 definitions and no universal consensus (Hey, 2001). For example, the Biological Species Concept, which currently defines species as groups which interbreed in natural populations and are reproductively isolated from other such groups (Mayr 2000). This definition has been adjusted over the years (Mayr, 1942, 1982, 1996), cannot be applied to some situations where its main criteria of reproductive isolation cannot be tested (fossils, or asexual organisms). But while the species concept is still being wrestled with, additional issues arise when we attempt

to classify variation within species. This issue of within-species variation is of particular relevance in conservation biology: when is a population or taxon sufficiently distinct to warrant conservation in its own right? And how do we conserve enough diversity to allow evolution to occur unconstrained? Many terms have arisen in an attempt to classify the variation below the level of species, for example, subspecies, evolutionary significant units, management units, ecotypes, or geographical races (Mayr, 1942; Mortiz 1994; Lowry, 2012). Debate has been ongoing around the definitions and usefulness of terms, partly because they can have significant implications that go beyond just taxonomic classifications (e.g informing conservation strategies). While a consensus on the definitions and uses of terms has not been reached in the literature, and resolving these definitions is beyond the scope of this thesis, relevant terms for the study system at hand and how they are used are reviewed below.

1.5.1 Lineages

A lineage has been defined as a series of entities descended from a single ancestor over time (Simpson, 1961; Hull, 1980). A lineage can represent entities at many levels, for example: genes, organelles, cells, organisms, or populations can all have lineages (de Queiroz, 1998; Hull 1980). Most commonly, a lineage represents the evolutionary history of an organism and these are usually represented in phylogenetic trees (de Queiroz, 1998, 1999, 2007). A clade or monophyletic group consists of several lineages (Figure 1.1). These are sometimes used interchangeably in the literature. Lineages have been linked to a related species concept and while species concepts are complicated and unresolved, here it will be briefly touched upon.

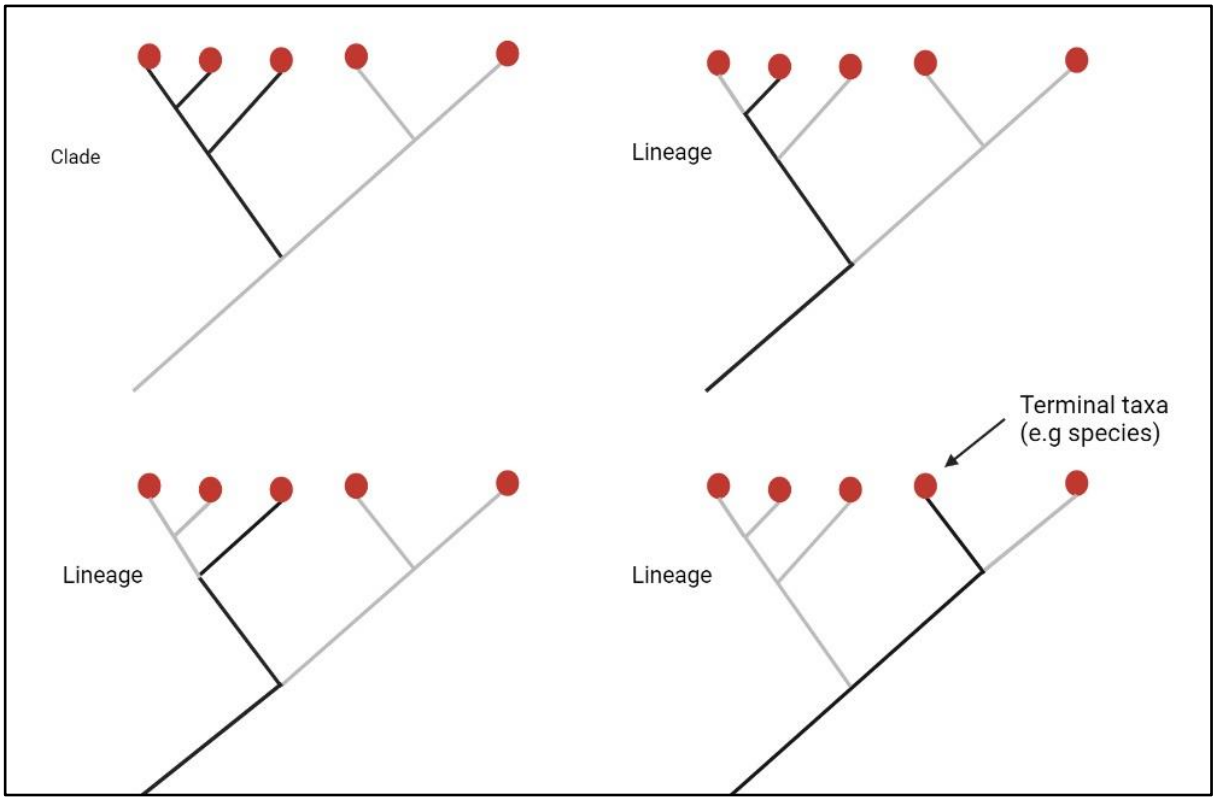


Figure 1.1 Lineages (bold lines) within clades displayed on a phylogenetic tree. Lineage is an ancestral descendant population.

The lineage species concept states that species are ancestral descendant population separated from other such groups (de Queiroz 1998; Henning, 1950; Ridley 1989; Simpson, 1951). When speciation occurs over time a species diverges into to two separate lineages producing two descendant species (Henning, 1950; Ridley 1989). Proponents of this species concept propose that a species is a single unbranched population sequence of ancestor-descendants through time (de Queiroz, 1998). The problem with this hypothesis is that the current method to build phylogenetic trees is often solely via DNA analysis of contemporaneous individuals, therefore we need to sample all potential species to confirm lineages, which is not always possible (for example with fossils or rare species). Additionally, if hybridisation or interbreeding occurs between branches there could potentially be a lack of clear branching. Finally

like many concepts it is not applicable to all study systems. The discussion from here on will focus on definitions and concepts regarding variation within species and subsequent discussion will focus on this theme.

1.5.2 Subspecies

The subspecies concept arose in the early 19th century to document varieties, races or forms of a species based on morphological differences (Braby, *et al,* 2012). Mayr (1942, 1963) described subspecies as conspecific populations that differ morphologically, are geographically isolated, and represent incipient species. However, Mayr also stated that if the subspecies are part of a continuous population lacking a geographical barrier the subspecies classification is purely a taxonomic device (Mayr, 1942, 1963). Essentially, Mayr's definition is that true species and subspecies would develop in allopatry (Mayr, 1996).

Criticism of the subspecies concept emerged on the basis that the classification was often poorly applied and geographical boundaries can be difficult to define, particularly in parapatric populations (Wilson and Brown, 1953; Gillham, 1956). Additionally, allocating subspecies can be subjective, arbitrary, and based on too few morphological features (Mayr, 1982). Critics argued that the subspecies classification was more useful as a taxonomic sorting method rather than as a useful evolutionary unit (Mallet, 2007).

The advent of genetic methods enabled the re-examination of subspecies classified using the traditional morphological method. Often, results from genetic analysis differed from the previously classified subspecies (Burbrink, Lawson, Slowinski, 2000; Zink 2004). For example, the subspecies concept was first applied in ornithology (Winker, 2010) in an attempt to classify the variation seen within avian

species. As an example, this approach had classified the cactus wren (*Campylorhynchus brunneicapillus*) into six subspecies. Phylogenetic analysis using mitochondrial DNA (mtDNA) haplotypes revealed just two groups and these groups were inconsistent with modern subspecies boundaries (Zink, 2004). These inconsistencies between genetic analysis and the morphologically defined subspecies were also revealed in the North American rat snake (*Elaphe obsolete*). Originally this species was divided into eight subspecies based on colour patterning in adult snakes (Burbrink, *et al.,* 2000). Individuals were sampled from across the species range in the United States, and examined using two mtDNA genes. Four lineages were identified, three containing seven subspecies, and one subspecies (*E. o. bairdi*) forming its own lineage. The lineages corresponded to geographical locations occupied by the snakes (Burbrink, *et al.,* 2000). Further work, using 67 morphological measurements (Burbrink, 2001) and nuclear markers (Burbrink, Crother and Lawson 2007) eventually recognized the lineages as four separate species (Burbrink, *et al.,* 2021). This is a good example of how categorising subspecies based on one or a few morphological characteristics can lead to poor categorisation of species and subspecies.

1.5.3 Evolutionary significant units

The evolutionary significant unit (ESU) was conceived in an attempt to conserve evolutionary potential and genetic resources within species (Ryder 1986; Avise 1989). The term was first used by Ryder (1986) who defined the ESU as “a subset of the more inclusive entity species, which possess genetic attributes significant for present and future generations of the species in question”. Ryder (1986) offered no advice on how to assess the criteria he had set out, leaving the concept without a fully demonstrable concept. How to assess criteria for ESU designation is important

because different data can yield different results as evident in the above section (1.5.2, where snake colour led scientists to one conclusion and mtDNA to another). The concept was later developed to contain an adaptive element. In particular, Waples (1991) redefined an ESU as a population that is "substantially reproductively isolated from other conspecific populations' while it 'represents an important component in the evolutionary legacy of a species'. Again, the issues surrounding how to identify populations that meet the criteria for this definition was left wanting. However, Mortiz (1994) offered a testable definition for the ESU stating that ESU's are populations that "are reciprocally monophyletic for mtDNA alleles and demonstrate significant divergence of allele frequencies at nuclear loci". While this definition offered methods to test the definition it was not without criticism. Critics asserted that there is not one method best for phylogenetic tree construction and, no method allows for the most likely phylogeny in every situation (Waples, 1995). Also, critics highlighted that the criteria of reciprocal mtDNA could be too stringent: while it accounts for historical restrictions to gene flow (Crandall, *et al.,* 2000; Kizirian and Donnelly, 2004), it does not account for ecological differences. Crandall *et al.* (2000) illustrated the issue using spiders (*Nesticus* sp.) from the southern Appalachian Mountains, USA. These spiders previously had a continuous distribution at lower elevation, but changes in climate since the Pleistocene era resulted in a change in the spiders' distribution. Some spiders shifted to a higher altitude, surviving on north facing boulder fields and rocky gorges (Gertsch 1984; Coyle and McGarity 1991), resulting in a distribution on the tops of mountains (*Nesticus mimus*) with no gene flow between these mountain top populations. Other spiders survived by entering caves and are now cave-dwelling species (*Nesticus carolinensis)*. The cave spiders have adaptations consistent with cave-dwelling, such as increased leg length, a reduction in pigmentation and eye size (Crandall, *et al.,* 2000; Hedin, 1997). The cave

dwelling spider is only present in a small area and the populations were the result of colonisation from nearby the mountain top spiders (figure 1.2). The result of this is a paraphyletic relationship of the cave spiders to the mountain top spiders (figure 1.2). Even though the cave spiders are morphometrically and ecologically different to the mountain top spiders they would not meet the criteria of reciprocal monophyly and would not be classified as an ESU, despite the adaptations they possess (Crandall, *et al.,* 2000; Hedin, 1997).

Effectively this definition places an emphasis on historical demographic processes and could potentially ignore adaptive differences. Though there still remains no consensus on how to define an ESU, one thing that can be agreed on is that the overall goal of identifying an ESU is to maintain and protect evolutionary potential. Fortunately, the advent of next generation sequencing and genome wide methods can now access large numbers of both neutral and adaptive loci (Coates *et al.,* 2018). Assessment of both neutral and adaptive markers allows for the capture of demographic and functional information to be taken into account when making conservation decisions (Funk, *et al.,* 2012).

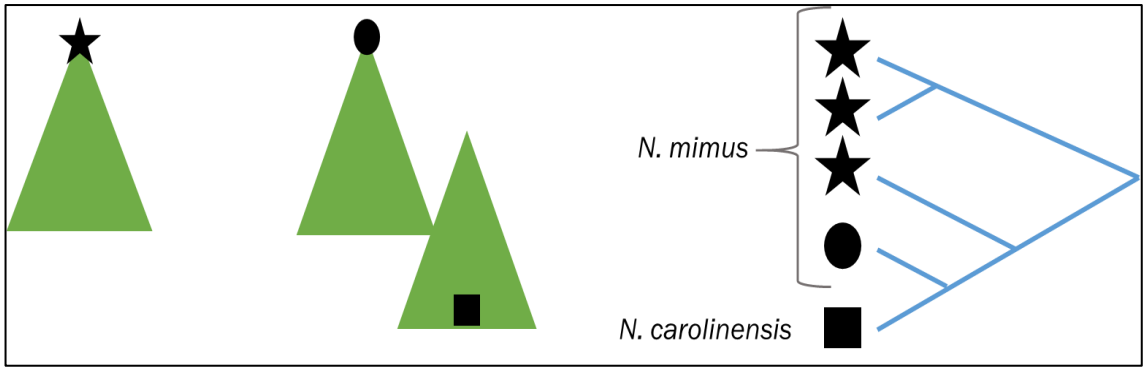


Figure 1.2 Crandall *et al* (2000) and Hedin (1997) mountain top and cave spiders

1.5.4 Ecotypes

The term ecotype was first used in relation to heritable traits in plants when Turesson (1922) investigated plants collected from inland and coastal habitats grown under the same set of conditions. The aim was to compare the effects of environmental factors on the genotype of the populations of the same species, where those populations originated from different habitats. The experiments revealed that plants from different habitats varied from each other in multiple traits (Turesson, 1922). To refer to the plant populations that exhibited these trait differences the word ecotype was used and was defined as ecologically distinct populations (Turesson, 1922). Ecotypes were seen to be the result of natural selection and evidence of adaptation to local environments (Turesson, 1925). Some researchers saw this difference between populations as demonstrating the importance of barriers to gene flow in the process of speciation, as the populations were far enough apart not to interbreed, and a new definition emerged. Later, Gregor *et al*. (1936) provided a definition that specifically mentions reproductive isolation, asserting that an ecotype is 'a population distinguished by morphological and physiological characteristics, most frequently of a quantitative nature' and 'interfertile with other ecotypes and ecospecies, but prevented from exchanging genes by ecological barriers' (Lowry, 2012). This definition meant that while reproductive isolation was important, there is a prerequisite that they can but do not exchange genetic material. Clausen *et al* (1951) expanded on plant ecotype experiments confirming ecotypes in many species and the term 'ecological races' was first coined (Clausen, 1951). As the evidence of ecotypes and ecological races emerged further, evolutionary biologists began to factor it into their discussions around the process of speciation. Dobzhansky saw these ecological races as 'populations of a species, which differ in the frequencies of

one or more genetic variants, alleles, or chromosomal structure' (Dobzhansky, 1937). He also acknowledged these ecological races were the same as the previously discussed ecotypes and described them as a fundamental stage of speciation (Lowry, 2012; Dobzhansky, 1940). Mayr, however, initially thought the discussions around ecotypes and ecological races had become taxonomic and too typological (Mayr, 1958). Mayr's opinions changed over time and he refered to species as consisting of geographical races and that 'all geographic races are also ecological races' and vice-versa (Mayr, 1947; Lowry 2012). The current definition of ecotypes offered by Lowry (2012) is that 'ecotypes, also known as ecological races or geographical races, are populations (or subspecies) that are adapted to an environment via variation in multiple traits and allele frequencies across loci' (Lowry, 2012). These discussions cemented the idea of ecological race and its potential role in ecological speciation in the literature (Lowry, 2012; Clausen 1951; Ortiz, 2020).

An example of ecotypes outside of the plant science discipline is the shore gastropod, the rough periwinkle (*Littorina saxatilis*) (Reid 1996). This ovoviviparous species has no pelagic larval distribution (Reid 1996; Rolan-Alvarez, *et al.,* 1998) and this limits gene flow between populations along coastlines. There are two recognised ecotypes of this species, a small shell phenotype adapted to withstand high wave exposure and a larger ridged shelled phenotype adapted to withstand crab predation (Johannesson, *et al* 2010). These two ecotypes occur in different habitats, either exposed or sheltered rocky beaches. Both ecotypes interact in areas where the habitats overlap, and have been genetically assessed in a number of studies (Butlin, *et al.,* 2014; Westram, *et al.,* 2014; Ravinet, *et al.,* 2016). This is a well-studied example of ecotypes and the forces that have limited gene flow between the two. Forces such as divergent selection (Johannesson, *et al* 2010), assortative mating

(Perini, *et al.,* 2020) and environmental selection (Rolan-Alvarez, *et al.,* 1998; Janson, *et al.,* 1983) are the barriers to gene flow between the two ecotypes where the habitats cross over. These ecotypes are phenotypically distinct, have barriers to gene flow and occupy distinctly different environments. So, what makes these periwinkle ecotypes 'ecotypes' and not subspecies or species, and would they qualify as an ESU? How is the rough periwinkle example different from the example of the cave spiders? Defining the real difference between a subspecies, an ecotype or an ESU is no easy task. Issues surrounding how units of the evolutionary process are defined will not be resolved in this thesis, but defining the key terms in the study system at hand is important for clarity.

1.5.5 *Apis mellifera* lineages, subspecies, ecotypes and ESUs

The honey bee, *Apis mellifera* has a large natural range spanning Africa and Eurasia. Across its range *A. mellifera* exhibits substantial variation and there have been lineages, subspecies and ecotypes designated (Meixner, *et al.,* 2013; (De la Rúa, *et al.,* 2009). However, with regard to the structure of lineages and subspecies a number of issues remain unresolved. In *Apis melifera,* taxonomic and evolutionary units were originally based largely on work by Ruttner (1988) who employed 36 morphological measurements to classify the species in Europe. Since then some have been confirmed and revised with the addition of genetic data (Franck, *et al*, 2000; Dall'Olio, *et al.,* 2007; Rortais, *et al.,* 2011; Pinto, *et al*., 2014).

1.5.5.1. *Apis mellifera* lineages

Apis mellifera has been split into lineages each containing a number of subspecies and within those subspecies some ecotypes have also been defined (Ilyasov, *et al* 2020; Meixner, *et al* 2013; Dall'Olio, *et al.,* 2007). Assigning several subspecies

within a lineage strays from the usual pattern of classifications set out in the preceding discussion and additionally these evolutionary lineages have not been completely resolved.

Initially, *A. mellifera* was divided into 3 lineages, A, M and C, using morphological measurements, and within these 3 lineages approximately 20 subspecies were described (Ruttner, 1978; Moritz, *et al.,* 1986; Garnery *et al* 1992). Subsequently, using a combination of morphological and genetic analysis the species was divided into 4 lineages, A, M, C and O, containing approximately 24 subspecies (Figure 1.3) (Ruttner, 1988; Arias and Sheppard, 1996; Palmer *et al.,* 2000; Estoup, *et al.,* 1995).

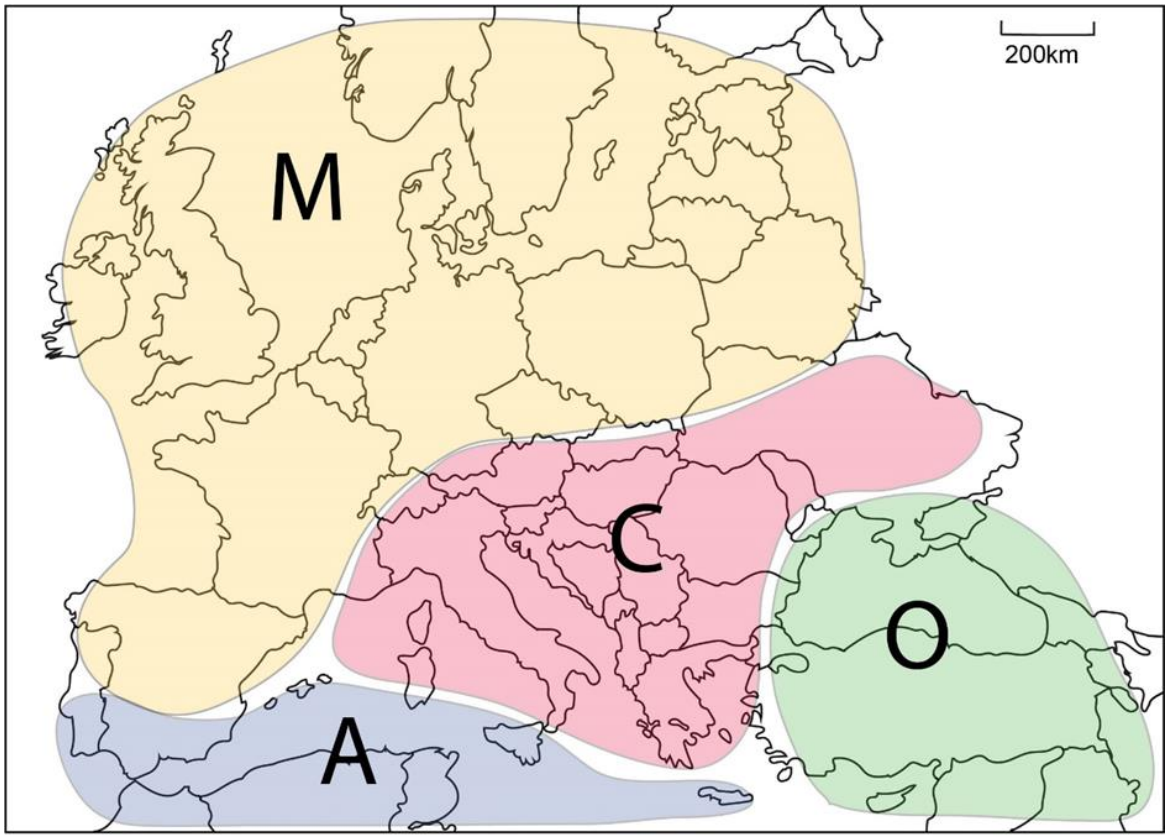


Figure 1.3. Distribution of the A, M, C and O *Apis mellifera* lineages across Europe

Currently, there are 5 or 6 lineages, A, M, C, O, Y, along with a sub-lineage Z, with more than 30 subspecies divided among them (Lineages Y and Z are located in Africa) (Alburaki, *et al.,* 2011; Meixner *et al.,* 2013; Ilyasov, *et al* 2020; Tihelka, *et al.,* 2020).

The composition of these lineages is still unclear. For example, in Europe there are two evolutionary lineages, lineage M and lineage C. The M lineage is generally described as containing three subspecies, *Apis mellifera mellifera, Apis mellifera iberiensis* and *Apis mellifera sinisxinyuan* (Mexiner, *et al* 2013; Ilyasov, *et al* 2020). In lineage C there are currently 10 subspecies, including two of the most commercially important subspecies, *Apis meliffera ligustica* and *Apis mellifera carnica* (Mexiner, *et al* 2013; Ilyasov, *et al* 2020). Tihelka *et al* (2020) attempted to clarify the phylogeny of this species by examining mtDNA from 18 subspecies. This resulted in the M lineage containing just two of the three subspecies, the northern European *A. m. mellifera* subspecies and the Chinese subspecies *A. m. sinisxinyuan.* This grouping of *A. m. mellifera* and *A. m. sinisxinyuan* is difficult to interpret as *A. m. sinisxinyuan* is found in the north west of China (Chen, *et al.,* 2020). Tihelka *et al* (2020) placed *A. m. iberiensis,* the Iberian peninsula subspecies, with lineage A. The A lineage's range is across North Africa. This is not unusual, previous studies using both nDNA and mtDNA found evidence for lineage A grouped in the southern range of this subspecies (Chávez-Galarza, *et al.,* 2015). An interesting question is whether these A lineage haplotypes are the result of natural contact zones, or the result of human transhumance; at present we simply do not know. Resolving the structure of the lineages will require wider sampling of subspecies from across the geographical range and more genetic information from across the genome.

1.5.5.2. *Apis mellifera* subspecies

With over 30 subspecies of *Apis mellifera* described worldwide, do these subspecies fit within the widely used definitions of a subspecies in the literature?

As discussed, Mayr (1942, 1963) suggested that to be classified as a subspecies geographical isolation is required. The distribution of *Apis mellifera* across Europe is

large and continuous, much of which lacks any geographical isolation (with the exception of some mountain ranges, for example the Alps). However, geographical boundaries change over time and it has been suggested that honey bee populations were previously more fragmented. Ruttner (1988) described honey bees retreating towards the Iberian Peninsula and the Balkan Peninsula during the last glacial period approximately 110,000 years ago, where they remained until approximately 12,000 ago when the glacial ice receded allowing the bees to recolonise Europe. The Iberian bees then recolonised northern and western Europe resulting in lineage M and the Balkan bees recolonised eastern and central Europe resulting in lineage C (Ruttner, 1988; Miguel, *et al* 2007). This geographical separation explains the formation of the lineages and but does not account for the current subspecies within the lineages. The geographical barrier criterion is unhelpful in clarifying subspecies delineation here: using that criterion few honey bees would be defined as subspecies. Nevertheless, subspecies have been assessed using a combination of methods.

1.5.5.2.1.Morphological categorisation of subspecies

In honey bee studies there are two broad methods for morphological categorisation. One is the method designed by Ruttner (1988) which uses 36 full body characteristics, and the other is wing morphology (Miguel, *et al.,* 2011; Tofilski, 2004; DuPraw, 1964). Wing morphology is used in a number of different ways. Bee keepers often employ two measurements, cubital index and discoidal shift, while researchers often use geometric wing morphometry which measures up to 41 wing vein angles (Tofilski, 2004; DuPraw, 1964; Miguel, *et al.,* 2011). As seen in the preceding discussion the use of too few morphological characteristics has led to subspecies being poorly defined. However many of the subspecies defined using Ruttner's 36 morphological traits have been confirmed using molecular data (See section

1.5.5.2.2). There has also been consistency between the use of the full suite of wing morphometry and genetic data (Miguel, *et al.,* 2010). Abdominal colour is sometimes used to identify subspecies by bee keepers. In the UK a black phenotype is used to identify the M lineage bee, *A. m. mellifera,* while in the Azores a yellow phenotype is used to select for *A. m. ligustica*. Henriques *et al* (2020) examined whether colour could be used as a proxy for C-lineage introgression in M lineage honey bees. They compared abdominal colour phenotypes to introgression values resulting from genetic data (SNPs). *A. m. mellifera* individuals with yellower phenotypes did exhibit a higher average level of introgression, however, the two most dissimilar phenotypes (black and mostly yellow) had overlapping introgression ranges. There were also several occasions where individuals with a black phenotype had higher introgression levels than individuals with most yellow phenotypes (Henriques *et al* 2020). Effectively, a phenotypically yellow individual may help identify introgression in *A. m. mellifera*, but a phenotypically black individual does not necessarily indicate that the individual has no introgression. This could be due to a number of factors: firstly, while colour is a heritable trait there is some evidence that during pupal development temperature can play a role in the colour of honey bees, with darker phenotypes developing at lower temperatures (DeGrandi-Hoffman, *et al* 1993; Spivak *et al* 1992). Secondly, there is only one yellow honey bee in Europe (Ruttner, 1988), which is *A. m. ligustica*. This could explain why the black phenotype is not necessarily indicative of *A. m. mellifera* but yellow tergites are potentially indicative of introgression. A phenotypically black honey bee with a high introgression value could be a result of introgression from *A. m. carnica* or another European subspecies that is not yellow (Henriques *et al* 2020).

Additionally some subspecies are based purely on bee keeper descriptions and morphological measurements, especially for subspecies outside of Europe. For example, many subspecies in Africa and the Middle East rely exclusively on morphometric analysis or bee keeper descriptions (Fotso Kenmogne, *et al.,* 2021; Khoshraftar, *et al.,* 2021; Hailu, *et al.,* 2021). For confirmation of subspecies identified via morphological methods genetic data is used (Miguel, *et al.,* 2010). Although genetic analysis also has its limitations regarding genetic marker choice, specialist skills are required and higher costs are involved for this kind of work.

1.5.5.2.2.Genetic categorisation of subspecies

Many methods have been used to identify subspecies, for example, mtDNA haplotype analysis, microsatellite or genome wide SNPs data coupled with Bayesian clustering algorithms. These have all allowed researchers to investigate the genetic differentiation and population structure exhibited by subspecies (Dall'Olio, *et al.,* 2007; Munoz, *et al.,* 2015) and subsequently allowed researchers to investigate introgression in populations (Ellis, *et al.,* 2018; Oleska, *et al.,* 2011; Franck, *et al.,* 2000).

While some subspecies have been genetically identified, for example *A. m. ligustica* (Dall'Olio, *et al.,* 2007), or *A. m. mellifera* (Pinto, *et al.,* 2014), many have not and some methods may not be very helpful in identifying where subspecies remain. Ellis *et al* (2018) examined *A. m. mellifera* (lineage M) populations in Cornwall, England, using mtDNA and microsatellites. Some samples possessed C lineage mtDNA while the nuclear DNA grouped with M lineage (above 0.95 probability of belonging to lineage M). This could potentially be due to the human-induced mixing of populations and means potentially using mtDNA to identify subspecies could be problematic (as

mtDNA is passed down only through the maternal line whereas markers such as microsatellites estimate nuclear divergence).

While genetic studies have begun to identify subspecies, many subspecies have never been sampled and the classification of a subspecies is largely unknown. Additionally human-mediated movement of honey bees can confuse matters further.

1.5.5.3. *Apis mellifera* ecotypes

While subspecies level traits are the result of adaptation to a general and wide set of ecological environments there is also variation within subspecies. For example, *Apis mellifera mellifera (A. m. m.)* has the largest original geographical distribution of any European subspecies. The range of *A. m. m.* stretches from Ireland in the west as far as Russia in the east, from Sweden in the north to France in the south. Across this range, there is reported variation within this subspecies where adaptations to local environmental conditions occur and this would potentially represent ecotypes (Meixner, *et al*., 2013; De la Rúa, *et al.,* 2009). There has been one ecotype of *A. m. m* described and reported in the scientific literature. In Landes, France, *A. m. m* has shown an adaptation to local flowering timings in the form of brood cycle timings (Louveaux *et al.,* 1966; Strange *et al.,* 2007). However, this is only one example of an ecotype within this subspecies and the adaptive characteristics of many honey bee subspecies remains poorly understood and largely unstudied.

Additionally, whether trait variation is due to genetic differences within and between ecotypes or the result of plasticity across the range is completely unstudied. Would the thirty subspecies across the range be better described as geographical races or ecotypes?

1.5.5.4. Evolutionary significant units in *Apis mellifera*

There has been no substantial discussion in the *Apis mellifera* conservation literature concerning ESUs, but previously described definitions of an ESU are based on three main ideas. ESUs are 'a subset of the more inclusive entity of a species, which possess genetic attributes significant for present and future generations of the species in question' (Ryder, 1986) or 'represents an important component of the evolutionary legacy of the species' (Waples, 1991) and 'are reciprocal monophyletic for mtDNA alleles and demonstrate significant divergence of allele frequencies at nuclear loci' (Mortiz, 1994). These attempted definitions and the wider idea of an ESU would fit the descriptions of the variation we see in *Apis mellifera* and could be a good description of what a subspecies in *Apis mellifera* represents. However, these ideas of ESUs did not arrive into the wider biological discussions until around the same time Ruttner's (1988) classifications had already been formed, so while there is no discussion of ESU's in the honey bee literature, the concept is still important and may be a relevant way to frame honey bee conservation.

1.5.6 Classification of *Apis mellifera* unit for this study

There are issues in defining subspecies or units for conservation and resolving these terms is a difficult task for biologists. It is also evident that there are many confounding factors in the study of honey bees regarding the lineages, subspecies, and ecotypes. In this study and in an effort to remain in keeping with the current literature around *Apis mellifera* lineage 'M' will be referred to as containing subspecies *A. m. mellifera*, *A. m. iberiensis* and *A. m. sinisxinyuan* and lineage 'C' containing the subspecies *A. m. carnica* and *A. m. ligustica.*

1.6 An overview of the study system

This thesis investigates the topics of introgression and adaptation in honey bees (*Apis mellifera*). Honey bees are a predominantly human-managed pollinator that are now distributed over every continent, save Antarctica (Requier *et al.,* 2019). Where *Apis mellifera* originated is still debated and has been suggested to be either Asia (Wallberg *et al.,* 2014), Africa (Whitfield *et al*., 2006) or northern Africa and the Middle East (Tihelka, *et al.,* 2020). Its native range spans Africa, Europe, Asia, and the Middle East (Ruttner, 1988; Requier *et al.,* 2019). *A. mellifera* is a eusocial insect belonging to the order Hymenoptera. *A. mellifera* are haplodiploid and arrhenotokous, where, males (drones) develop from unfertilized haploid eggs and, females (workers) from diploid fertilised eggs. *A. mellifera* live in large colonies of between 10,000 and 35,000 individuals (Eckert, *et al* 1994; Palmer and Oldroyd, 2001). This species has a matriarchal structure consisting of three bee castes: queens, workers and drones (Culliney, 1984). A female diploid queen is the only member of the colony that lays eggs (although there are some specific colony breakdown circumstances when workers may also do so). During the reproductive phase of the colony drones and new queens will be produced. These drones and queens produced by colonies will mate with drones and queens from the local area at drone congregation sites and go on to produce new colonies (Elekonich and Roberts, 2005). Queens will mate multiple times with up to 7-28 different drones from the surrounding area (Estoup, *et al.,* 1994; Palmer and Oldroyd, 2000).

A. mellifera has been categorised into different evolutionary lineages (Ruttner 1988; Dognatzis and Zayed, 2019; Tihelka, *et al.,* 2020) and each lineage contains a

number of subspecies (Ruttner 1988; Franck *et al*., 1998; Alburaki *et al.,* 2001; Chávez-Galarza *et al.,* 2017). In Europe *A. mellifera* has been grouped in to four evolutionary lineages, M, A, C, and O (Figure 1.3). As mentioned in section 1.5.5.2, Ruttner (1988) described how the last glacial period resulted in lineage M and lineage C. The O lineage is present in the Middle East while the A lineage is present in southern Spain and North Africa. Within these lineages are subspecies, for example, *Apis mellifera mellifera* sit within the M lineage, while *Apis mellifera ligustica* and *Apis mellifera carnica* are in the C lineage (Whitfield *et al.,* 2006; Meixner *et al.,* 2013). There have been over 27 honey bee subspecies described worldwide (Dogantzis and Zayed 2019).

1.7 *Apis mellifera* and its native status in the UK

Questions have been raised in the literature and the wider media about whether *Apis mellifera* is native to the UK (Goulson 2003; Carreck, 2008; Ollerton 2013). Below the history of *Apis* in Europe and evidence in relation to its native status in the UK is reviewed.

First the question of what 'native' means must be addressed, but the scientific community has not reached a consensus. Webb (1985) suggested that a native plant is one which evolved in Britain or arrived before the Neolithic period independently of human activity, suggesting that pre-Neolithic humans were part of nature and the effect they had on species dispersal was equivalent to that of other animals. However, it is now known that Mesolithic people's lifestyle did alter the environment, e.g. they may have cleared woodland using fire to attract grazing animals for hunting (Smout 2014). Other authors have tried to provide further clarity. For example, Usher (2000) suggested six categories to classify species (native, formerly native, locally

non-native, long established, recently arrived or non-native), while Crees and Turvey (2015) suggested a nine category system based on temporal periods and mode of colonisation. In short, ascertaining what species are native and is not as simple task.

Regarding honey bees, the oldest fossils representing the genus *Apis* were found in Europe. Fossils from France, Spain and Germany date back to the Oligocene (33.9 million years ago to 23 million years ago) and are morphologically similar to *Apis dorsata* (Engel, 1998; Kotthoff *et al.,* 2011; Kotthoff *et al.,* 2013; Nel *et al.,* 1999), confirming the presence of the ancestral *Apis* genus in Europe during this time. There are other *Apis* species fossils from Italy (Handlirsch 1907), Austria (Nel *et al.,* 1999), Czechia (Nel 2003), China (Hong 1983; Zhang 1990) and the United States (Engel *et al.,* 2009), yet there is only one *Apis mellifera* fossil. This single fossil dates back to the late Pleistocene (approximately 2 million years old) and is preserved in East African tree resin (Baker and Chmielewski, 2003; Zeuner and Manning, 1976).

Information from cave paintings across the world during Mesolithic period suggests that people practiced “honey hunting” rather than managing hives (Eva Crane, 2005; Garlake, 1995; Pager, 1973; Strickland, 1982; Gordon 1958; Mathpal, 1984). Indeed, the historical presence of honey bees in Southern Europe is also documented by the Araña cave painting (Bicorp, Spain), which depicts a Mesolithic figure (pre-Neolithic hunter gatherers) collecting honey from a wild colony (Dams, 1978; Kritsky 2014) and dates back to approximately 8000 years ago. The Mesolithic use of honey bees is interesting in relation to the question of whether honey bees are native to the UK because the land strait that connected the UK to mainland Europe, Doggerland, was shut off approximately 8000 years ago (Coles, 1998; Walker *et al.,* 2020) and we know Doggerland was home to generations of Mesolithic people for thousands of years (Gaffney, *et al* 2009; Amkreutz and Spithoven., 2019).

Carreck (2008) reviewed evidence that demonstrated the presence of honey bees in the British Isles for the last 4000 years. This includes archaeological evidence of bees wax inside approximately 4000 year old Neolithic pots located in Berkshire (Needham and Evans, 1987; Carreck 2008). Since then evidence of beeswax in Neolithic pots has been found in two additional sites in the south of Britain, at Eton rowing lake and at Bulford Torstone, dating from approximately 6000 years ago (Roffet-Salque, *et al.,* 2015). Clearly, this evidence contradicts arguments for introductions later than that date.

Conversely, an argument for *Apis mellifera* not being native is their close association with humans (Goulson 2003) and that, today, they are not often found living in the wild (Ollerton 2013), thus it is likely that honey bees were always an introduced semi-domesticated species. Humans and honey bees do share a close relationship, but does the eventual domestication by Neolithic people mean the honey bee is not native to Britain?

In summary, honey bees could have come over Doggerland, with or without the help of pre-Neolithic humans. Honey bees were present in Europe before the land bridge shut and there is evidence of the use of bees wax by Neolithic humans in Britain 2000 years after Britain was cut off. If hunter-gatherers did transport honey bees across Doggerland during the thousands of years it was open, these pre-Neolithic people could be seen as performing a natural process, similar to *Pheidole* and *Solenopsis* ants that farm mealybugs, *Dysmicoccus*, (Beardsley and Gonzalez-Hernandez, 2003) or *Macrotermitinae* termites domesticating fungus (Poulsen 2015).

1.8 Locally adapted *Apis mellifera*

Recently, bee keepers have become aware of the benefits and importance of locally adapted forms and are questioning the use of imported subspecies. It has been suggested that *A. m. m.* bees may possess a number of traits that are of advantage in UK conditions (Box 1). This suggestion raises questions not only about whether or not *A. m. m.* is locally adapted, but about how we measure and define local adaptation, and subspecies, and whether subspecies are a useful concept in honey bee conservation. How does human management or semi-domestication of honey bees affect traits or local adaptation? How do we assess introgression and admixture in this biological system, and, how do introgression and admixture affect local adaptation in honey bee subspecies?

1.9 Imports to the United Kingdom

Importation of honey bees is in to the UK is commonplace. This is thought to have stemmed from the population crash event that occurred in the 1920s, often referred to as 'the Isle of Wight disease'. The causes of the Isle of Wight disease have been attributed to several sources, from *Acarapis woodi* (Adam, 1968), *Nosema apis* (Fantham and Porter, 1912) or to a combination of many causes (Bailey, 1964). As a consequence, there was a need to replenish bee stocks and beekeepers resorted to importation from continental Europe, which presented an opportunity to purchase bees that possessed perceived desirable characteristics, for example, docility or increased honey yield. This importation culture persists today. Over the timeline of

this study the number of imports of batched queens, (a queen accompanied by a batch of workers) into the UK, has remained at a constant of approximately 15,000 queens a year (Beebase, 2021, BeeBase 2022). Since Britain's exit from the EU, imports can no longer consist of larger packages, such as nuclei or full colonies (DEFRA, 2021; Adam Parker, 2021). Despite this importation still persists and in 2022 so far 15,457 queens have been imported with high numbers of queens originating from Italy and Malta (BeeBase 2022). This support for non-native subspecies has led to a continued genetic influx in to local UK populations, which is thought to lead to the breakdown of combinations of alleles built up by local adaptation (Edmands, *et al.,* 2007; Harper, *et al.,* 2013). Currently, legislation in the UK allows for imports from European Union (EU) member states and New Zealand (Adam Parker, 2021). Imports from the EU may consist of queens with up to 20 workers accompanied by a health certificate (DEFRA, 2021).

A related issue is that the demand for honey bee queens is not met by breeders in the UK. When bee keepers need to replace new queens there are not enough honey bee queen breeders to meet demand (B4, personal communication). This is especially true of *A. m. mellifera*. Even though there are conservation groups in the UK none of these widely supply *A. m. mellifera* queens for purchase for the general public.

Studies of UK honey bee introgression present an opportunity to inform UK government, conservation bodies and beekeeper stakeholders regarding importation and breeding policy.

Box 1.
Apis mellifera mellifera specific traits

There are a number of traits that bee keepers consider to be *A. m. m* specific based on their experience. They are lacking empirical evidence.

Low tendency to swarm: Reduced swarming is a desirable characteristic for bee keepers. Swarming results in the colony splitting, where the old queen leaves with approximately 50% of the workers and a new queen stays with the remaining workers. Swarming is costly because possessing limited resources and labour, a swarm must create a new comb, build up new food reserves, and begin rearing a work force in order to replace and replenish the old workers.
Supersedure: the mating of a new virgin queen who then proceeds to lay eggs in the presence of the old queen, who does not relocate with a portion of the workers (Cooper, 1986). Essentially, it is the replacement of the queen by one of her daughters without swarming. This can be either the replacement or co-existence.
Brood cycle: The brood cycle of locally adapted ecotypes of *A. m. m.* colonies has been observed to be in phenological rhythm with the local flora (Louveaux *et al.,* 1966; Ruttner, 1988; Strange *et al.,* 2007). It is possible that imported subspecies of honey bee are automatically at a disadvantage as the brood timing and phenology has evolved and adapted to another locality. (see section 1.5.5.3)
Prolificacy: It is believed that some imported queens build brood and expand their colonies rapidly in the good weather and when the nectar is readily available. Conversely, when the weather becomes changeable, a frequent occurrence in the UK, in order to sustain worker numbers the imported colony consumes stores heavily leading to heavy loss of production and possible starvation. It is believed by *A. m. m.* bee keepers that as a result of being adapted to the UK's changeable climate, *A. m. m.* bees are conservative with stores and non-prolific in laying, even going off laying in the poor weather spells.
Drone brood production and cycles: Drones are produced at times of high resource abundance as drones do not contribute to the colony; they do not go out to forage, contribute to nursery duties or guard the colony and are therefore a costly expenditure for the colony to rear. Keepers of *A. m. m.* claim that these bees have different drone brood timing when compared with imported bees in the UK. It is thought that drone timings could be affected by the thrifty and non-prolific nature of native bees. If imported colonies build up more rapidly when compared with native colonies then they will reach the strength at which the colony can afford to produce the energy costly drones more quickly than *A. m. m*.
Pollen collection and thrifty nature: Bee keepers have reported that *A. m. m.* stores more pollen and are conservative (bee keepers use the term thrifty) with their pollen stores allowing them to survive periods of poor weather. Bees require a varied pollen diet to stave off vitamin, mineral and protein deficiencies (Schmidt *et al.*, 1995), in addition to this, poor pollen nutrition can result in a colony becoming more susceptible to disease and pesticide exposure (Fries, 1993; Brodschneider and Crailsheim, 2010; DeGrandi-Hoffman *et al.,* 2010; Wu, Anelli and Sheppard, 2011).
High longevity: Assertions that *A. m. m*. workers live longer than imported honey bee workers could be a beneficial trait due to the cost incurred by continually raising workers.
Increased flight strength in cold weather: Bee keepers have suggested that A. *m. m*. can forage in cooler climate such as the UK and northern Europe, unlike other subspecies.
Temperament: *A. m. m*. is said by some to be more aggressive than other subspecies. Bee keepers prefer honey bees to be calmer for ease of working the colony. Some suggest that this aggression arises from introgression between *A. m. m.* and other subspecies.
Heightened defensive behaviour: Wasps and hornets are significant enemies of honey bees and can invade and attack persistently, weakening colonies. *A. m. m.* is suggested to display increased defensive behaviour and so potentially be better equipped to ward off these predators.
Apiary vicinity mating (AVM): Sometimes referred to as alternate mating behaviour, AVM is often referred to by bee keepers who keep *A. m. m.* Ordinarily, honey bee queens mate at specific drone congregation sites (Koeniger *et al.,* 2005; Brockmann *et al.,* 2006) that form in hot sunny weather (Cooper, 1986; Widdicombe, 2015). AVM is when the queen mates with a number of drones that reside in the same apiary rather than locating a drone congregation area (Cooper, 1986). AVM is thought to be a strategy used by *A. m. m.* to breed in unsettled weather conditions when drone assemblies cannot form and imported queens will fail to mate.

Box 1. Traits of *A. m. mellifera.* that bee keepers have observed and suggested to be subspecies traits.

1.10 Conservation of UK Apis mellifera mellifera

There are no formal guidelines regarding the conservation or breeding of *A. m. m.* in England, Wales, and Northern Ireland. In Scotland, *A. m. m.* is protected by the Scottish Parliament solely on the Inner Hebridean islands of Colonsay and Oronsay via the Bee Keeping Order 2013 which came into effect on January the 1st 2014 and prohibits the keeping of bees other than *A. m. m.* on the island (The Bee Keeping Order, 2013).

The majority of *A. m. m.* conservation in the UK is in the hands of a network of dedicated hobbyist beekeepers, enthusiasts and a few commercial keepers who maintain *A. m. m*. bee stocks, some of which are lost when bee keepers withdraw from the pursuit. In Britain, the Bee Improvement and Bee Breeders Association (BIBBA) was founded in 1964 and they state the purpose of the group is conserving, restoring, studying, selecting and improving native and near native honey bees in Britain. BIBBA publishes guides and runs courses and conferences to inform the wider bee keeping community and the public of their mission. In Scotland the Scottish Native Honey Bee Society (SNHBS) is a charity encouraging the keeping, conservation, reinstatement, and scientific research of the dark honey bee.

1.11 Genetic studies on *Apis mellifera* subspecies

1.11.1 Mitochondrial haplotypes for lineage assignment

For the purposes of ascertaining honey bee lineage membership mitochondrial DNA (mtDNA) has been utilised extensively due to its maternal inheritance and a unique non-coding intergenic spacer located between the COI and the COII genes (Crozier,

Crozier and Mackinlay, 1989; Cornuet, *et al.,* 1991) that varies within and between honey bee subspecies. Also known as the tRNAleu and COII intergenic spacer, this intergenic spacer consists of variation in what are termed 'P' and 'Q' sequences (Cornuet, *et al.,* 1991). The P sequence has two forms P and P_o and the Q sequence is often repeated more than once although haplotypes vary in length (Cornuet, *et al.,* 1991). The C lineage contains a haplotype lacking the P sequence altogether with a single Q sequence (Garnery *et al.*, 1998). Other haplotypes in the M lineage, M4, M4' and M4'', contain a P followed by differing number of Q sequence repeats (Franck *et al.*, 2001). These mtDNA studies on honey bee subspecies often rely on a PCR-RFLP method using the restriction enzyme *Dra*I. Known as the '*Dra*I test', the method has been successfully used to study genetic diversity within lineages (Garnery *et al.*, 1998; Franck *et al.*, 2001; Jensen *et al.*, 2005). The DraI test does have its limitations as haplotypes can exist across subspecies within the same lineage. For example, the C1 haplotype is present in populations of both *A. m. carnica* and *A. m. ligustica* (Muñoz *et al.*., 2009; Meixner *et al.*., 2013). The A1 haplotype has also been shown to be present in both *A. m. iberiensis* and *A. m. adansonii* (Franck *et al.*, 2001; Ellis *et al.,* 2018).

Full sequencing of these mtDNA haplotypes allows for a more detailed investigation of honey bee diversity within lineages (Arias and Sheppard, 1996; Jensen *et al.*, 2005; Alburaki *et al.*, 2011; Pinto *et al.*, 2014; Ilyasov *et al.*, 2016), revealing deletions, insertions and single point mutations. For example, sequencing of the haplotype M7 revealed the presence of two separate variants present at different frequencies within two different populations (Franck *et al.*, 2000). Rortais *et al.* (2011) reviewed mtDNA haplotypes of *A. m. m.* populations utilising both the DraI test and sequencing, which revealed that the previously published haplotypes M14 and M26 were actually one haplotype.

1.11.2 SNPs for lineage assignment

Numerous genetic studies of *Apis mellifera* have been carried out using single nucleotide polymorphisms (SNPs) (e.g. Whitfield *et al.*, 2006; Chávez-Galarza *et al.*, 2013; Harpur *et al.*, 2014), and a number of studies utilizing SNPs have been carried out specifically investigating the *A. m. m.* subspecies in Europe (Pinto *et al.*, 2014; Ilyasov *et al.*, 2016; Parejo *et al.*, 2016). Frequently, these studies focus on the introgression and admixture of subspecies revealing the extent of hybridization amongst European populations. While the cost is continually decreasing for genome-wide SNP scans and next generation sequencing, it is still expensive to perform routinely. Genome wide SNP discovery has also recently been optimised for sub-species identification by the use of reduced sets of SNPs that have been shown to be the most informative for that purpose. Often referred to as 'ancestry-informative markers' (AIMs) (Muñoz *et al.*, 2015; Parejo *et al.*, 2016), these allow for accurate identification of subspecies and measures of hybridization at a reduced cost. Muñoz *et al.* (2015) used five analytical methods to rank 1183 SNPs to inform the combinations of SNPs that would create AIMs panels specifically to investigate admixture in *A. m. m.* from the C lineage honey bees. They then created panels consisting of 48, 96, 144, 192 and 384 AIMs, subsequently suggesting that the chosen AIMs were sufficient in accurately estimating introgression from the C lineage. Parejo *et al.* (2016) reported that as few as 50 SNPs were sufficient to assign honey bees as *A. m. m.* or other subspecies. A further investigation of the AIMs SNP panel allowed for a cost effective and ready to use assay that was developed on the iPLEX MassARRAY system and consists of the 144 SNP AIMs panel (Henriques *et al.,* 2018). The aim of this was to open up the platform commercially to bee keepers to help preserve genetic resources and safeguarding populations with unique alleles (Henriques *et al.,* 2018).

1.11.3 Signatures of selection

SNP based genome scans coupled with F_{ST} based outlier tests is one approach used to identify areas of the genome exhibiting signatures of selection (Zayed and Whitfield, 2008; Chávez-Galarza *et al.,* 2013; Wragg *et al.,* 2016; Parejo *et al.,* 2017). Chávez-Galarza *et al.* (2013) investigated signatures of selection in *A. m. iberiensis* (Iberian honey bee) and detected 69 outlier loci each identified by at least one of the four F_{ST} based outlier methods and 17 outliers detected by all four methods and were deemed the best candidates for loci under selection (Chávez-Galarza *et al.,* 2013). Of these 17 loci, 15 were strongly associated with one or more environmental variable. Gene ontology allowed some of the outlier's putative function to be investigated. Outliers associated with vision, xenobiotics, and innate immune responses were revealed (Chávez-Galarza *et al.,* 2013).

Parejo et al. (2017) investigated signatures of selection in *A. m. m.* from France and Switzerland, identifying eight genes on five chromosomes that showed evidence of selection. The WNT4 gene located on chromosome 1 exhibited a strong selection signature, which in *Drosophila melanogaster* has been associated with wing cross-vein differentiation and abdominal phenotypes. As bee keepers that keep *A. m. m.* often use both wing morphology (Tofilski, 2004; Oleksa and Tofilski, 2015) and abdominal colour as gauges of purity and select to breed from colonies using those criteria, the WNT4 allele is a potential candidate of human-mediated selection in honey bees (Parejo *et al.,* 2017).

There has to date been no comparative investigation into signatures of selection between different subspecies but it would be expected that their natural distribution would subject them to differing selection pressures. This highlights a difficultly in studying selection in this particular system. Honey bees are subject to both natural

and artificial selection pressures, and teasing the two apart is challenging, requiring a truly wild population for comparison (Kim *et al.,* 2016; Li et al., 2016). There are records of wild honey bee populations in Poland and Germany, and it is suspected that wild honey bees do exist in the Urals, France, Ireland and Italy but empirical data are lacking (Requier *et al.,* 2020). Further, if wild honey bees do exist and are within breeding range of semi-domesticated colonies it is likely that these signatures would still be intertwined.

1.11.4 Genome-wide association studies

Genome-wide association studies (GWAS) allow researchers to connect a phenotypic trait to its underlying associated genetic determination (if the trait is heritable). Such analyses are undertaken to identify alleles involved in traits expression or to determine the genomic architecture of the trait. GWAS generally work by correlating marker variants, usually SNP markers, with variation in the trait of interest (Bush and Moore, 2012).

GWAS experiments concerning honey bee traits have been used to assess various behaviours and phenotypes, such as the loci associated with hygienic behaviour linked to *V. destructor* resistance (Spötter *et al.,* 2016), royal jelly production (Wragg, *et al.,* 2016) and altitude adaptation (Wallberg *et al.,* 2017). Studies that take into account the subspecies of the honey bee being examined in the GWAS are less common but there has been one that explored calmness and gentleness in *A. m. m.* located in Switzerland (Guichard *et al.,* 2021). A total of 5 loci were identified and associations were found with genes LOC409692, the Abscam gene, a gene upstream of LOC102655631, with LOC413669 and LOC413416. These genes are mainly involved in determining the olfactory and nervous system of honey bees (Guichard *et al.,* 2021).

Importantly, the 'calmness' study was performed on pooled colonies unlike previous studies which used individuals, allowing the variation of the entire colony to be captured. Avalos *et al.* (2020) compared the use of pooled colony versus individual sequencing in a GWAS of honey bees and aggression. That study found no significant correlation between individual aggression and individual alleles, but strong correlations between colony level aggression and colony level allele frequencies. This could be explained by indirect genetic effects where an individual's behaviour is influenced by the rest of the colony, and the colony is a product of colony level allele frequencies (Avalos *et al.,* 2020; Sokolowski, 2020). Specifically, social interactions are when two or more individuals send and receive multisensory cues to one another (Moore, Brodie and Wolf, 1997). This is a complex behaviour where by the sender of the social cues changes the behaviour of the recipient of the information. In this way, the genetics of the sender of the social cue effects the behaviour of the other (Wolf, *et al.,* 1998; Schneider, *et al.,* 2017). Therefore, to perform a GWAS on colony level phenotypes it is vital to assess the colony level genotype as the colony is subject to group genetic effects (Sokolowski, 2020).

1.12 Aims

In light of the preceding discussion this project aims to:

(1) Assess the effectiveness of an *A. m. mellifera* breeding program in the south west of England.

(2) Compare three different approaches for methods for measuring introgression. Firstly, using an AIMs SNP array designed for assessing an individual honey bee. Secondly, using a genome wide SNP discovery method to assess an individual honey bee. Lastly, using the genome wide SNP discovery method to assess a pooled colony.

(3) Investigate the current extent of *A. m. mellifera* across the UK and Ireland outside of specific conservation and breeding programs using whole genome data.

(4) In collaboration with researchers in Scandinavia running a phenotypic monitoring experiment, investigate subspecies purity and the underlying genetics differences between the subspecies using whole genome data.

2. STRUCTURE analyses and genetic assignment tests reveal temporal decreases in introgression in an *Apis mellifera mellifera* (L.) conservation programme

2.1 Introduction

Conservation and breeding programmes require the management of wild or captive populations. This management is often reliant on information about diversity, origin, and kinship of individuals within the population in question (Fienieg and Galbusera, 2013). Conservation programmes also aim to retain genetic variation in the population, avoiding inbreeding and inbreeding depression in subsequent generations and outbreeding effects (Fraser, 2008), as well as hybridisation and introgression events (Templeton, *et al.,* 1986; Rhymer and Simberloff, 1996).

Hybridisation is defined here as the interbreeding between members of genetically distinct lineages (Futuyma and Kirkpatrick, 2017). In parallel, introgression is defined as the transfer of alleles between genetically distinct lineages via hybridisation and subsequent backcrossing of the offspring into one, or both, of the parental lineages (Anderson, 1949). Hybridisation and introgression can be problematic for conservation efforts as they can cause the erosion of genetic integrity and the disruption of adaptive traits, including genomic extinction (Allendorf, *et al.,* 2004).

Genomic extinction is where hybridised and/or introgressed individuals replace one or both parental lineages and the intact parental genomes no longer exist. Although individual parental alleles are present in hybrid and introgressed individuals, genomic extinction is more than a loss of a unique allele at a single locus. Specifically, it is the loss of combinations of alleles and genotypes that have evolved over time and are lost due to genetic swamping via hybridisation/introgression with another lineage, resulting in the loss of a legacy of an evolutionary lineage (Allendorf and Luikart 2009).

The human-mediated movement of species can cause or increase hybridisation and introgression, potentially threatening endangered species (Ottenburghs, 2021). This process has been observed across many different taxa, for example in the red-legged partridge (*Alectoria rufa*) (Casas, *et al.,* 2012), whitefish (*Coregonus spp.)* (Dierking, *et al.,* 2014), amphibians (Talarico, *et al.,* 2020), wildcats (*Felis silvastris*) and dingos (*Canis lupus dingo*) (Daniels and Corbett, 2003). Conversely, however, hybridisation and introgression can also potentially increase genetic diversity, bringing increased adaptive potential, fitness and the possibility for adaptive introgression (Tigano and Friesen, 2016; Hendrick, 2013) as well as being significant mechanisms of evolutionary divergence (e.g. speciation events) (Rieseberg 2006; Saitoh, Chen, Mayden, 2010; Patton *et al* 2020) (see chapter 1, general introduction, for a fuller discussion on these subjects).

The use of genetic markers has allowed conservation and breeding programs to monitor populations in order to avoid inbreeding, hybridisation and introgression, and allow for estimates of population differentiation and individual relatedness (Frankham., 2010) with the ultimate aim being to maintain genetic integrity and diversity within lineages (Schwartz, Luikart and Waples., 2007). Microsatellites are single sequence repeats (SSRs) or short tandem repeats located in non-coding

regions of the genome and they have been commonly employed in this context (Chase, Kesseli and Bawa., 1996; Jehle and Arntzen., 2002; Vashistha, *et al.,* 2020). They are highly variable and codominant and thus ideally suited for many conservation uses with a focus on understanding demographic changes (Tautz, 1980; Miesfeld, *et al.,* 1981; Field and Wills, 1996). Microsatellite markers are also an excellent choice for monitoring conservation programs since they are often readily available and accessible to the conservation group or breeder via commercial companies (for example, in this study through the company Apigenix). They are also relatively cost effective and straightforward to analyse and interpret in contrast with more extensive genomic datasets, which require specialist computational resources, large data storage space and specialist skills in bioinformatics to analyse and interpret the data.

Insect pollinators provide a special set of challenges for conservation and breeding programs. A growing body of evidence indicates that many insects are in decline (e.g. Hallmann, et al., 2017; Goulson, 2019). For pollinator groups such as bumblebees, solitary bees, and other insect pollinators, conservation efforts typically concentrate on *in situ* approaches such as the preservation, or creation, of habitats, enhancing habitat connectivity, and provision of forage resources (e.g. Redhead, *et al.,* 2016; Pywell, *et al.,* 2011; Carvell, *et al.,* 2012). While the majority of insect pollinators are naturally distributed and these goals serve to increase range and connectivity, some insect pollinators are regarded as semi-domesticated, e.g. the Western honey bee (*Apis mellifera*L.). Concerns over the loss of managed honey bees have been ongoing over the last twenty years (vanEngeldrop *et al.,* 2009; Williams *et al.,* 2010). Honey bees are under threat from, amongst other factors, parasites (Varroa destructor, *Nosema* spp., *Melissococcus plutonius*), viruses (*Dicistroviruses*, for example, bee paralysis viruses) (Genersch, 2010; VanEngelsdrop and Meixner, 2010), land use

changes (Biesmeijer, *et al.,* 2006; Flynn, *et al.,* 2009), pesticides (Thompson, et al., 2003) and beekeeping practices (De la Rúa, *et al.,* 2009). Since *Apis mellifera* reside in human-managed colonies at apiary sites, its conservation and management poses especial challenges. Especially relevant in the context of introgression is the human-mediated trade of preferred subspecies (due to perceived more desirable traits or increased productivity) within and between countries. This has given rise to concerns regarding the loss of locally- adapted ecotypes and subspecies via gene flow due to the repeated importation of honey bee subspecies (De La Rúa, *et al.,* 2013).

In the UK there is a well-established culture of importation of honey bees of different subspecies, particularly popular C-lineage honey bees *A. m. carnica* and *A. m. ligustica* (in comparison to the UK subspecies *A. m. mellifera* belonging to the M lineage. See chapter 1 for full details). Due to the potential negative impacts of the importation of these commercially attractive subspecies, beekeepers are becoming increasingly interested in the conservation of local diversity and subspecies. Several groups have been set up with this goal such as the Societas Internationalis pro Conservatione *Apis mellifera mellifera* (SICAMM), Save Local Bees, The Native Irish Honey bee Society (NIHBS), Fédération Européenne des Conservatoires de l'abeille noire (FEDCAN), and in the South West England a Community Interest Company, Bringing Back Black Bees (B4). There is thus a growing demand for a cost-effective service and a feasible timeline for beekeepers to monitor their managed colonies when setting up or maintaining a breeding programme.

As for many other taxa, genetic approaches have been particularly helpful in the investigation of introgression in honey bee subspecies. Mitochondrial haplotypes (Rortais, *et al.,* 2011), microsatellites (Solignac, *et al.,* 2003; Oleksa and Tofilski, 2015; Ellis, *et al.,* 2018) and single nucleotide polymorphisms (SNPs) (Chávez-Galarza *et al.,* 2013; Harpur, *et al.,* 2014; Pinto, *et al.,* 2014) have all been used to

gain insight into the “purity” of colonies and populations. These methods all vary in initial investment costs, running costs, precision, and resolution (Mace and Purvis, 2008). Assignment analysis, which makes use of genetic informationto infer groups and assign individuals to those groups is often applied to these genetic data to evaluate introgression (Peter, 2016; Lawson, *et al.,* 2018).

The B4 project was founded by a group of beekeepers whose aim is to conserve, protect and increase the population of the honey bee subspecies *Apis mellifera mellifera* in the UK. The project aims to set up reserves with like-minded beekeepers and identify the genetic purity of honey bees within the project and establish a breeding programme. To achieve this, the project uses isolated mating apiaries and drone flooding areas to prevent mating with any imported subspecies present locally in combination with genetic testing of the honey bees in the programme.

Previous work at the University of Plymouth used microsatellite genotyping and mtDNA data to assess samples from the B4 project aiming to identify any pure *Apis mellifera mellifera* colonies and assess the state of introgression of bees in the South-West (Ellis, *et al.,* 2018). The study established that the colonies were generally admixed and showed introgression from the subspecies *Apis mellifera carnica*.

The aim of this present study is to assess temporal changes in genetic diversity and introgression of these same honey bee apiaries, asking whether there have been changes in the genetic background of honey bees in the programme and assessing the effectiveness of the breeding efforts over time.

2.2 Methods

2.2.1 Sample collection

Honey bee samples were collected by members of the B4 network from managed apiaries in Cornwall in 2015 (n=11) and again in 2018 (n=15). The colonies sampled were putative members of the subspecies *Apis mellifera mellifera* being monitored as part of B4's breeding and conservation effort. A total of 30 drone brood were sampled per colony by removal of the right antenna and preserving that in absolute ethanol. Genotyping haploid males, instead of worker larvae, allows assessment of the genotype of the queen(s) without harming the colony and avoids patrilines entering the analysis from the queen(s) mates when workers are accidentally processed. Samples of European subspecies obtained by Soland-Reckeweg *et al.,* (2009) were included as standards to enable the assessment of admixture and introgression of the Cornish samples. The standards were chosen from a wide geographic range and from locations where low levels of introgression were known: *A. m. carnica* from Slovenia (n=21) where this subspecies has country-wide legal protection, and three populations from Austria (Austria n= 44, Austria Wurm n=36 and Austria 2015 n=102) that also have legal protection; *A. m. mellifera* from a legally protected conservation area in Norway (n=18), a secluded French (n=24) population with no known introductions of foreign subspecies, Sweden (n=10), two populations from Switzerland (Glarus n=10, Schistal n=12) and Ireland (n=22); *A. m. ligustica* from Italy (n=55).

2.2.2 DNA extractions and microsatellite PCR

DNA was extracted using a DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. All genotyping was performed by Apigenix in Switzerland. Amplification of 8 microsatellite loci (A007, A28, Ac306, Ap226, Ap273, Ap289, Ap33 and B24) (Solignac, *et al.,* 2003) was performed using two multiplex PCR reactions (Soland-Reckeweg, *et al*., 2009). Reactions were carried out in a total volume of 10µl containing 2- 10 ng of genomic DNA, 5µl Hot-StarTaq Master Mix, double distilled water, and a final concentration of 10µM of each forward and reverse primer.

2.3 Statistical analysis of microsatellite data

2.3.1 Quality control and diversity measures

Microsatellite data were grouped by location and checked for any genotyping errors (e.g. stutter effects, null alleles) using Microchecker version 2.2.3 (Van Oosterhout, *et al.,* 2004). Data were then explored for departures from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) in Arlequin version 3.5 (Excoffier, *et al.,* 2005). In Arelquin, analyses were run for 1,000,000 Markov chain steps and 100,000 dememorization steps to estimate HWE and LD (Ellis, *et al.,* 2018; Pfeiler, *et al.,* 2017; Seitz, *et al.,* 2021). Arlequin was then used to measure observed and expected heterozygosity. Allelic richness was calculated across loci using FSTAT (Goudet, 2001), based on a minimum of 10 samples per population.

2.3.2 Relatedness between individuals

The presence of related individuals in the Cornwall 2015 and the Cornwall 2018 samples was tested using two programs. First, the ML-Relate program (Kalinowski, 2008) was used, which employs a maximum likelihood approach to estimate relatedness. ML-Relate is designed for microsatellite data and can accommodate null alleles (Kalinowski, 2008). Second, COANCESTRY (Wang, 2011), which estimates relatedness (r_{xy}) using seven relatedness estimators and allows for the selection of the most appropriate estimator, was also used. COANCESTRY uses two maximum likelihood estimators and five moment estimators along with bootstrapping to obtain P-values. The five moment estimators differ in the interpretation and use of multilocus genotype information (Queller and Goodnight, 1989; Li, *et al.,* 1993; Ritland., 1996; Lynch and Ritland., 1999; Wang., 2002). Unlike the five moment estimators the maximum likelihood estimators in both COANCESRTY and ML-relate allow for inbred individuals in the dataset. Both packages were used as COANCESTRY is recommended as a complementary approach to other programmes (Wang, 2011). In order to select the most appropriate estimator in COANCESTRY, a total of 400 dyads consisting of 100 dyads per relationship were simulated. Relationships simulated were parent-offspring or full-sibling (rxy =0.5), half-sibling or grandparent-grandchild (rxy =0.25) and unrelated (rxy =0). The program does not distinguish between relationships that would produce the same relatedness value, for example, parent-offspring and full-sibling, or between half-sibling and grandparent-grandchild. Pearson's correlations were performed between the expected relatedness values of the simulated relationships, (for example the expect relationship value for a simulated parent-offspring is 0.5) and the relatedness values resulting from the estimators. The estimator that had a Pearson's correlation closest to one was chosen as the best.

2.3.3 Investigation of population structure and introgression

To assess subspecies assignment two approaches were used: inference of population clusters using STRUCTURE (Pritchard, *et al.,* 2000) and genetic assignment analysis using ONCOR (Kalinowski, *et al.,* 2007). STRUCTURE assigns samples to the most likely clusters using a fully Bayesian method and utilizes all of the samples available (Pritchard, *et al.,* 2000). ONCOR is a genetic assignment tool and uses a partially Bayesian method to assign samples to a reference baseline (Kalinowski, *et al.,* 2007). While fully Bayesian methods have been suggested to be more rigorous than partially Bayesian methods, they rely on the assumption that the population of origin was sampled (Manel, *et al.,* 2002). Manel *et al.,* (2002) recommend employing a partially Bayesian method (such as ONCOR) alongside a fully Bayesian method as implemented in STRUCTURE where this assumption may not be fully met.

Markov chain Monte Carlo (MCMC) estimations were implemented in STRUCTURE to estimate the number of genetic clusters (K) and individual membership to each cluster (Q) (Pritchard, *et al.,* 2000). Three iterations of K from 2-12 were run with the admixture model, without geographical information (no Locprior). The burn-in period was 50,000 and 500,000 MCMC replications were used in lines with recommended protocols (Janes, *et al.,* 2017). Each iteration of K assigned a membership value (Q) for each individual to each cluster, with the mean Q value from all individuals forming the overall population level membership Q value to each cluster. To investigate the best K value the log probability of the data was used to infer the most probable K value via the deltaK value calculated using Structure Harvester (Earl, *et al.,* 2012). The degree of introgression was examined at the population and individual levels

through mean Q values and standard deviations for K=3 as this proved to be the most informative for investigating the structure of populations (see Results).

STRUCTURE assumes that loci are independent and related individuals thus have the potential to cause the algorithm to perform sub-optimally (Rodríguez-Ramilo, *et al.,* 2014). Therefore, the focal samples from each year (Cornwall 2015 and Cornwall 2018) were examined in separate STRUCTURE analyses.

In the second approach, control samples for each subspecies (listed above) were used to generate baseline populations in ONCOR and both Cornish samples sets were assigned. Assignments are based on the probability that the given multilocus genotype would be produced by one of the baseline populations. Subsequently the ONCOR option 'Leave One Out' test was performed on the baseline population to evaluate the reproducibility of the baseline population assignments by iteratively removing and reassigning one individual per baseline population. Then the samples from Cornwall 2015 and Cornwall 2018 were placed through ONCOR's Mixture Analysis to estimate the proportion of the whole population that belongs to each baseline population and an assignment test was employed to estimate the origin of each individual sample (Kalinowski, *et al.,* 2007).

2.4 Results

2.4.1 Genetic Diversity

No departure from HWE or any linkage between loci was detected. The Cornish samples showed similar values of genetic diversity (allelic richness and observed heterozygosity) over the two sample periods (Table 2.1). Both Cornish samples showed the highest genetic diversity values across all populations analysed.

Table 2.1 Genetic diversity of *A. mellifera* populations (SD = standard deviation).

Population	Subspecies	Average allelic richness		Observed Heterozygosity			Expected Heterozygosity		
		Mean	SD	Mean	SD	SE	Mean	SD	SE
Cornwall 2015	Unknown	5.16	2.80	0.59	0.26	0.09	0.62	0.22	0.08
Cornwall 2018	Unknown	5.06	2.87	0.61	0.23	0.08	0.62	0.23	0.08
Italy	*A. m. ligustica*	3.53	2.72	0.35	0.35	0.12	0.38	0.37	0.13
Austria	*A. m. carnica*	4.41	1.97	0.54	0.24	0.08	0.57	0.20	0.07
Austria Wurm	*A. m. carnica*	3.93	1.02	0.57	0.15	0.05	0.59	0.12	0.04
Slovenia	*A. m. carnica*	3.63	1.78	0.45	0.26	0.09	0.44	0.27	0.09
Sweden	*A. m. mellifera*	3.00	1.41	0.38	0.27	0.10	0.46	0.33	0.12
France	*A. m. mellifera*	3.92	3.67	0.38	0.30	0.11	0.39	0.34	0.12
Norway	*A. m. mellifera*	3.59	2.79	0.28	0.31	0.11	0.40	0.35	0.12
Switzerland Glarus	*A. m. mellifera*	3.88	3.00	0.41	0.34	0.12	0.43	0.33	0.12
Switzerland Schistal	*A. m. mellifera*	3.62	2.42	0.43	0.33	0.12	0.40	0.32	0.11
Ireland	*A. m. mellifera*	3.67	2.62	0.36	0.26	0.09	0.39	0.32	0.11
Austria 2015	*A. m. carnica*	4.28	1.90	0.50	0.21	0.08	0.52	0.24	0.08

2.4.2 Relatedness of individuals

The COANCESTRY relatedness estimator that was identified as the best was the dyadic likelihood estimator (DyadML) (Table 2.2). The remaining six estimators were removed from subsequent analysis. Coincidentally, the DyadML estimator is the same estimator used by the ML-relate program and ML-relate and COANCESTRY produced identical results in all but one pairwise relationship (Table 2.3; COANCESTRY detected one relationship that ML-relate did not detect (Illogan 2015 and Trethill A). There were no parent, offspring or full sibling relationships found, by either programme, comparing the Cornwall 2015 and Cornwall 2018 samples; the highest relatedness value between the 2015 and 2018 samples was rxy = 0.32. ML-relate categorised nine half-sibling relationships between Cornwall 2015 and Cornwall 2018 (Table 2.3). In total, 32 pairwise relationships between the Cornwall 2015 and Cornwall 2018 samples were given estimated relatedness values above zero by both ML-Relate and COANCESTRY (Table 2.3). Relatedness was detected between colonies from the same apiary sampled in different years (Antony, Insworke, Trethill, Illogan and Insworke. Table 2.3) as well as between apiaries (Scooner 2015 and Rosewarne A, Slaters 2015, and Antony B, Illogan 2015 and Trethill B, St. Agnes 2015 and Maker Cornwall and St Agnes 2015 and Lizard. Table 2.3).

Table 2.2 Performance of COANCESTRY relatedness estimators from simulation module

Estimator	Pearsons r	P-value
TrioML	0.74	1.15E-71
Wang	0.73	7.90E-67
LynchLi	0.70	9.69E-60
LynchRd	0.73	5.01E-67
Ritland	0.69	7.46E-58
QuellerGt	0.69	7.46E-58
DyadML	0.76	2.57E-77

Table 2.3 Estimates of pairwise relatedness and most likely relationships between Cornwall 2015 and Cornwall 2018 produced by ML-relate and COANCESTRY. LR is the most likely relations produced by ML-Relate, U= Unrelated and HS= Half sibling/Grandparent/Grandchild. R_{xy} are pairwise relatedness produced by ML-Relate and COANCESTRY. *This value was produced only by COANCESTRY

Cornwall 2018	Cornwall 2015																			
	Antony		Insworke		Maker Cornwall		Sconner		Tregantle		Trethill		Rosewarne		Slaters		Illogan		St. Agnes		Lizard	
	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}	LR	R_{xy}
Insworke A	U	0.07	**HS**	0.25	U	0	U	0	U	0	U	0	U	0	U	0.07	U	0	U	0	U	0
Insworke B	U	0	U	0.1	U	0	U	0	U	0.07	U	0.21	U	0.04	U	0	U	0.02*	U	0	U	0
Maker Cornwall	U	0	U	0	U	0.1	U	0	U	0.13	U	0	U	0.12	U	0	U	0	**HS**	0.1	U	0
Trethill A	U	0.12	U	0	U	0	U	0.18	U	0	**HS**	0.26	U	0	U	0	U	0.3	U	0	U	0
Trethill B	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	**HS**	0.3	U	0	U	0
Slaters Falmouth	U	0	U	0.07	U	0	U	0	U	0	U	0	U	0	U	0	U	0.2	U	0	U	0
Tregantle	U	0.06	U	0	U	0.07	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0
Sconner	U	0	U	0	U	0	U	0	U	0.1	U	0.09	U	0	U	0	U	0	U	0	U	0
Antony A	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0
Antony B	**HS**	0.18	U	0	U	0	U	0.01	U	0	U	0	U	0	**HS**	0.25	U	0	U	0	U	0
Lizard	U	0	U	0	U	0.06	U	0	U	0	U	0	U	0	U	0	U	0	**HS**	0.2	U	0
Illogan	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	**HS**	0.3	U	0	U	0
Rosewarne A	U	0	U	0	U	0	**HS**	0.23	U	0	U	0	U	0	U	0	U	0	U	0	U	0
Rosewarn B	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0
St Agnes	U	0	U	0	U	0	U	0	U	0	U	0	U	0	U	0.06	U	0	U	0	U	0

2.4.3 Population structure

2.4.3.1. STRUCTURE analysis

The STRUCTURE analysis identified K=3 as the most suitable K value. This is unsurprising as it differentiates the subspecies *A. m. mellifera, A. m. ligustica* and *A. m. carnica*. At K=3 the Cornwall 2018 showed the largest mean membership coefficient (Q value 0.82) with the European *A. m. mellifera* samples (Sweden, France, Norway, both Switzerland samples and Ireland, (Figure 2.1 and Table 2.4). Cornwall 2018 additionally showed membership to the *A. m. carnica* cluster (mean Q value 0.15). The same pattern was found in the Cornwall 2015 samples, which showed the largest mean membership coefficient for European *A. m. mellifera* (0.71) and an additional membership to *A. m. carnica* (0.27 mean Q values) at K=3 (Figure 2.2 and Table 2.5). A comparison of mean Q values of apiaries that were sampled in 2015 and 2018 indicated that four apiaries had a reduction in the mean membership to the *A. m. mellifera* cluster (Insworke B, Maker Cornwall, Slaters and Antony A, Table 2.6) while eleven apiaries saw an increase in the mean membership to the *A. m. mellifera* cluster (Table 2.6). The largest increases between 2015 and 2018 in mean Q value membership to the *A. m. mellifera* were observed in the Trethill and Rosewarne samples (Table 2.6). The largest decrease in *A. m. mellifera* mean Q value membership was seen in the Insworke site (Table 2.6). The overall change in subspecies assignment across the two time periods can be seen by the average change in mean Q values between 2015 and 2018. This change is away from *A. m. carnica* cluster (-0.16) towards the *A. m. mellifera* cluster (+0.16)(Table 2.6).

2.4.3.2. ONCOR analysis

Evaluations of the baseline populations via the Leave One Out test in ONCOR reassigned 100% of the *A. m. ligustica* back to the *A. m. ligustica* baseline population, 98.5% of the *A. m. carnica* individuals back to the *A. m. carnica* baseline population and 100% of the *A. m. mellifera* individuals back to the *A. m. mellifera* baseline population (Table 2.7). The mixture analysis estimated Cornwall 2015 to belong to the *A. m. mellifera* baseline (0.64) and the to the *A. m. carnica* baseline (0.36) (Table 2.8). The mixture analysis estimated the Cornwall 2018 samples to belong to the *A. m. mellifera* baseline (0.86) and *A. m. carnica* (0.14) (Table 2.8). Individual assignment probabilities showed nine apiary sites had an increase in assignment probabilities to the *A. m. mellifera* baseline from the Cornwall 2015 to the Cornwall 2018 samples (Table 2.9). The ONCOR individual assignments also showed two apiary sites had a reduction in assignment to the *A. m. mellifera* baseline. The largest increase in assignment to the *A. m. mellifera* baseline was seen in the Trethill and Rosewarne apiaries (Trethill 0.999 and 1.0, Rosewarne 1.0 and 1.0). The largest decrease in *A. m. mellifera* baseline assignment was seen in the Insworke B apiary sample, where the sample was assigned to *A. m. carnica* (1.0). Overall, four apiaries indicated no change in assignment to the *A. m. mellifera* baseline and seven saw an increase in baseline assignment to *A. m. mellifera*.

Both STRUCTURE and ONCOR detected the largest increase in *A. m. mellifera* assignment in the Trethill and Rosewarne sites and the largest decreases in the Insworke site (Table 2.10). In four sites STRUCTURE presented a different pattern to ONCOR. In two instances STRUCTURE detected that a site had had a decrease in mean Q value when ONCOR detected no change (Maker Cornwall and Slaters Falmouth). Twice STRUCTURE indicated a site increasing in mean Q value while

ONCOR, again, detected no change in baseline assignment (Lizard and St Agnes) (Table 2.10).

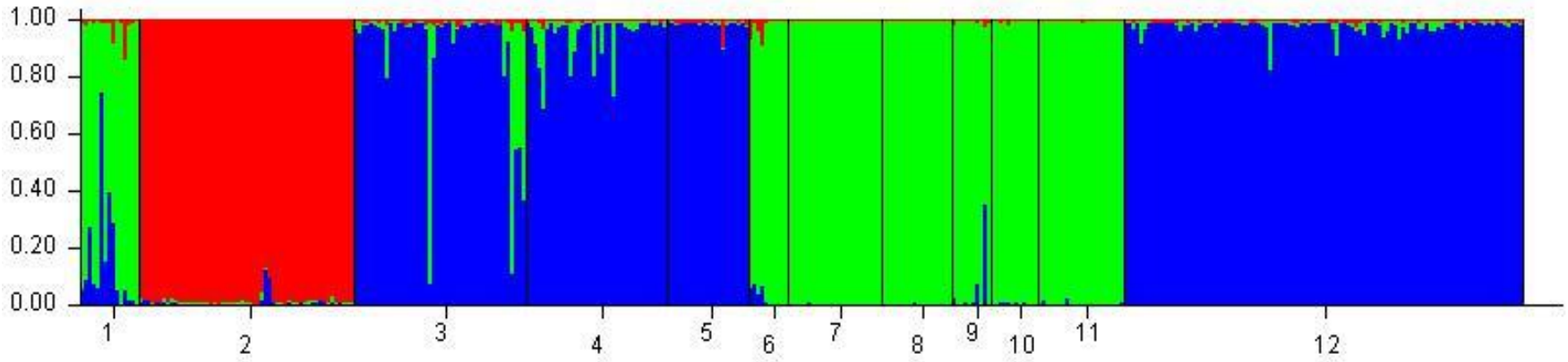


Figure 2.1 STRUCTURE analysis of Cornwall 2018 samples. K=3. Population key- 1, Cornwall 2018. 2, Italy, *A. m. ligustica*. 3, Austria, *A. m. carnica*. 4, Austria Wurm, *A. m. carnica*. 5, Slovenia, *A. m. carnica*. 6, Sweden, *A. m. mellifera*. 7, France, *A. m. mellifera*. 8, Norway, *A. m. mellifera*. 9, Switzerland Galrus, *A. m. mellifera*. 10, Switzerland Schistal, *A. m. mellifera*. 11, Ireland, *A. m. mellifera*. 12, Austria 2015, *A. m. carnica*

Table 2.4. Mean Q values of membership of populations to clusters identified in STRUCTURE at K=3 for the Cornwall 2018 analysis

Population	Clusters					
	1		2		3	
	Average Q	SD	Average Q	SD	Average Q	SD
Cornwall 2018	**0.82**	0.000	0.02	0.000	**0.15**	0.000
Italy	0.01	0.000	**0.98**	0.000	0.01	0.000
Austria	0.10	0.001	0.01	0.000	**0.89**	0.000
Austria Wurm	0.05	0.000	0.01	0.000	**0.94**	0.000
Slovenia	0.00	0.001	0.01	0.001	**0.98**	0.000
Sweden	**0.95**	0.000	0.02	0.000	0.03	0.000
France	**0.99**	0.000	0.00	0.000	0.00	0.000
Norway	**0.99**	0.000	0.00	0.000	0.00	0.000
Switzerland Glarus	**0.94**	0.000	0.01	0.000	0.05	0.001
Switzerland Schistal	**0.99**	0.000	0.01	0.000	0.01	0.000
Ireland	**0.99**	0.000	0.00	0.000	0.01	0.000
Austria 2015	0.01	0.000	0.01	0.000	**0.98**	0.000

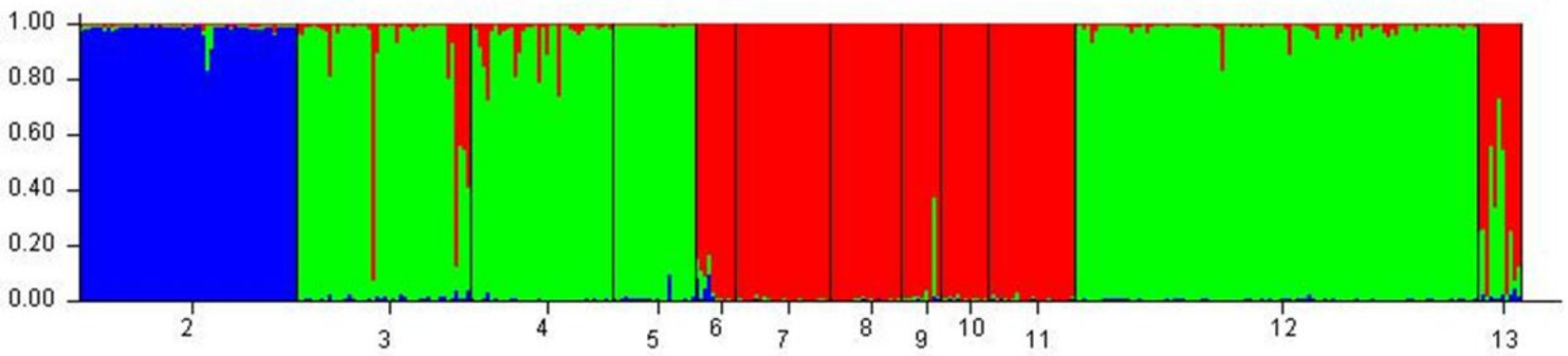


Figure 2.2 STRUCTURE analysis of Cornwall 2015 samples. K=3. Population key- 2, Italy, *A. m. ligustica*. 3, Austria, *A. m. carnica*. 4, Austria Wurn, *A. m. carnica*. 5, Slovenia, *A. m. carnica*. 6, Sweden, *A. m. mellifera*. 7, France, *A. m. mellifera*. 8, Norway, *A. m. mellifera*. 9, Switzerland Galrus, *A. m. mellifera*. 10, Switzerland Schistal, *A. m. mellifera*. 11, Ireland, *A. m. mellifera*. 12, Austria 2015, *A. m. carnica*. 13, Cornwall 2015

Table 2.5. Mean Q value of membership of populations to clusters identified in STRUCTURE at K=3 for Cornwall 2015 analysis

Population	Clusters					
	1		2		3	
	Average Q	SD	Average Q	SD	Average Q	SD
Italy	0.01	0.000	0.01	0.000	**0.98**	0.000
Austria	0.10	0.001	**0.89**	0.000	0.01	0.000
Austria Wurm	0.05	0.000	**0.94**	0.000	0.01	0.000
Slovenia	0.01	0.000	**0.98**	0.000	0.01	0.000
Sweden	**0.94**	0.000	0.03	0.001	0.03	0.001
France	**0.99**	0.000	0.00	0.001	0.00	0.000
Norway	**0.99**	0.000	0.00	0.000	0.00	0.000
Switzerland Glarus	**0.95**	0.000	0.05	0.000	0.01	0.000
Switzerland Schistal	**0.99**	0.000	0.01	0.001	0.01	0.000
Ireland	**0.99**	0.000	0.01	0.000	0.00	0.000
Austria 2015	0.01	0.000	**0.98**	0.000	0.01	0.001
Cornwall 2015	**0.71**	0.002	**0.28**	0.002	0.02	0.000

Table 2.6. STRUCTURE membership Q values to cluster produced by K=3

Sample	Cluster memberships						Change in mean Q from 2015 to 2018		
	A. m. mellifera		*A. m. carnica*		*A. m. ligustica*				
	mean Q	SD	mean Q	SD	mean Q	SD	*A. m. mellifera*	*A. m. carnica*	*A. m. ligustica*
Insworke 2015	0.74	0.003	0.24	0.004	0.02	0.001			
Insworke 2018 A	0.94	0.001	0.05	0.000	0.01	0.000	0.20	-0.19	-0.01
Insworke 2018 B	0.24	0.002	0.75	0.002	0.01	0.000	-0.50	0.51	-0.01
Maker Cornwall 2015	0.98	0.000	0.01	0.000	0.01	0.000			
Maker Cornwall 2018	0.89	0.001	0.09	0.002	0.02	0.001	-0.09	0.08	0.01
Trethill 2015	0.27	0.002	0.73	0.001	0.01	0.000			
Trethill 2018 A	0.72	0.001	0.27	0.001	0.01	0.000	0.46	-0.46	0.00
Trethill 2018 B	0.84	0.001	0.16	0.001	0.01	0.000	0.57	-0.57	0.00
Slaters Falmouth 2015	0.98	0.000	0.02	0.000	0.01	0.000			
Slaters, Falmouth 2018	0.92	0.001	0.07	0.001	0.01	0.000	-0.06	0.05	0.01
Tregantle 2015	0.66	0.003	0.33	0.003	0.01	0.001			
Tregantle 2018	0.93	0.001	0.06	0.000	0.01	0.000	0.27	-0.27	0.00
Sconner 2015	0.44	0.002	0.54	0.003	0.02	0.002			
Sconner 2018	0.59	0.000	0.40	0.000	0.01	0.000	0.15	-0.14	-0.01
Antony 2015	0.74	0.003	0.25	0.003	0.00	0.001			
Antony 2018 A	0.63	0.000	0.29	0.000	0.09	0.000	-0.12	0.04	0.08
Antony 2018 B	0.94	0.000	0.05	0.001	0.01	0.000	0.20	-0.20	0.00
Lizard 2015	0.87	0.002	0.11	0.002	0.02	0.001			
Lizard 2018	0.98	0.001	0.01	0.000	0.01	0.001	0.11	-0.10	-0.01
Illogan 2015	0.74	0.002	0.23	0.003	0.02	0.001			
Illogan 2018	0.81	0.000	0.06	0.001	0.14	0.001	0.06	-0.18	0.12
Rosewarne 2015	0.46	0.002	0.52	0.002	0.03	0.001			
Rosewarne 2018 A	0.96	0.001	0.02	0.001	0.02	0.000	0.51	-0.50	-0.01
Rosewarne 2018 B	0.97	0.000	0.02	0.001	0.01	0.000	0.51	-0.50	-0.02
St Agnes 2015	0.92	0.002	0.03	0.001	0.05	0.001			
St Agnes 2018	0.99	0.000	0.01	0.000	0.00	0.000	0.07	-0.03	-0.04
	Average of mean Q values						Average change in mean Q values from 2015 to 2018		
2018	**0.82**		**0.15**		**0.02**		**0.16**	**-0.16**	**0.01**
2015	**0.71**		**0.27**		**0.02**				

Table 2.7 Leave One Out baseline test results of subspecies controls from ONCOR

ONCOR Leave One Out test		
Baseline group	Number of samples	% correctly reassigned
A. m. ligustica	50	100.00%
A. m. carnica	201	98.50%
A. m. mellifera	93	100.00%

Table 2.8 ONCOR mixture analysis of Cornwall 2015 and Cornwall 2018 populations

ONCOR population mixture analysis		
Baseline population	Mixture estimates 2015	Mixture estimates 2018
A. m. ligustica	0.00	0.00
A. m. carnica	0.36	0.14
A. m. mellifera	0.64	0.86

Table 2.9 ONCOR probabilities of each sample belonging to each baseline population

Sample	Baseline assignment			Change in assignment from 2015 to 2018		
	A. m. mellifera	*A. m. carnica*	*A. m. ligustica*	*A. m. mellifera*	*A. m. carnica*	*A. m. ligustica*
Insworke 2015	0.980	0.020	0			
Insworke 2018 A	1.000	0.000	0	0.020	-0.020	0.000
Insworke 2018 B	0.000	1.000	0	-0.980	0.980	0.000
Maker Cornwall 2015	1.000	0.000	0			
Maker Cornwall 2018	1.000	0.000	0	0.000	0.000	0.000
Trethill 2015	0.000	1.000	0			
Trethill 2018 A	0.999	0.001	0	0.999	-0.999	0.000
Trethill 2018 B	1.000	0.000	0	1.000	-1.000	0.000
Slaters Falmouth 2015	1.000	0.000	0			
Slaters, Falmouth 2018	1.000	0.000	0	0.000	0.000	0.000
Tregantle 2015	0.011	0.989	0			
Tregantle 2018	1.000	0.000	0	0.989	-0.989	0.000
Sconner 2015	0.021	0.979	0			
Sconner 2018	0.140	0.860	0	0.120	-0.120	0.000
Antony 2015	0.997	0.003	0			
Antony 2018 A	0.798	0.202	0	-0.199	0.199	0.000
Antony 2018 B	1.000	0.000	0	0.003	-0.003	0.000
Lizard 2015	1.000	0.000	0			
Lizard 2018	1.000	0.000	0	0.000	0.000	0.000
Illogan 2015	0.998	0.002	0			
Illogan 2018	1.000	0.000	0	0.002	-0.002	0.000
Rosewarne 2015	0.001	0.999	0			
Rosewarne 2018 A	1.000	0.000	0	0.999	-0.999	0.000
Rosewarne 2018 B	1.000	0.000	0	0.999	-0.999	0.000
St Agnes 2015	1.000	0.000	0			
St Agnes 2018	1.000	0.000	0	0.000	0.000	0.000
	Average baseline assignment			Average change from 2015 to 2018		
2018	**0.81**	**0.14**	**0.00**	**0.26**	**-0.26**	**0**
2015	**0.64**	**0.36**	**0.00**			

Table 2.10. Comparison of STRUCTURE mean Q values and ONCOR baseline assignment results for Cornwall 2015 and Cornwall 2018

Sample	*A. m. mellifera*		*A. m. carnica*		*A. m. ligustica*	
	ONCOR Baseline assignment	STRUCTURE mean Q value	ONCOR Baseline assignment	STRUCTURE mean Q value	ONCOR Baseline assignment	STRUCTURE mean Q value
Insworke 2015	0.980	0.74	0.020	0.24	0	0.02
Insworke 2018 A	1.000	0.94	0.000	0.05	0	0.01
Insworke 2018 B	0.000	0.24	1.000	0.75	0	0.01
Maker Cornwall 2015	1.000	0.98	0.000	0.01	0	0.01
Maker Cornwall 2018	1.000	0.89	0.000	0.09	0	0.02
Trethill 2015	0.000	0.27	1.000	0.73	0	0.01
Trethill 2018 A	0.999	0.72	0.001	0.27	0	0.01
Trethill 2018 B	1.000	0.84	0.000	0.16	0	0.01
Slaters Falmouth 2015	1.000	0.98	0.000	0.02	0	0.01
Slaters, Falmouth 2018	1.000	0.92	0.000	0.07	0	0.01
Tregantle 2015	0.011	0.66	0.989	0.33	0	0.01
Tregantle 2018	1.000	0.93	0.000	0.06	0	0.01
Sconner 2015	0.021	0.44	0.979	0.54	0	0.02
Sconner 2018	0.140	0.59	0.860	0.40	0	0.01
Antony 2015	0.997	0.74	0.003	0.25	0	0.00
Antony 2018 A	0.798	0.63	0.202	0.29	0	0.09
Antony 2018 B	1.000	0.94	0.000	0.05	0	0.01
Lizard 2015	1.000	0.87	0.000	0.11	0	0.02
Lizard 2018	1.000	0.98	0.000	0.01	0	0.01
Illogan 2015	0.998	0.74	0.002	0.23	0	0.02
Illogan 2018	1.000	0.81	0.000	0.06	0	0.14
Rosewarne 2015	0.001	0.46	0.999	0.52	0	0.03
Rosewarne 2018 A	1.000	0.96	0.000	0.02	0	0.02
Rosewarne 2018 B	1.000	0.97	0.000	0.02	0	0.01
St Agnes 2015	1.000	0.92	0.000	0.03	0	0.05
St Agnes 2018	1.000	0.99	0.000	0.01	0	0.00

2.5 Discussion

This study evaluated the change in introgression, diversity and relatedness over time in an *A. m. mellifera* conservation programme. The levels of introgression reported by both analysis approaches, STRUCTURE and ONCOR, were congruent and detected introgression with *A. m. carnica*. However, there was an increase in *A. m. mellifera* assignment at the population level from 2015 and 2018. Although the results should be considered with caution due to the small sample size, this study suggests that the conservation breeding program is making progress towards increasing the *A. m. mellifera* purity across the studied apiaries.

Estimates of the magnitude of the shift towards *A. m. mellifera* varied from 11% (STRUCTURE) to 22% (ONCOR). However, at the individual level, while the pattern of results for the majority of samples were consistent between the two approaches, four samples were reported by ONCOR as having no change in *A. m. mellifera* assignment whereas STRUCTURE reported them as having an increase (Lizard +0.11 and St. Agnes +0.07) or decrease (Maker -0.09 and Slaters Falmouth -0.06). STRUCTURE infers the probability of membership to a given K value by directly using all the samples and minimizing population genetic parameters such as departure from Hardy-Weinberg equilibrium and linkage disequilibrium, therefore Q values can change depending on which samples are included. On the other hand, ONCOR allows the creation of baseline populations separately and assigns individual samples to the baseline of pre-defined clusters. While it remains difficult to place a definitive number on the overall increase in *A. m. mellifera* assignment, these results indicate an increase in *A. m. mellifera* of the honey bee colonies within the breeding programme of the B4 initiative, thus the breeding programme appears to be making progress in its aims.

Beekeepers in the B4 project aim to isolate mating apiaries, practice queen rearing, and practice drone flooding (Widdicombe, 2015). Isolated mating sites are used by this group as they operate in Cornwall, UK and utilise the peninsulas available, such as Rame Head and The Lizard, and often have the advantage of using the coastline to further isolate the apiary site. Drone flooding allows the beekeepers to fill an area with preferred drones or drones from a previously tested colony, although to be successful it may also require the cooperation of neighbouring beekeepers. Finally, the group will often rear queens from colonies previously tested to breed from that colony. However, it is important to note that the changes in introgression could be attributed to a number of confounding factors. For example, sampling effects, demographic variation within the populations or purposeful measures taken by the bee keepers could all explain the observed trends.

The genetic diversity measurements of allelic richness and observed heterozygosity suggest that the diversity of the Cornish population over the two sample periods were similar. The Cornish populations also possessed the highest diversity measurements in both allelic richness (Cornwall 2018, 5.1 and Cornwall 2015, 5.2) and observed heterozygosity (Cornwall 2018, 0.61 and Cornwall 2015, 0.59) compared to the European populations used in this study for comparison. The higher diversity could be a result of continued outbreeding with other subspecies. The Cornwall 2015 and 2018 populations did show membership to all three clusters in STRUCTURE more so than when compared with the European populations, showing the Cornish populations both in 2015 and in 2018 are more introgressed than the European populations.

The sample size of the study was small (Cornwall 2015, n=11 and Cornwall 2018, n=15) as these samples were specifically chosen by beekeepers for monitoring an ongoing conservation programme. It is important to note that this is not reflective of

the entire Cornish population which, at the time of sampling, would have consisted of >5500 colonies (Ellis, *et al.*, 2018).

One of the concerns in conservation is that of a small population size (Ouborg, 2010). Small populations are more susceptible to the loss of genetic diversity in the form of genetic drift and inbreeding, which in turn reduces the potential for adaptive responses to, for example, environmental change, demographic effects or disease (Frankham, 2005), as well as typically increasing genetic load. Care should be taken to ensure that the populations in this conservation program have adequate population sizes to prevent the loss of diversity while at the same time attempting to reduce introgression. There was a considerable amount of relatedness between sites and within the same sites in different years. This presents a challenge to beekeepers as they will want to avoid inbreeding and reducing the diversity in the *A. m. mellifera* stock. Conversely, the degree of relatedness within the honey bees of the B4 project was not associated with reductions in genetic diversity when compared to the European populations, at least at the current time. The extent of relatedness between apiaries is not surprising given that beekeepers from this conservation group often practice queen trading and transhumance of colonies between apiaries. The highest relatedness estimation value between the 2015 apiaries and the 2018 apiaries was rxy = 0.26, suggesting that some apiaries were founded by half-siblings or grandparents to those sampled in 2018. Beekeepers often assert that queens live up to two years (B4, personal communication) and studies have shown that the lifespan of a sexually reproductive queen is a mean length of one year (Page and Peng, 2001). The study periods were four years apart and the relationships calculated are potentially more likely to be that of a grandparent to grandchild than half-siblings.

For the majority of breeding and conservation programmes to be successful, the end goal is for the target species to become self-sustaining in the long term (Frankham, *et al.,* 2010). However, in this instance, this end goal has never been the objective as honey bees are semi-domesticated. This will inevitably continue as honey bees are managed for their honey production and farmed on larger scales for crop pollination. Instead, the end goal here is to increase the population of *A. m. mellifera* in the South West of England, and to prevent the loss of this subspecies along with their locally adapted traits. However, further work is needed to establish whether *A. m. mellifera* is better adapted than the long established locally bred honey bees may or may not have high levels of introgression. It has been suggested that a large proportion of honey bee colonies (those that are not protected or in a conservation programme) are now considerably artificially hybridized because of human-mediated queen transport (Pinto *et al.,* 2014). Cornwall and the B4 project are geographically located not far from a famous human-led hybridization project that was a purposeful attempt to create a bee that was resilient to disease and possessed an increased honey yield, the 'Buckfast Bee ™'. This hybridization is a result of the cultural practices of beekeeping in the UK, where importation is commonplace (see section 1.9).

For breeding programs and interest groups to succeed in protecting populations of *A. m. mellifera* and expanding their range, continued genetic monitoring is required. While microsatellite analysis is of great use, recently an iPLEX SNP panel of ancestry informative markers (AIMs) has been created for accurate and cost-effective detection of introgression from C lineage honey bees in *A. m. mellifera* bees (Muñoz, *et al.,* 2015; Henriques *et al.,* 2018). Microsatellites are highly polymorphic markers distributed throughout the genome in non-coding regions (Tautz, 1980; Miesfeld, *et al.,* 1981; Field and Wills, 1996). As microsatellites appear in these non-coding regions across the genome they represent demographic information, rather than any

information about functional differences between the honey bee subspecies. SNPs, like microsatellites, are distributed across the genome but can be representative of both demographic and functional processes (Allendorf, 2017; Morin, *et al.,* 2004). The SNPs on the iPLEX panel are largely in non-coding regions. Of the 117 SNPs on the panel, 90 are situated in non-coding regions and 27 are situated in coding regions, and only two of those coding SNPs are non-synonymous significant changes (Muñoz, *et al.,* 2015; Henriques, *et al.,* 2018). Studies that wish to uncover any functional difference (indicative of functionality) between subspecies and allow insights into local adaptation or to investigate functional information in introgressed regions, should consider using methods that capture sequences from across the genome and contain sufficient functional and demographic information (Zimmerman, *et al.,* 2020). The challenge remaining is to decrease turnaround time to facilitate breeding programme management in a timely manner and set up a workflow that can be readily used as a service.

Future work is required to conduct a UK survey across a wide geographical scale and possibly identify any previously unknown remaining pockets of *A. m. mellifera.*

3. Assessing introgression in *Apis mellifera mellifera* colonies: pool-seq analyses outperform individual-based methods.

3.1 Introduction

Introgression can promote adaptive potential by increasing genetic variance or it can disrupt adaptation via the breakdown of co-adapted gene complexes (Hedrick, 2013; Lawson, *et al.,* 2017). Consequently, assessing and monitoring introgression is an important topic in evolutionary and conservation biology (Chapter 1). Before questions can be addressed such as *'when and in which circumstances does introgression break down local adaptation?'*, determining the best approach to measure introgression is necessary. However, there are different genotyping approaches available to measure genetic introgression and these approaches vary in the information they capture. Careful consideration of appropriate genetic approaches to assay introgression is especially relevant when studying eusocial organisms such as honey bees where queens are polyandrous and multiple patrilines exist (Inbar *et al.,* 2020).

Currently, there are two broad approaches for assessing introgression: assessing introgression at the individual level or by pooling individuals from the same colony ('pool-seq'). The incentive for the use of pools is distinctly different in this context to original use of the method. Traditionally, pool-seq has been used to sample a population of unrelated individuals and came about to reduce the cost of population

genetic studies (Gautier *et al.,* 2013; Rode *et al.,* 2018). But it can also be used to represent a colony as a composite genome, which can then go on to be used to investigate the collective colony genotype (Inbar *et al.,* 2020). In this method the pooled individuals are not independent samples from a population but a group of siblings or half-siblings. Pooled colony approaches have been used in studies of both honey bees and ants (Alalos *et al.,* 2020; Inbar *et al.,* 2020). In honey bee studies, pooled colony methods have been used to examine the genetic basis of traits such as aggression and calmness (Alalos *et al.,* 2020; Guichard *et al.,* 2021), and parasite defence (Guichard *et al.,* 2021). As well as using pools of workers (Henriques *et al.,* 2018; Regan *et al.,* 2018; Gmel *et al.,* 2021), studies of introgression in honey bees have also been conducted using individual drones (Pinto *et al.,* 2014; Munoz *et al.,* 2015; Parejo *et al.,* 2016) and individual workers (Jensen *et al.,* 2005; Munoz *et al.,* 2021)

A comparison of introgression values resulting from pooled and individual genotyping has been performed on a small scale specifically on the MassARRAY platform (Ellis *et al.,* 2017; Henriques *et al.,* 2018) during the development of the C-lineage introgression SNP array. As described in chapter 1, a SNP array was there developed to measure introgression from C-lineage honey bees (*A. m. carnica* and *A. m. ligustica*) to *A. m. mellifera* honey bees (Munoz *et al.,* 2015; Henriques *et al.,* 2018). SNP arrays are SNP genotyping microchips or microarrays (sometimes called 'SNP chips' or SNP arrays). These arrays are developed to aid cost-effective sequencing of known sets of polymorphic markers (LaFramboise 2009). The development of SNP arrays requires large-scale whole-genome genotyping from a diverse number of individuals (Peterson *et al.,* 2012). SNP arrays are designed for cost effective sequencing and are used in genome wide association studies (Liu *et al.,* 2014; Li *et al.,* 2008), introgression (Hernandez *et al.,* 2020; Przewieslik-Allen *et al.,* 2019) and

population genetic studies more generally (Zhang *et al.,* 2018; Perry *et al.,* 2020). The SNP arrays are largely designed for individual analysis but Henriques *et al*. (2018) examined the effects of pooled data on the SNP array using two methods: testing the sensitivity of the MassARRAY system to different pooled samples and by applying the SNP array to four colonies in both individual and pooled form. First, pooled samples were examined on the MassARRAY system (a SNP array platform). Combinations of 2, 3, 4, and 8 individuals were pooled and examined using three SNP combination (29 SNPs, 62 SNPs, and 117 SNPs). When pools of 8 individuals were tested on the SNP array introgression measurements deviated from what was expected (Henriques *et al.,* 2018, data from supplementary table 8). Secondly, they applied the SNP array to four colonies and four individuals belonging to that colony. Introgression values were consistently lower in the colony pool than in the individual (Henriques *et al.,* 2018 data from supplementary table 10).

Henriques *et al* (2018) noted that the finding merits further investigation using a larger sample size. So far this has been the only empirical test of introgression level differences between pooled colony and individual workers and it was specific to this SNP array platform. Regan *et al.,* (2018) also compared ADMIXTURE results from pooled whole genome sequencing to simulated individuals and found that up to K=3 the results were consistent with those compared to the pooled genotypes, while at K=4 two samples showed an assignment difference and at K=5 one additional sample showed a Q value difference. These results imply that at lower K values ADMIXTURE can be used on pooled samples.

Alongside the choice of how to best sample a colony, the choice of genotyping approach is also important. Common genotyping methods for studying introgression are mitochondrial sequencing (Bachtrog *et al.,* 2006; McGuire *et al.,* 2007; Darras and Aron 2015), microsatellite analysis (Randi and Lucchini 2002, Jug *et al.,* 2005;

Hutchinson *et al.,* 2018), restriction site-associated DNA sequencing (RAD-seq) (Kim *et al.,* 2018; Beeler *et al.,* 2020; Natola *et al.,* 2021) and whole genome sequencing (WGS) (Hanna *et al.,* 2018; Upadhyay *et al.,* 2021; Smeds *et al.,* 2021).

Microsatellites and mitochondrial DNA (mtDNA) markers are considered to be neutral markers and confer no fitness advantage, though microsatellites can be adjacent to coding regions and used to infer selection in some cases, and mtDNA codes for proteins that facilitate energy transformation (Bogenhagen *et al.,* 2003; Haasl *et al.,* 2012).

The mitochondrial genome (mtDNA) is usually maternally inherited in eukaryotic organisms (Wallace 2007) (although there are exceptions known, for example in chickens, Alexander *et al.,* 2015, and cicada species, Fontaine *et al.,* 2007). Microsatellites are polymorphic markers consisting of short sequence repeats distributed throughout the nuclear genome. Both mtDNA and microsatellite markers can be used to infer demographic processes (e.g. diversity, gene flow, inter population genetic relationships) and introgression, although the use of mtDNA in introgression studies has been called into question, since comparisons of introgression levels examined with the use of nuclear DNA and mtDNA have in some cases yielded conflicting results (Bonnet *et al.,* 2017). These discordant results have been observed across taxa, in fish (Willis *et al.,* 2014) invertebrates (Gompert *et al.,* 2008), mammals (Hedrick 2010; Melo-Ferreira *et al.,* 2009; Moa and Rossiter 2020), reptiles (Rato *et al.,* 2010; Zieliński *et al.,* 2013), and crustaceans (Lima *et al.,* 2019).

Conflicting results between nuclear and mtDNA are known to arise especially where historical hybridisation and subsequent back crossing into a parental group occur (Bonnet *et al.,* 2017) and have been observed in some honey bee studies (Garnery *et al.,* 1998; Pinto *et al.,* 2014; Ellis *et al.,* 2018). Regardless of the causes of these discrepancies, mtDNA may be an unreliable marker in the study of introgression. This

could be especially true in honey bees as this species is not naturally distributed, has undergone much human mediated movement, yet breed openly with the population around them and many of the natural boundaries have been broken down. Regarding microsatellite markers, SNP based methods have been seen to outperform microsatellites markers for estimating honey bee introgression (Muñoz *et al.,* 2017; Parejo *et al.,* 2018).

The most accurate method for assessing introgression would be to capture genome wide SNP markers. This can be achieved through a number of genotyping methods. 'RAD-seq' is one of a group of methods sometimes referred to as genotype-by-sequencing or reduced representation methods (these include, among others, exome capture or transcriptome sequencing). RAD-seq involves sequencing a subset of the genome using restriction enzymes (Peterson *et al.,* 2012). The use of restriction enzymes allows for the targeting of the same loci across many individuals or pools without prior knowledge of the genome (Peterson *et al.,* 2012). Importantly, as RAD-seq is sampling a subset of the nuclear genome, 25% to 2% depending on restriction enzymes chosen (Lowry *et al.,* 2017), the per-sample cost of sequencing is lower compared to WGS. RAD-seq can therefore allow an increased number of samples to be sequenced (Peterson *et al.,* 2012). The benefit of this is especially important in population genetics as some statistics and comparative analysis rely on a larger number of individuals or samples to be sequenced (Peterson *et al.,* 2012; Inbar *et al.,* 2020). While WGS costs have fallen greatly in the last 20 years it is still prohibitively expensive for most projects to sequence many samples (Inbar *et al.,* 2020).

Given the extent of imports of non-native honey bee subspecies (largely *A. m. carnica* and *A. m. ligustica*) into the UK (BeeBase, 2021) and the possibility of introgression and loss of local adaptation in the face of this (Chapter 1), the aim of this study is to

compare introgression measurements in colonies in the south west of England across three genotyping approaches. Colonies were assessed using an individual and pooled method. The same individual was processed twice, in both the SNP array and the individual RADseq approach and a pool of 30 workers from the same colony was processed in pooled colony RADseq approach. These were then assessed using the ADMIXTURE programme and an ABBA BABA approach.

3.2 Methods

3.2.1 Sampling

Honey bees were sampled in the south west of England (Figure 3.1). Bee keepers were contacted through a number of special *A. m. mellifera* interest bee keeping groups. The Bee Improvement Programme for Cornwall (BIPCo), the B4 Project and the Bee Improvement and Bee Breeders association (BIBBA) were all involved in contacting bee keepers. Bee keepers were asked if they suspected they had native *A. m. mellifera* or near native honey bees, and did not import.

Sampling tubes of 70% ethanol were sent to bee keepers to sample colonies. Bee keepers were asked to provide 35 worker honey bees from each colony. Samples were then posted back to the University of Plymouth for processing.

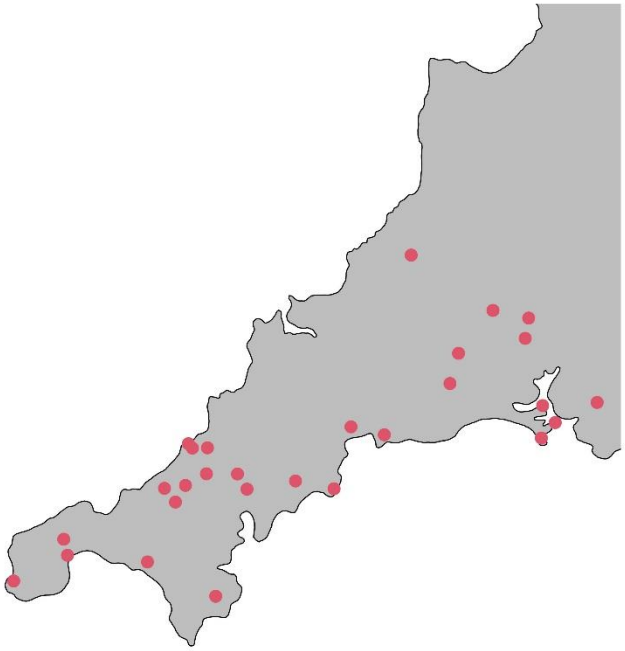


Figure 3.1. Sampling locations of samples collected across the South West of England.

3.2.2 Overview of sequencing approaches

This study is comprised of novel data and data downloaded from the sequence read archive (SRA; Table 3.1). Novel sequence data (see below) were generated for bees sampled from SW England by pooled RAD-Seq, individual RAD-Seq and iPLEX, Data downloaded from SRA used other platforms and sampling approaches (Table 3.1). To account for this, different bioinformatics workflows were implemented to generate the data required for analysis: (1) subspecies standards and SW ddRADSeq data using pooled samples (Figure 3.2); (2) subspecies standards and SW ddRADSeq for individuals (Figure 3.2); (3) combination of these data with WGS outgroup data (from *A. cerana;* Figure 3.3). An outgroup was necessary for later ABBA BABBA analysis (Table 3.1). The subspecies standards, generated using pooled colonies and individuals are necessary for assessing introgression in the pooled colony and individual South West samples. Throughout this chapter the pooled colony samples

are referred to as ‘colony’ samples and the individual worker samples are referred to as ‘individual’ samples.

Table 3.1 Overview of data sets used in this study. The samples represented three groups. The South West represents the putative population of *A. m. mellifera* in the South West of England. Subspecies standards, comprising of three subspecies from across Europe (*A. m. carnica, A. m. ligustica* and *A. m. mellifera*) and the outgroup representing the phylogenetic outgroup *A. cerana*. Genotyping methods used were restriction site associated DNA sequencing (RADseq), ancestry informative markers in the form of single nucleotide polymorphism (AIMs SNPs)(Henriques *et al.,* 2018) and whole genome sequencing (WGS). Sequencing platforms used to generate the data were Illumina (Novaseq and HiSeq 2500), a MassARRAY iPLEX , a Beijing Genomics Institute (BGI) platform and a SOLiD platform which generates colorspace data. Sampling approaches are either pooled colony, where 30 workers from the same colony are pooled for sequencing, or individual, where a single worker is sequenced

Samples representing	Genotyping method	Sequencing platform	Sampling approach	Data source
South West	RADseq	Illumina (Novaseq)	Pooled colony	Generated in this study
South West	RADseq	Illumina (Novaseq)	Individual worker	Generated in this study
South West	AIMs SNPs	iPLEX?	Individual worker	Generated in this study
Subspecies standards	WGS	BGISEQ-500	Pooled Colony	Generated in this study
Subspecies standards	WGS	SOLiD 5500xl	Individual worker	Downloaded (SRA)
Outgroup	WGS	Illumina(HiSeq 2500)	Individual worker	Downloaded (SRA)

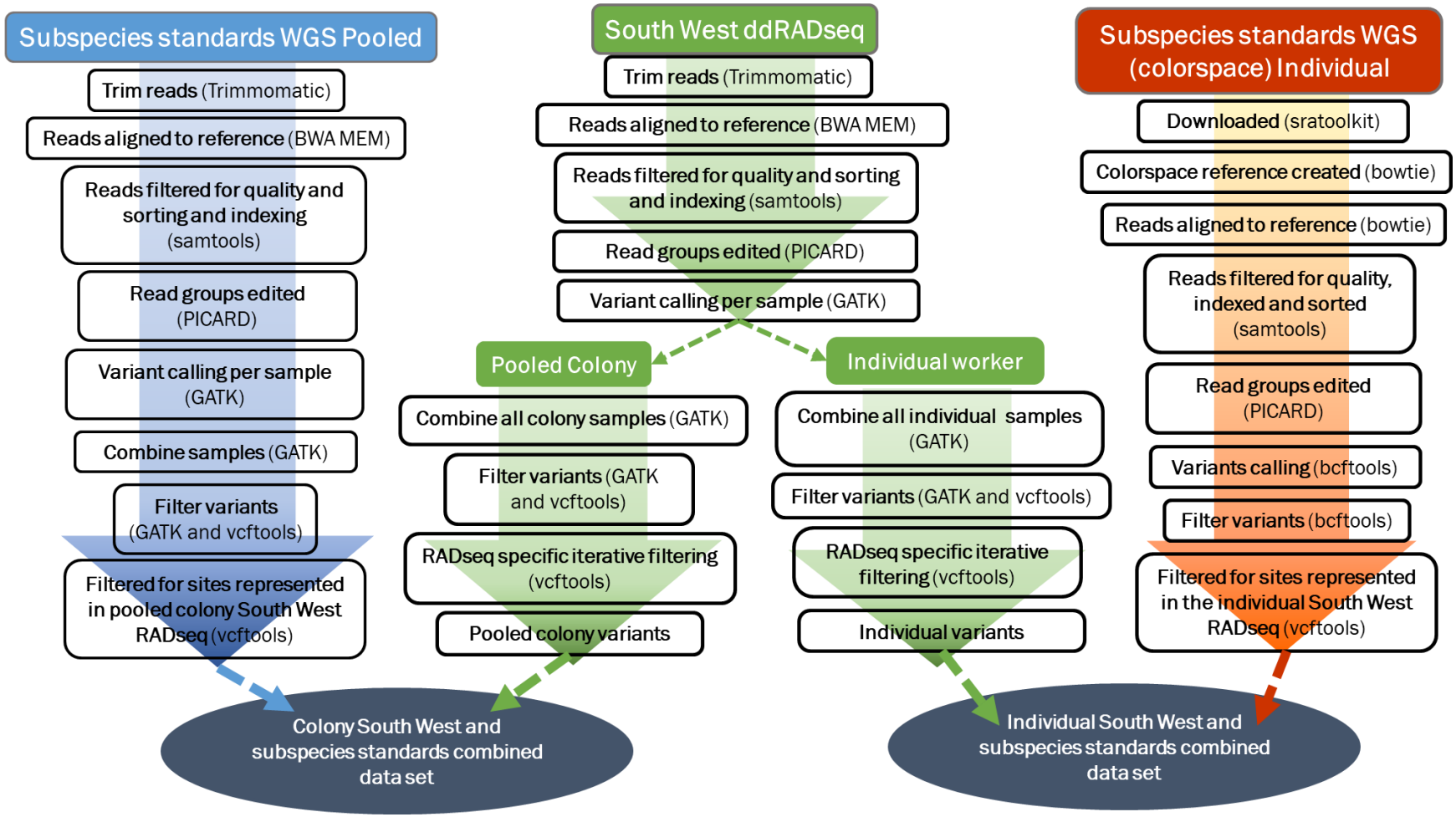


Figure 3.2 Overview Overview of South West England ddRADseq samples and subspecies standards bioinformatics workflows. Samples were processed based on the sampling approach

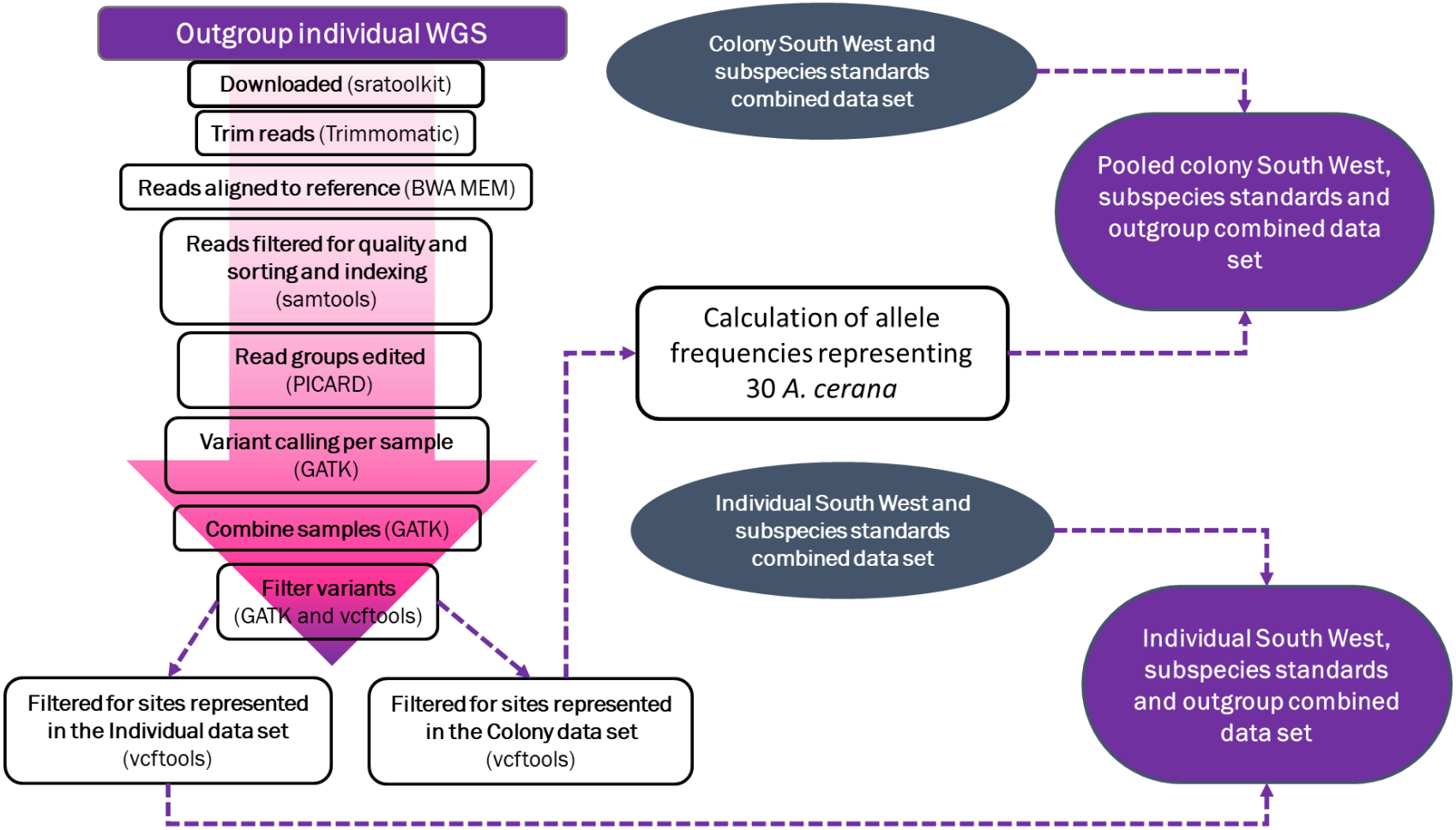


Figure 3.3. Overview of outgroup bioinformatics workflows and merging with ddRADseq and subspecies standards datasets previously created in Figure 3.2

3.2.3 Restriction site associated DNA methods

3.2.3.1. Library preparation overview

Library preparation is the process of preparing DNA sequences to be read by the sequencing platform and allow samples to be identified in downstream analysis. Library preparation of individual and pooled colony bees sampled from south west England was adapted from Peterson *et al.,* (2012). The process for both the individual and pooled colony double digest restriction associate DNA sequencing (ddRADseq) consisted of DNA extraction, DNA double enzyme digestion, adaptor annealing, adaptor ligation, size selection and a PCR library enrichment (Figure 3.4).

Briefly, genomic DNA is extracted and then digested using two restriction enzymes resulting in a DNA fragment with enzyme cut sites at either end (Figure 3.4 A) . Two oligos (single strands of synthetic DNA) are annealed together to form an adaptor. Two adaptors are constructed, called P1 and P2 (Figure 3.4 B). The adaptors are designed to adhere to the enzyme cutting sites. The P1 adaptor contains one of twelve unique barcodes and each sample (either DNA from an individual worker bee or a pool of workers) receives one of these barcodes. The P2 adaptor is a common adaptor (i.e. contains no sequence variation). The adaptors are ligated to the enzyme cleaved DNA fragments which adds a barcode to each DNA samples' fragments along with a primer binding site (Figure 3.4 C). Multiple DNA samples are placed into batches. Each batch consist of twelve samples, every sample in a batch has a unique P1 barcode ligated to it (numbered from one to twelve) (Figure 3.4 D). These batches are then placed through a size selecting process. This process filters the adaptor ligated fragments for a length of base pairs that is appropriate for use with the chosen sequencing platform (Figure 3.4 D). The batches then go through a PCR amplification (Figure 3.4 E). This is performed to add sequencing platform annealing sequences, sequencing platform primers regions and a unique index for each batch (Figure 3.4

E). This index in combination with the twelve unique P1 adaptors allows for the multiplexing (a large number of samples to be sequenced in one run while allowing for the identification of the samples at the end of sequencing) (Figure 3.4 F). The final library is ready for equimolar mass pooling, where all samples are combined into one final multiplexed library (Figure 3.4 F).

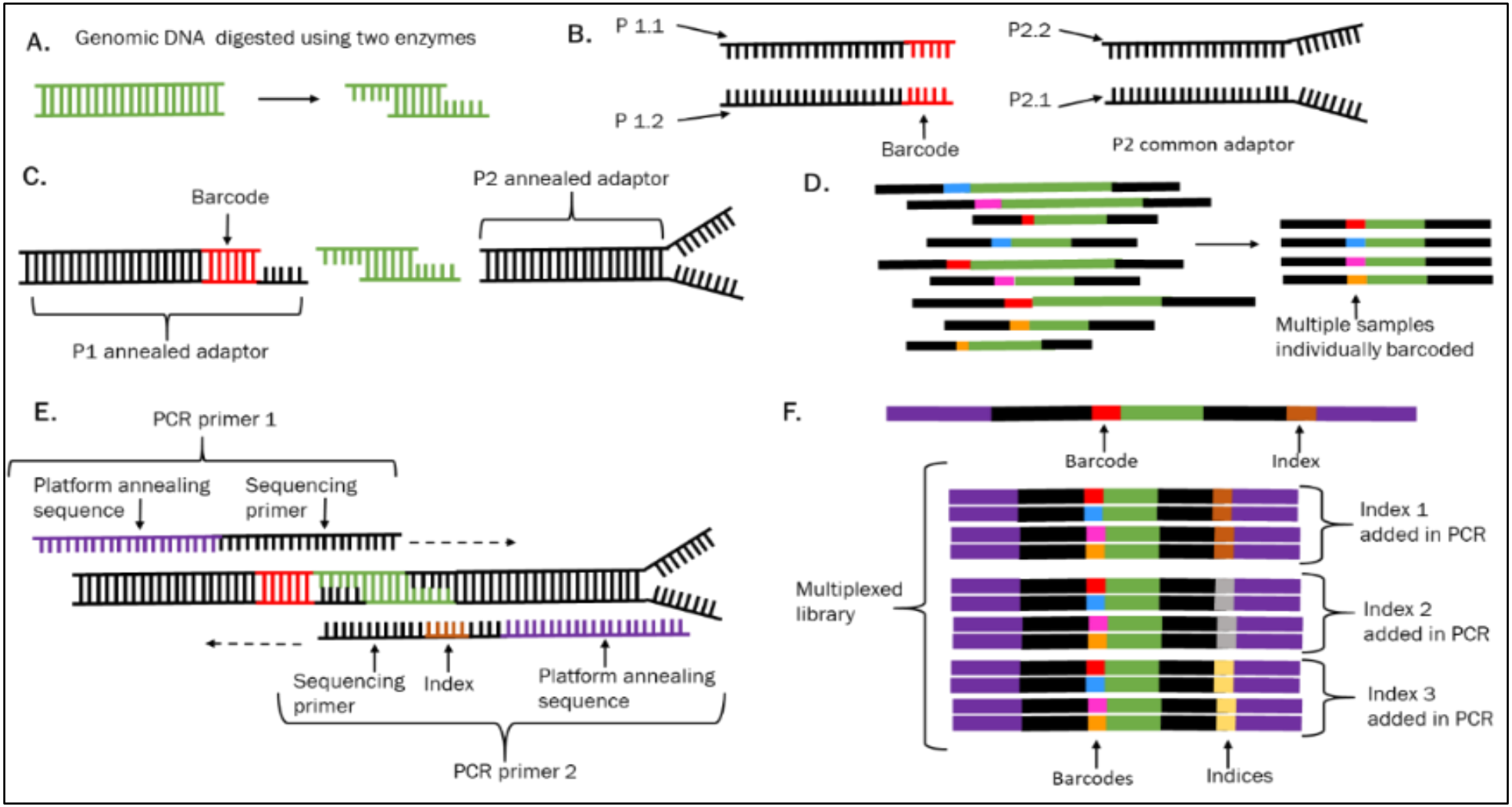


Figure 3.4. An overview of library preparation for the RADseq protocol used in the pooled colony and individual samples from the south west

3.2.3.2. DNA extraction

DNA was extracted from a standardised weight of thorax tissue from individual bees using an ammonium acetate protocol (Nicholls *et al.,* 2000). For colony extracts every thorax was weighed and all honey bees donated an equal amount of tissue to each extraction. Five thoraxes were pooled per extraction and six extractions were performed per colony. This approach was used due to practicalities in the laboratory. The pooled thoraxes were placed in 250µl of digestion buffer Digsol (20ml of 0.5 EDTA at pH 8.0, 3.425g NaCl, 25ml 1M Tris-HCl at pH 8.0, 430ml pure water and 25ml 20% SDS), with 25µl of Proteinase K and placed in a 55°C oven overnight. After digestion 250µl of ammonium acetate (pH 7.5) was added and mixed by vortexing four times over a 25 minute period. The extractions were then centrifuged at maximum speed (13,200 rpm) the supernatant was then aspirated into a clean tube and the pellet discarded. Two washing steps then followed, first adding 1000µl of ice cold 100% ethanol, inverting several times to mix then storing at -20°C for 2 hours. The extractions were then again centrifuged at 13,200 rpm for ten minutes and the supernatant was discarded. For the second wash the same procedure was repeated using 70% ethanol and without a freezing period. The pellet was then dried on a heat block at 70°C before being resuspended in 20µl molecular grade water, incubated at room temperature for 2 hours. All colony extractions were then treated with 4µl RNase (New England Biolabs) in each extraction and incubated for 2 hours at 37°C. The individual extractions were placed through the same protocol with two minor variations. The individual thoraxes were incubated with 12µl of Proteinase K and treated with 2 µl of RNase. DNA extractions were quality checked using gel electrophoresis. Gels were examined for a high molecular weight band with a low smear, indicating largely intact high quality genome extraction. All DNA extractions

were then kept at -20°C for long term storage. DNA extractions for pooled samples yielded an average of 1357ng of DNA and Individual extractions yielded an average of 124ng of DNA.

3.2.3.3. Double enzyme digest

DNA extractions were quantified using a Qubit Fluorometer® 2.0. For the colony samples the six extraction were equimolarly pooled to create a final DNA pool representing 30 workers. Each of the six pooled extractions donated 166.66ng of DNA for a total of 1000ng total DNA content representing each colony. The individual worker sample extractions donated 10ng of DNA to the digest. In order to remove any remaining unwanted contaminants and to obtain the volumes required, all samples were cleaned up using High Prep PCR Clean-up System (MAGBIO). Pooled colony DNA samples were eluted to a volume of 40 µl and individual DNA samples were eluted into a volume of 8µl ready for the double digest. The pooled colony extracts were then digested with restriction enzymes at 37°C for 3 hours in a 50µl reaction volume. The colony digest contained 1.5µl MluCl and 1.5µl MspI (restriction enzymes), 5µl CutSmart Buffer (10x) (New England Biolabs), 2µl of molecular free water and 40µl x 1000ng pool of DNA. Individuals were similarly digested but in 10µl reaction volumes containing 0.5µl MluCl, 0.5µl MspI, 1µl of cutsmart buffer (10x) (New England Biolabs) and the 8µl of 10ng DNA. All samples were then again cleaned using HighPrep PCR Clean-up Stystem (MAGBIO) beads at 1.2 x the volume of the sample in order to remove any leftover enzymes or salts. Both colony and individual DNA samples were eluted off with 25µl of molecular grade H2O.

3.2.3.4. Adaptor annealing

Each of the 12 different P1 adaptors was produced by annealing the complimentary 1.1 and 1.2 oligos together. The P2 adaptor was also annealed by combining its 2.1

and 2.2 oligos together (3.3 B). Annealing was performed to a final concentration of 40mM. Adaptor annealing was performed in a reaction volume of 100µl, each containing 40µl of adaptor P1.1 (100mM), 40µl of adaptor P1.2 (100mM), 10µl of annealing buffer (100Mm of Tris HCL pH 8.0, 500mM of NaCl and 10mM of EDTA) and 10µl of molecular grade H_2O. Reactions incubated at 97°C for 2 minutes 30 seconds, then cooled at a rate of 1.5°C every 30 seconds until they reached 21°C. Ligated adaptors were then held at 4°C and subsequently stored in the fridge.

3.2.3.5. Adaptor ligation

All samples had adaptors ligated to the enzyme digested DNA. This was performed in batches. Each batch was made up of 12 samples with each being assigned one of the 12 uniquely barcoded P1 adaptors. The batches were not mixed in terms of extraction approaches, i.e. the batches consisted either solely colony samples or solely individual samples. All samples were ligated to the common P2 adaptor.

The individual DNA samples had adaptors ligated in a 40µl reaction containing 0.4µl rATP 100mM (Promega), 1µl assigned P1 adaptor for that sample (4mM), 1µl P2 (4mM), 0.5 ligase (New England Biolabs), 4µl ligation buffer (New England Biolabs), 25µl of enzyme digested DNA and 8.1µ molecular grade H_2O. The colony samples adaptors were ligated 0.4µl rATP 100mM (Promega), 2µl assigned P1 adaptor for that sample (4mM), 2µl P2 (4mM), 0.5 ligase (New England Biolabs), 4µl ligation buffer (New England Biolabs), 25µl of enzyme digested DNA and 6.1µl molecular grade H_2O.

All reactions were placed into a thermocycler and incubated at 23°C for 30 minutes, 65°C for 10minutes then decreased in temperature at a rate of 2°C every 1 minute 30 seconds until reaching 23°C.

3.2.3.6. Pooling barcoded samples

Both the colony batches and the individual batches, consisting of the now barcoded samples each are pooled into a single tube, where each sample within is identifiable by the P1 barcode. These pooled batches were then cleaned up using High Prep PCR Clean-up (MAGBIO) at 1.2x the sample volume in order to effectively remove any unligated or adaptor-adaptor ligated products. The pooled batches samples were eluted off at a volume of 30µl ready for insertion in to the Pippin Prep for size selection.

3.2.3.7. Fragment size selection

All batches were taken through a size selection process using a Pippin Prep (www.sagescience.com). The Pippin prep machine was set to elute off read lengths between 150bp to 500bp.

3.2.3.8. PCR amplification

A PCR was performed on each batch separately using a Phusion PCR kit (Thermofisher). The purpose of this PCR step to enrich the library sequences, to add flowcell annealing sequences, primers regions specific to the Illumina platform, and multiplexing indices to all fragments (Figure 3.4 F). A flowcell is a component of the Illumina sequencing platform. Flowcells are channels made up of a number of lanes (1 to 4 lanes per flowcell) which the library is loaded on to. Each lane has adaptors attached to the surface which bind to the complimentary flowcell annealing sequence. Primer regions are allow for the amplification of fragments during sequencing. Finally, the multiplexing the batches allows the entire librbary to be combind by assigning a uniquely indexed reverse primer to each batch to create a unique combination of index and P1 combinations. Both individual and colony batched PCR reactions were set up in volumes of 25µl each containing 1µl of reverse primer, 0.25µl phusion polymerase, 10µl of pooled adaptor ligated size selected DNA,

5µl of phusion buffer (5x), 0.5µl dNTPs (10mM), 1µl forward primer, 0.75µl DMSO and 6.5µl molecular free H2O. The reactions were then placed into a thermocycler for 98°C for 3minutes, then 16 cycles at 98°C for 1 minute, 63.5°C for 1minute 30seconds and 72°C for 1 minute, the cycles were followed by a final extension of 72°C for 3 minutes and an infinity hold at 4°C. After PCR the entire library was equimolarly pooled and quantified.

3.2.4 Sequencing and bioinformatics

3.2.4.1. Sequencing

A total of 36 pooled colonies and 33 individual samples were processed and sent for sequencing. Sequencing of the prepared library was performed in Hong Kong at the Beijing Genomic Institute (BGI) on an Illumina Novaseq platform using 150bp paired end reads.

3.2.4.2. Data processing and filtering

Raw reads were de-multiplexed in stacks with basic quality filters. The demultiplexed data consisted of an average of 14 million reads per pooled colony samples and 7 million reads per individual sample. Using Trimmomatic (version 0.39) (Bolger *et al.,* 2014), reads were trimmed at the ends if quality dropped below a quality score of 4, and a 4 base pair (bp) sliding window trimmed sections if the average quality dropped below 15 (Bolger *et al.,* 2014). Unpaired reads were discarded. Paired reads were mapped to the *Apis mellifera* reference genome (Amel_HAv3.1 assembly) using the Burrow-Wheeler Aligner (BWA) MEM aligner (version 0.7.17) (Li 2013). Reads were discarded if they aligned to more than one position on the reference genome. The subsequent alignment files were converted to the bam file format (Li *et al.,* 2009). Samtools was then used to filter reads with a mapping quality score >20 and the files were sorted (by genomic coordinates, staring with the first position on chromosome

one and ending with the last position chromosome 16) and indexed (where overlapping reads are placed together for faster accessing of the files information) (Li *et al.,* 2009). Each individual or colony sample's data bam file was edited to create one read group per sample (PICARD). The data were then processed using the genome analysis tool kit's (GATK, version 4.1.9.0) best practices pipeline (Van der Auwera and O'Connor, 2020). Firstly, Haplotype Caller was used to create the genomic variant call format (gvcf) files for every sample (both colony and individual) (Poplin *et al.,* 2017). Then GATK's GenomicsDBImport was used to create two intermediate databases, one for colony data samples and one for individual data samples.

This was followed by the creation of a vcf per chromosome, one for colony data samples and one for individual data samples, using GATK's GenotypeGVCFs, then the per chromosome vcfs where merged into one vcf containing all samples and all sites using GatherVcfs (Poplin *et al*., 2017). Using GATK's SelectVariants command, indels were removed and only SNP variants were kept.

Following the data processing a filtering pipeline was implemented. The pipeline consisted of initial variant filtration performed using the recommended GATK hard filtering thresholds (Table 3.2), then an iterative filtering method was performed in vcftools (version 0.1.16) to prevent erroneous SNP calls (O'Leary *et al.,* 2018; Danecek *et al.,* 2011). This robust filtering mitigates errors in downstream data analysis that can be caused by allelic dropout (O'Leary *et al.,* 2018; Cerca *et al.,* 2021). As well as quality filtering, SNPs were first filtered for depth and filtered based on their proportion of representation across all samples. Samples were filtered based on the proportion of all SNPs they contained. SNPs were filtered with the ' --max-missing' command, and samples were filtered using the '--missing-indv' command in

vcftools (Table 3.2). Finally, any monomorphic sites were removed using a custom python script.

Table 3.2. Data filtering performed in GATK and vcftools. GATK filtering is based on recommended filtering parameters from GATK best practices pipeline. vcftools filtering based on O'Leary *et al* (2018)

Filter name and programme	Filter description	Threshold
QualByDepth (GATK)	Variant confidence divided by the unfiltered depth. Filtered at GATK recommended value.	< 2
Quality (GATK)	Variant quality confidence. Filtered at GATK recommended value.	< 30
FisherStrand (GATK)	Phred-scaled probability of strand bias. Filtered at GATK recommended value.	> 60
StrandOddsRatio (GATK)	Strand bias test that compensates for where FisherStrand filter penalises variants at the end of exons. Filtered at GATK recommended value.	> 3
RMSMAppingQuality (GATK)	Root mean square mapping quality over all reads at each site. Filtered at GATK recommended value.	< 40
MappingQualityRankSumTest (GATK)	Rank sum test for mapping qualities. This compares mapping qualities of the reads supporting the reference allele at the alternate allele. Filtered at GATK recommended value.	< - 12.5
ReadPosRankSumTest (GATK)	The rank sum test for site position within reads. This compares whether positions of the reference and alternate alleles are different within the reads. Filtered at GATK recommended value.	< - 8.0
--remove-indels (vcftools)	Remove any insertions or deletions in the data, leaving only SNPs	All removed
--maxDP and –minDP (vcftools)	Filters sites based on read depth, removing any below the minimum threshold and above the maximum threshold.	>5 and <500
--minQ (vcftools)	Retain sites with quality value above this threshold.	>20
--max-missing (vcftools)	SNPs excluded based on a proportion of missingness across all samples	>0.5
--missing-indv (vcftools)	Samples excludes based on a proportion of missing SNPs	<0.9
--max-missing (vcftools)	SNPs excluded based on a proportion of missingness across all samples	>0.6

Table 3.2. Data filtering continued

Filter name and programme	Filter description	Threshold
--missing-indv (vcftools)	Samples excludes based on a proportion of missing SNPs	<0.7
--max-missing (vcftools)	SNPs excluded based on a proportion of missingness across all samples	>0.7
	Data filtering continued	
--missing-indv (vcftools)	Samples excludes based on a proportion of missing SNPs	<0.5
--missing-indv (vcftools)	Samples excludes based on a proportion of missing SNPs	>0.25
--max-missing (vcftools)	SNPs excluded based on a proportion of missingness across all samples	<0.95
Filter_monomorphic.py	Custom python code that removed any monomorphic sites	All removed

3.2.5 AIMs SNP Array

After DNA was extracted to perform the individual RADseq the rest of the honey bee tissue was sent to the Roslin Institute. The Roslin institute performed DNA extraction and library preparation for the SNP array platform. The SNP array is accompanied data was accompanied with standards belonging to the Henriques *et al* (2018) assay. The standards supplied are from across Europe. *A. m. mellifera, A. m. carnica* and *A. m. ligustica* honey bees.

3.2.5.1. AIMs SNP Array variant filtering

SNP array raw data were formatted using a custom python code into plink ped and map format for missingness filtering (Purcell *et al.,* 2007). Genotypes were filtered in plink (version 1.07) to obtain a genotyping rate of 0.9 using the '–geno' command (Parejo *et al.,* 2016; Wragg *et al.,* 2016). Samples were filtered to contain a proportion of 0.9 of all SNPs using the '–mind' command. Data were then converted to the binary bed format using the '–make-bed' command in plink.

3.2.6 Processing of subspecies standards and outgroup data

3.2.6.1. Generation of subspecies standards

3.2.6.1.1.Individual subspecies standards data

Publically available data from subspecies standards were obtained for comparison with the individual RADseq samples. Whole genome data from Wallberg *et al.,* (2014) was downloaded from the Sequence Read Archive (project number PRJNA236426) using the sratoolkit (version 2.11.1). The downloaded data were from *Apis mellifera mellifera* (n=20) originating from Sweden and Norway, *Apis mellifera ligustica* (n=10) from Italy and *Apis mellifera carnica* (n=9) from Austria (Table 3.3). These data were

generated using a SOLiD 5500xl platform (Life Technologies), which generates colorspace data. Due to the format of the raw data from this sequencing platform a different bioinformatics procedure was required. Samples were run across multiple lanes and consisted of multiple colorspace fasta files per sample. A colorspace reference was constructed from the reference genome Amel_HAv3.1 assembly using bowtie's 'build' command (version 1.2.3) (Langmead *et al.,* 2009). The reads were then aligned to the colorspace reference using bowtie, generating output in the sam format (Li *et al.,* 2009). Samtools (version 1.10) was then used to convert the reads to the bam format, merge the multiple bam files into their corresponding biological samples, then sort, and index the reads (Li *et al.,* 2009). Read groups were edited using PICARD (via GATK 4.2.0.0) and bcftools (version 1.8) resulting in an mpileup file, and a vcf file (Danecek, *et al.,* 2021). Vcftools (Danecek, *et al.,* 2011). Colorspace data is prone to higher error rates than Illumina platform sequencing. Cridland *et al.* (2017) examined this data set and found that the elevated error rate was associated with an excessive number of calls for triallelic sites when compared to background rates of triallelic calls in *Drosophila* genomes. They also found that higher sequence coverage was associated with this higher error rate when compared to an Illumina dataset. To control for this, triallelic sites were removed by filtering the data to contain only biallelic sites and none of the high coverage sequenced samples were used. Additionally, data were filtered for genotype missingness (>0.9) and sample missingness (>0.9). The vcf was finally filtered to contain only sites that were present in the individual RADseq data. Using bcftools the data set was merged with the individual RADseq data for ADMIXTURE downstream analysis.

Table 3.3. Sample accession numbers for individual worker whole genome samples of European subspecies from Wallberg *et al.,* (2014).

SRA sample accession code	Subspecies
SAMN02596338	*A. m. carnica*
SAMN02596340	*A. m. carnica*
SAMN02596341	*A. m. carnica*
SAMN02596342	*A. m. carnica*
SAMN02596343	*A. m. carnica*
SAMN02596344	*A. m. carnica*
SAMN02596345	*A. m. carnica*
SAMN02596346	*A. m. carnica*
SAMN02596347	*A. m. carnica*
SAMN02596288	*A. m. ligustica*
SAMN02596289	*A. m. ligustica*
SAMN02596290	*A. m. ligustica*
SAMN02596291	*A. m. ligustica*
SAMN02596292	*A. m. ligustica*
SAMN02596293	*A. m. ligustica*
SAMN02596294	*A. m. ligustica*
SAMN02596295	*A. m. ligustica*
SAMN02596296	*A. m. ligustica*
SAMN02596297	*A. m. ligustica*
SAMN02596328	*A. m. mellifera*
SAMN02596329	*A. m. mellifera*
SAMN02596330	*A. m. mellifera*
SAMN02596331	*A. m. mellifera*
SAMN02596332	*A. m. mellifera*
SAMN02596333	*A. m. mellifera*
SAMN02596334	*A. m. mellifera*
SAMN02596335	*A. m. mellifera*
SAMN02596336	*A. m. mellifera*
SAMN02596337	*A. m. mellifera*
SAMN02596414	*A. m. mellifera*
SAMN02596415	*A. m. mellifera*
SAMN02596416	*A. m. mellifera*
SAMN02596417	*A. m. mellifera*
SAMN02596418	*A. m. mellifera*
SAMN02596419	*A. m. mellifera*
SAMN02596420	*A. m. mellifera*
SAMN02596421	*A. m. mellifera*
SAMN02596422	*A. m. mellifera*
SAMN02596423	*A. m. mellifera*

3.2.6.1.2.Pooled colony subspecies standards data

Standards required for the analysis of the pooled colony data were obtained from whole genome sequencing. This data set consisted of *Apis mellifera mellifera* (total n=28) sampled in Sweden (n=11), Norway (n=10) and Switzerland (n=7), *Apis mellifera ligustica* (total n=15) from Italy (n=5) and Sweden (n=10), and *Apis mellifera carnica* (total n=26) from Germany (n=7), Sweden (n=9) and Norway (n=10).

DNA extraction was performed using the same protocol as the pooled RADseq data (section 2.1.1.1). Library prep and sequencing was performed at BGI. This library prep is patent protected (Fang *et al.,* 2018; https://www.protocols.io/view/bgiseq-500-library-construction-protocol-3byl458ovo5d/v1). In brief, genomic DNA was fragmented using a Covaris ultrasonicator (Covaris). Fragment sizes were examined by agarose gel electrophoresis or on an Agilent 2100 BioAnalyzer (Agilent). The sheared DNA was then cleaned up and size selected using AMpure XP magnetic beads (Beckman Coulter). End repair and A-tailing was performed using KAPA HyperPrep enzyme mix and buffer (Roche) to produce fragment ends ready for adaptor ligation. The adaptors were ligated with specific barcodes using KAPA kits in a thermocycler (KAPA biosystems, Roche). The adaptor ligated DNA samples were cleaned up using AMpure XP magnetic beads to remove excess adaptors and fragments, before a PCR enrichment stage followed by another clean up. The library was loaded onto the BGISEQ-500 machine and sequenced at 100bp paired end sequencing.

The raw data were received from BGI in fastq format. Trimmomatic (Bolger *et al.,* 2014) was used to remove the ends of reads if they dropped below a threshold quality of 4. Reads were removed if they were below a length of 50bp or had no paired read. A sliding window of 4bp trimmed reads if the average quality of the window dropped

below a threshold of a phred score of 20. After trimming the same procedure as the RADseq data (section 2.1.2.2) was followed. The vcftools filtering was not performed in the same way as for the RADseq data as the iterative filtering system is designed to stop erroneous calls in RADseq data, not for whole genome data. Instead, the data were first filtered for excessive depth. The mean depth (I.depth) was visualised in R, and sites were removed at cut-off point of double the mean depth (minimum depth 5, maximum depth 65) (Danecek *et al.,* 2011; Purcell *et al.,* 2007). The data were then filtered for missingness. SNPs were required to be represented in 0.9 proportion of all samples (vctools max-missing) and each sample was required to contain a proportion of 0.9 of all SNPs in the data set (vcftools missing-indv)(Danecek *et al.,* 2011). Finally, the data were filtered for sites that are present in the RADseq data and merged using bcftools.

3.2.6.2. Generation of outgroup data

The outgroup chosen for the ABBA BABA analysis was *Apis cerana*, the eastern honey bee. This has been used as an outgroup for *A. mellifera* in previous studies (Cridland *et al.,* 2017; Han *et al.,* 2012; Chen *et al.,* 2018). Chen *et al* (2018) prepared a paired-end library of whole genome DNA that was sequenced on an Illumina HiSeq 2500. Using the sratoolkit, 30 *Apis cerana* worker bee samples (Table 3.4) were downloaded in fastq format (Project accession PRJNA418874). Here, the bioinformatics preparation was identical to the pooled colony whole genome data (section 2.2.1.2). Importantly, the reads were aligned to the *Apis mellifera* genome for direct comparison of sites, therefore not all reads align and this limited the sites represented across the data sets. The resulting vcf file was filtered in the same manner as the whole genome data and then filtered to contain the RADseq sites. Two files were created, one for individual and one for the pooled data. The vcf file that contained the individual RADseq sites was then merged with the individual data. The

A. cerana vcf file that contained the pooled colony RADseq sites was not merged with the pooled data but used to generate the population allele frequencies for every site representing all 30 *A. cerana* individuals (Danecek *et al.,* 2011).

Table 3.4. Accession numbers of *A. cerana* samples downloaded from the SRA archive.

Biological Samples Downloaded (Project accession PRJNA418874)
SRR6301264
SRR6301265
SRR6301266
SRR6301267
SRR6301268
SRR6301269
SRR6301270
SRR6301271
SRR6301272
SRR6301273
SRR6301274
SRR6301275
SRR6301276
SRR6301277
SRR6301278
SRR6301279
SRR6301280
SRR6301281
SRR6301282
SRR6301283
SRR6301284
SRR6301285
SRR6301286
SRR6301287
SRR6301288
SRR6301289
SRR6301290
SRR6301291
SRR6301292
SRR6301293

3.3 Introgression estimators overview

Subspecies standards and outgroup data were required to examine introgression in the south west samples. Two methods were used to assess introgression: ancestry clustering using ADMIXTURE (Alexander, *et al.,* 2009) and a genome wide test for introgression based on incomplete lineage sorting using Patterson's D statistic and *f* statistic, also known as an ABBA BABA test (Green *et al,.* 2010;Patterson *et al.,* 2012).

3.3.1 ADMIXTURE clustering program as an introgression estimator

ADXMITURE (Alexander *et al.,* 2009 ; version 1.3.0) is a clustering algorithm that uses a maximum-likelihood model to estimate sample ancestry. ADMIXTURE estimates ancestry membership proportions (Q values) of individuals to clusters that represent ancestral populations (K). ADMIXTURE uses a cross-validation (CV) procedure to inform the most likely K value for the data. The most likely K value will exhibit a lower CV value than other K values (Alexander, Novembre and Lange, 2009). For this program to be used for introgression estimation, standards are included in the analysis to view the clusters they form and the samples' membership to these clusters.

ADMIXTURE was run on bed files. For the Individual and pooled colony RADseq bed files were created from the vcf files using the vcftools command –plink (Danecek *et al.,* 2011) and then the plink command –make-bed (Purcell *et al.,* 2007). The SNP Array data raw data was received in an excel format and changed to a .ped and .map file using a custom python script. The .ped and .map files were then converted to .bed files using the plink command using the plink command –make-bed (Purcell *et al.,* 2007). All ADMIXTURE runs were performed using default settings.

3.3.2 The ABBA BABA approach

The ABBA BABA approach in contrast involves fitting four populations onto a phylogenetic tree and is based on examining derived and ancestral allele patterns brought about by incomplete lineage sorting (ILS) versus patterns of gene-flow and introgression (Paterson *et al.,* 2012; Durand *et al.,* 2011; Green *et al* 2010; Reich *et al.,* 2010). ILS occurs when species or lineages undergo diversification into separate groups but there has been insufficient time for complete genetic differentiation of those groups. As a result the gene tree differs from the overarching lineage, species tree and as the alleles are not perfectly segregated into those diversified groups. (Carstens and Knowles, 2007; Degnan and Rosenberg, 2009).

ABBA BABA employs two related statistics: Paterson's D and *f*. Paterson's D examines deviations from the expected patterns of alleles resulting from ILS. The D statistic compares SNPs across the genome between three in-group populations (P1, P2, P3) and one outgroup population (Po) that match ABBA and BABA genotype patterns (Figure 3.5). An ABBA pattern is where population P1 has the ancestral allele (represented by 'A'), while P2 and P3 share a derived allele (represented by 'B') (Figure 3.5). A BABA pattern is when P2 has the ancestral allele and P1 and P3 share the derived allele. Counting the occurrences of these patterns across all sites allows us to investigate if the total number of shared derived alleles between two populations is greater than expected by chance. Effectively we are asking if P1 and P3 share an excess of derived alleles or if P2 and P3 share an excess of derived alleles.

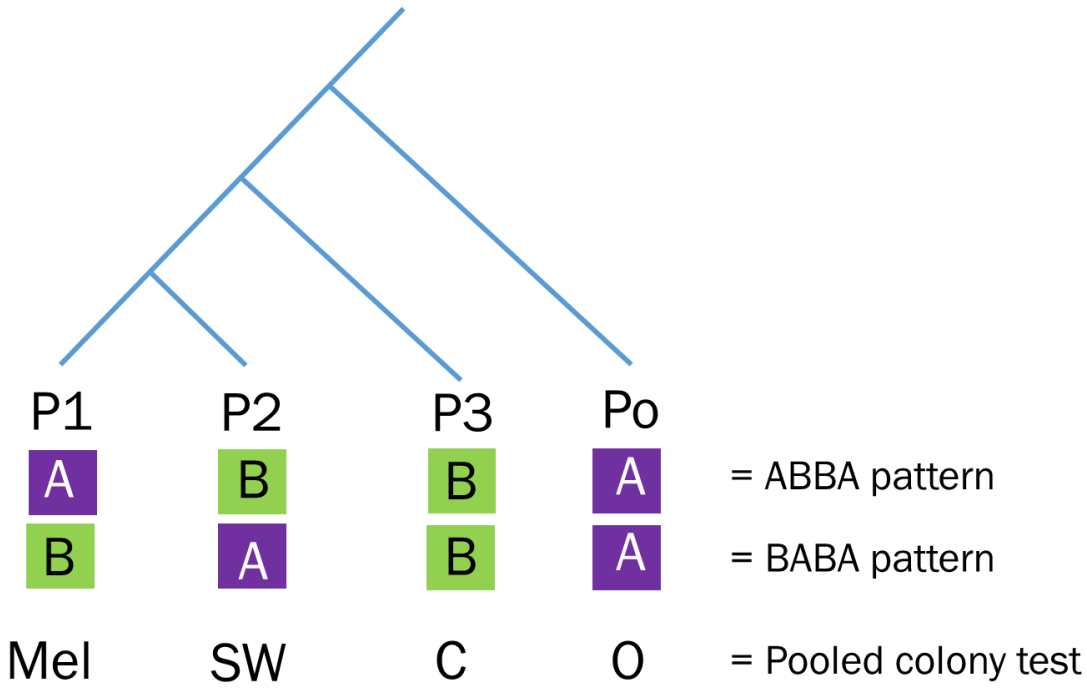


Figure 3.5 The principle behind the Paterson D and f statistic. Example of allele patterns of derived ('B') and ancestral ('A') alleles across the P1, P2, and P3 groups. The ABBA pattern is when P2 and P3 share the derived allele, while P1 has the ancestral allele from the outgroup. The BABA pattern is when P1 and P3 share the derived allele, while P2 shares the ancestral allele with the outgroup. In this paper we will test pooled colonies using *A. m. mellifera* as P1, south west England as P2, and a c-lineage colony in the P3 position

In a scenario where there is no gene flow between the tested populations, ABBA and BABA patterns are expected to occur in roughly equal proportions due to ILS (Patterson *et al.,* 2012; Durand *et al.,* 2011; Green *et al* 2010; Reich *et al.,* 2010). ABBA and BABA patterns are calculated using allele frequencies at fixed sites in the outgroup:

$$ABBA = (1 - P1) \times P2 \times P3 \times 1 - Po$$

$$BABA = P1 \times (1 - P2) \times P3 \times 1 - Po$$

Paterson's D is calculated using the sum of ABBA and BABA patterns across all SNPs:

$$D = \frac{\sum(ABBA) - \sum(BABA)}{\sum(ABBA) + \sum(BABA)}$$

When D deviates from zero it can be indicative of introgression between populations. An excess of ABBA patterns (introgression between P2 and P3) and would result in a D value > 1 and an excess of BABA sites (introgression between P1 and P3) will result in a D value <1. In this study we are looking for significant positive D values, indicating introgression between P2, the south west, to P3, a C lineage honey bee (Figure 3.4). To test whether the D statistics significantly varies from zero, Z-scores and P-values are calculated (Reich *et al.,* 2009; Durand *et al.,* 2011; Green *et al.,* 2010). Z-scores are generated using block jack-knifing, which accounts for the non-independence of linked sites. During block jack-knifing data are divided into blocks of a particular genomic distance or number of SNPs, and the D-statistic is calculated for each of these blocks. Then, the overall D is compared to the standard error of D resulting from the blocks, and a Z-score is calculated. Importantly, RADseq data can contain linked groups of SNPs and this can confound the standard error of jackknife block. Additionally, implementing Z-scores assumes that the data is normally distributed, but often, D statistics resulting from jackknife blocks may not be. To obtain an approximately normally distributed standard error the variation of D over the blocks is calculated, multiplied by the number of blocks and the square root of that number is taken (Reich *et al*., 2009; Durand *et al.,* 2011; Green at al 2010). From this a Z-score is calculated:

$$Z\ score = \frac{D}{normally\ distributed\ \ Standard\ Error\ of\ D}$$

and then a p-value to estimate significance:

$$p\ value = 2 \times Log\ of\ the\ cumlative\ distibution\ function\ (-Z\ score)$$

The related *f* statistic estimates the overall proportion of admixture by comparing the excess of ABBA over BABA patterns to a scenario of complete admixture. To examine complete admixture P3 is split into two groups, P3a and P3b, and P2 is replaced with one of these P3 populations and used in the calculation (Figure 3.6).

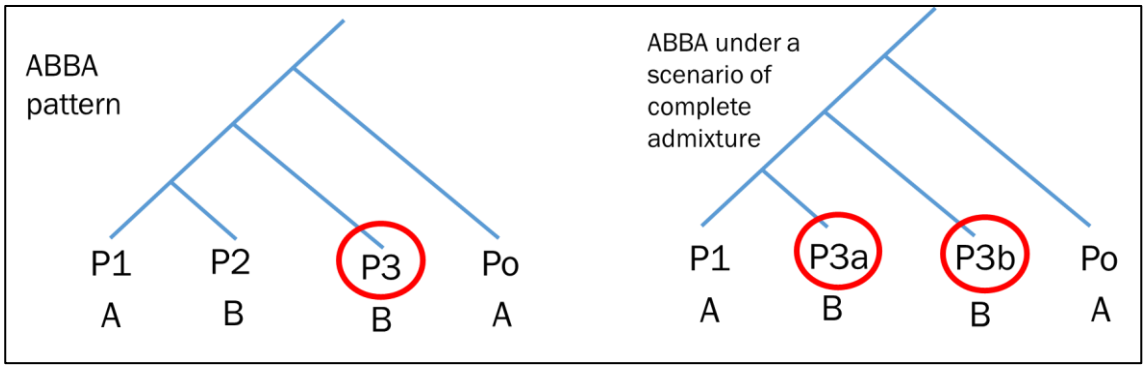


Figure 3.6 The f statistic compares a complete admixture scenario. Here to simulate complete admixture the P2 population has been replaced with another P3 population, labelled P3a.

The *f* statistic is calculated as:

$$f = \frac{\sum(ABBA) - \sum(BABA)}{\sum(ABBA) + \sum(BABA)}$$

Where:

$$ABBA_{numerator} = (1 - P1) \times P2 \times P3a$$

$$BABA_{numerator} = P1 \times (1 - P2) \times P3a$$

$$ABBA_{Denominator} = (1 - P1) \times P3b \times P3a$$

$$BABA_{Denominator} = P1 \times (1 - P3b) \times P3a$$

The pooled colony RADseq ABBA BABA calculations were performed using a custom python code. To calculate the D statistic and f statistic in the pooled data, colony level allele frequencies were calculated at each SNP using the AD and DP fields from the info column in the vcf file, this is the same method employed for pooled data by poolfstat (Gautier *et al.,* 2022). The individual worker RADseq calculations were performed in the software package Dsuite (Malinsky *et al.,* 2021). Dsuite performs a

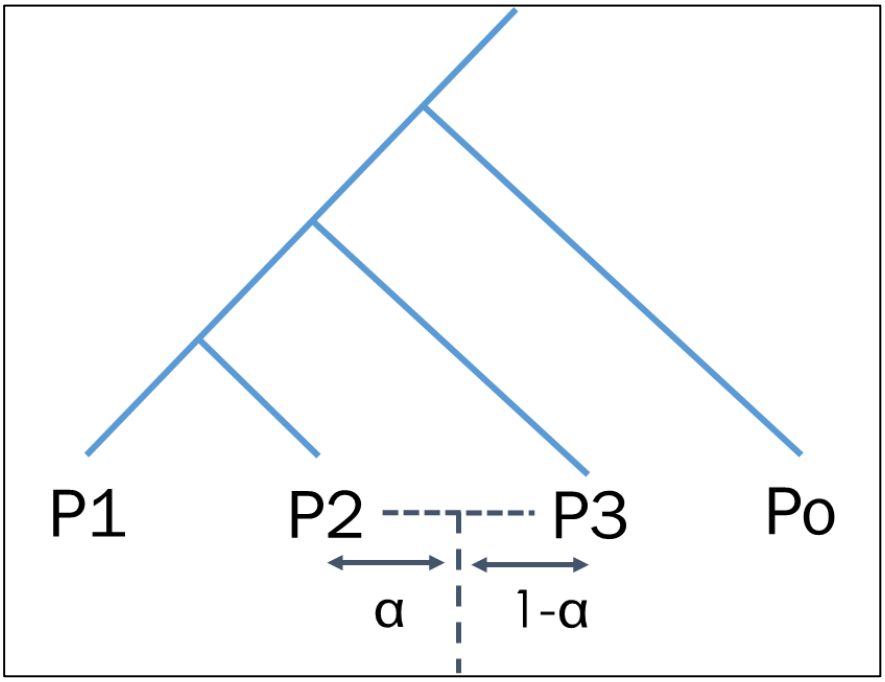


Figure 3.7. The f4-ratio calculated by Dsuite. The introgression from P2 to P3 is represented by α and the introgression from P3 to P2 is represented by 1-α.

related f statistic, the f4-ratio (Patterson *et al.,* 2012). The f4-ratio, just like f, estimates the proportion of admixture between P2 and P3 but here the result is a ratio where α represents the admixture from P2 to P3 and 1- α the admixture from P3 to P2 (Figure 3.7).

The admixture ratio (α) between P2 and P3 is calculated by again splitting P3 into two groups, P3a and P3b and replacing P2 with P3b to compare observed ABBA BABA

patterns to patterns of complete admixture. Specifically, if P3 were to be split into P3a and P3b, P3a and P3b represent the same subspecies and the admixture would be a proportion of 1.0, total admixture. Dsuite (Malinsky *et al.,* 2021) calculates the f4 ratio as:

$$f4ratio = \frac{\sum(P3a - 1) \times \sum(P2 - P1)}{\sum(P3a - 1) \times \sum(P3b - P1)}$$

Dsuite splits P3 by randomly sampling alleles from P3 at each SNP. Dsuite also calculates normalised Z scores and P-values.

3.4 Results

As the aim here is a comparison of methods the proceeding discussion will concentrate on samples that were in across methods, however for full results see appendix 1.

3.4.1 ADMIXTURE

The SNP array data consisted of 80 SNPs from the original 117SNPs, across a total of 197 samples with a genotyping rate of 0.98. ADMIXTURE analysis, identified the most likely K value as 2 (CV values, K=1: 0.97, K=2: 0.59, K=3: 0.60, K=4: 0.62, K=5: 0.64). This is unsurprising as the ancestry informative SNPs chosen for the SNP array are designed to distinguish between the two *Apis mellifera* lineages, C and M. To examine the accuracy of the SNP array Q values, the *A. m. m.* standards that accompany the SNP array were compared back to the original 117 SNP panel Q values in Henriques *et al*. (2018). There was a very strong correlation between the K=2 Q value results and the Q values presented in the 117 SNP results in Henriques *et al* (2018) (R^2=0.9879, Supplementary table 9).

The individual RADseq data consisted of 23,916 SNPs across 61 individuals. ADMIXTURE identified the lowest CV value at K=2 for the data identified it as the most likely K (CV values, K=1: 0.71, K=2: 0.49, K=3: 0.50, K=4: 0.56; K=5: 0.57; K=6: 0.69; K=7: 0.67, K=8: 0.76, K=9: 0.85, K=10: 0.86).

The pooled colony RADseq data consisted of 158,496 SNPS and 103 samples. ADMIXTURE analysis showed K=3 was the most likely number of clusters (CV values reposted, K=1: 0.34, K=2: 0.26, K=3: 0.26, K=4: 0.27, K=5: 0.27, K=6: 0.28, K=7: 0.29, K=8: 0.30, K=9: 0.31, K=10: 0.32). These three clusters broadly represent the English South West samples, the *A. m. m* samples and the C-lineage bees, *A. m. c* and *A. m. l*. The South West of England is unlikely to harbour its own subspecies and the K=3 result could be a result of a unique signature of admixture. It is important to use biological knowledge of systems when interpreting ADMIXTURE results (Lawson, Van Dorp and Falush, 2018) and in order to compare the pooled results to the individual RADseq and SNP Array results, here K=2 in the pooled RADseq is examined.

All South West samples in all methods showed some degree of introgression (Figure 3.8). There were 17 samples common to all three methods (SNP, individual RADseq and pooled RADseq, Table 3.5, and Figure 3.8). Across these 17 samples ancestry membership Q values for the M lineage range from 0.89 to 0.18 (Table 3.5 and Figure 3.8 D) and C lineage from 0.82 to 0.11 (Table 3.5). Overall, the individual RADseq results and the SNP array results were highly correlated (Figure 3.8 D), although pooled colony RADseq produced overall lower values than the SNP array (Figure 3.8 D, Table 3.5). Sample c22 obtained the highest C lineage Q value in all three methods (Table 3.5 and Figure 3.9) and sample c7 had the highest M lineage assignment in all three methods (Table 3.5), though in the pooled colony RADseq sample c25 was also assigned the same M lineage value (Table 3.5 and Figure 3.9). There were four

samples where the individual RADseq and SNP array Q values yielded the same results (c8, c17, c23, c21, Table 3.5 and Figure 3.9). Average Q values for each ancestry cluster estimated across the 17 samples were the same for both the individual RADseq and the SNP array (M lineage 0.70 and C lineage 0.30, Table 3.5). The pooled colony RADseq generated a higher average C-lineage Q value assignment compared to the individual methods (0.38) and lower Q value assignment to the M lineage cluster (0.62) (Table 3.5). The largest sample Q value difference across methods was between pooled colony RADseq and the SNP array in samples c11 and c13 (Table 3.5 and Figure 3.9). Sample c11 and c13 both had a C-lineage Q value difference of 0.21 between the SNP array and the pooled colony RADseq.

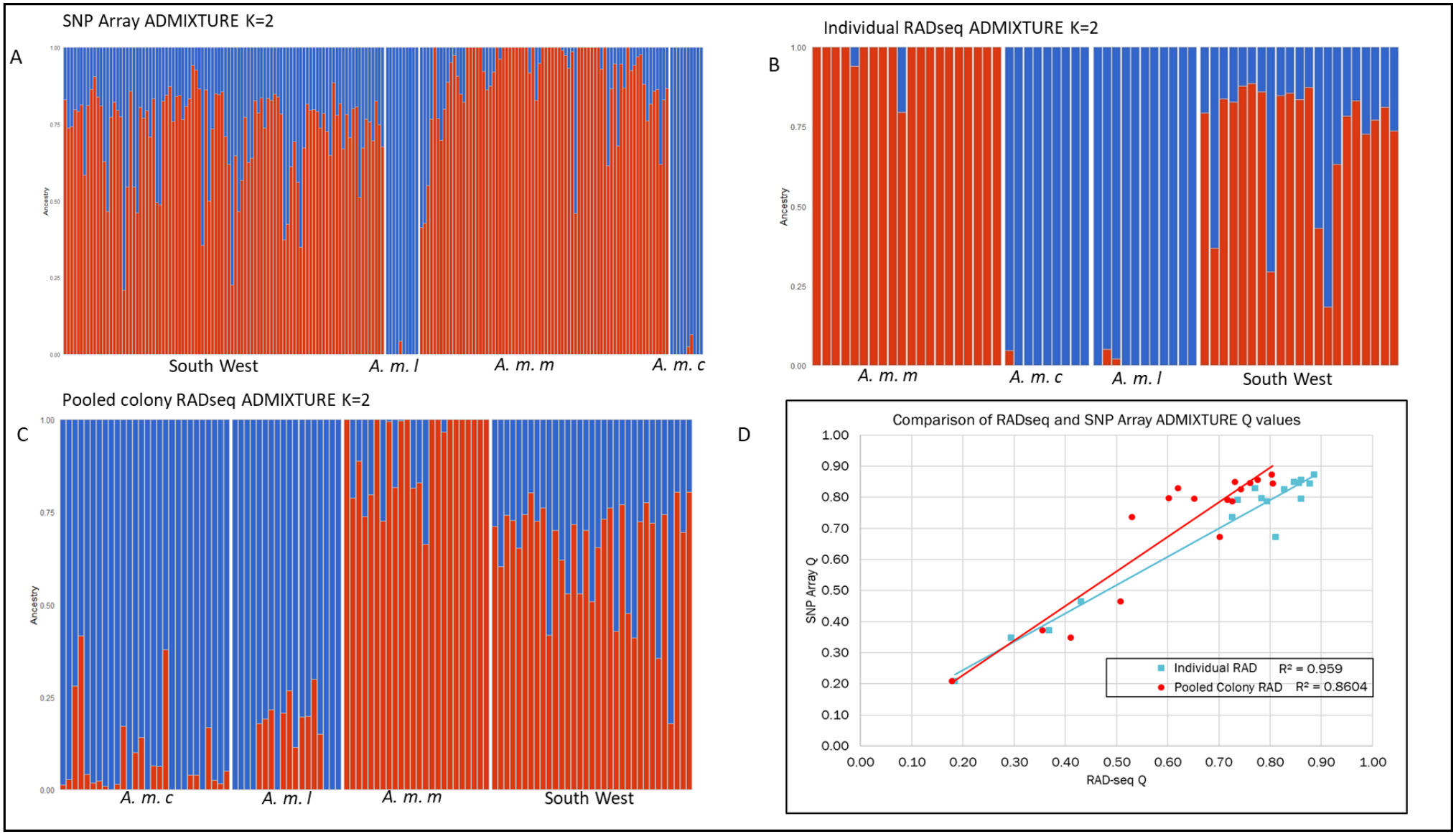


Figure 3.8. ADMIXTURE analysis of K=2 of honey bees in the south west of England (A) SNP Array,(B) Individual RADseq and (C) Pooled colony RADseq. Blue cluster represents the C-lineage honey bees, *A. m. carnica* and *A. m. ligustica*, the red cluster represents *A. m. mellifera*. (D) Comparison of RADseq Q values to SNP Array Q values for membership to *A. m. mellifera*

Table 3.5. South west samples Q values of membership to ancestral populations from ADMIXTURE at K=2 from SNP array, Individual RADseq and Pooled colony RADseq

Colony ID	ADMIXTURE K=2					
	SNP array		Individual RADseq		Colony pooled RADseq	
	M	C	M	C	M	C
c2	0.80	0.20	0.78	0.22	0.60	0.40
c5	0.79	0.21	0.86	0.14	0.65	0.35
c6	0.82	0.18	0.83	0.17	0.74	0.26
c7	0.87	0.13	0.89	0.11	0.80	0.20
c8	0.79	0.21	0.79	0.21	0.73	0.27
c12	0.67	0.33	0.81	0.19	0.70	0.30
c11	0.83	0.17	0.77	0.23	0.62	0.38
c13	0.74	0.26	0.73	0.27	0.53	0.47
c14	0.79	0.21	0.74	0.26	0.72	0.28
c16	0.47	0.53	0.43	0.57	0.51	0.49
c17	0.85	0.15	0.85	0.15	0.73	0.27
c18	0.85	0.15	0.86	0.14	0.76	0.24
c10	0.35	0.65	0.29	0.71	0.41	0.59
c23	0.86	0.14	0.86	0.14	0.78	0.22
c21	0.37	0.63	0.37	0.63	0.36	0.64
c22	0.21	0.79	0.18	0.82	0.18	0.82
c25	0.84	0.16	0.88	0.12	0.80	0.20
Average	0.70	0.30	0.70	0.30	0.62	0.38

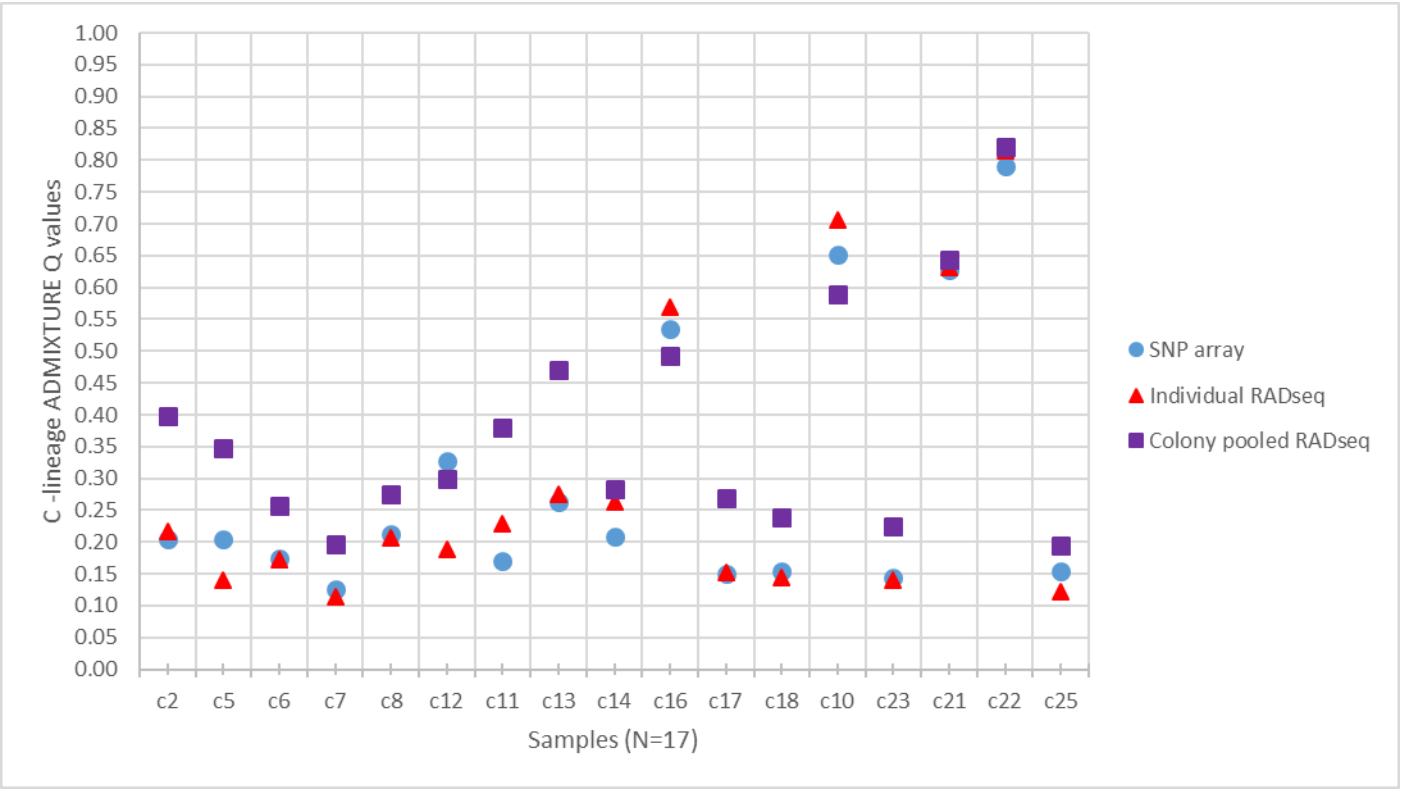


Figure 3.9. C-lineage proportion estimated by ADMIXTURE at K=2 using the SNP array (80SNPs), individual RADseq (23,916 SNPs), and pooled colony RADseq (158,496 SNPs) across the 17 samples examined in all three methods

3.4.2 ABBA BABA

The ABBA BABA calculations for the pooled colony RADseq were performed on 10,505 SNPs (sites at which the *A. cerana* outgroup is fixed). Each colony was tested 72 times (36 times each for *A. m. ligustica* and *A. m. carnica* introgression), against different combinations of standard colonies. In the trios Mel;SW;Car and Mel;SW;Lig, a total of 19 out of 28 colonies were significant for positive D values (p-values < 0.05, Z-scores >3) across all combinations indicating introgression from either *A. m. carncia* or *A. m. ligustica*. This indicates that these samples were significant regardless of the colonies chosen to represent the standards in the P1 (*A. m. mellifera*) and P3 positions (either *A. m. carnica* or *A. m. ligustica*). The highest proportion of admixture estimated in both the Mel;SW;Car and the Mel;SW;Lig trios was for sample c22, with average admixture proportion of 0.682 and 0.626 respectively (Figure 3.10). The lowest proportion of admixture observed in sample c25 with proportions of 0.06 (Mel;SW;Car) and 0.07 (Mel;SW;Lig) (Figure 3.10). The standards deviations of the f statistic were larger in samples that with higher introgression values (Figure 3.10). The admixture proportions from the two different trios were similar. The largest within sample difference in admixture proportion was in the c22 sample (a difference of 0.07). Of the 9 colonies that were not significant in all combinations, 4 were not significant for both *A. m. ligutsica* and *A. m. carnica*, while 5 were not significant only for *A. m. carnica* introgression. All colonies that were not significant for introgression had admixture proportions below 0.1.

Dsuite compared 4832 SNPs in the individual RADseq data set. D values deviated significantly from zero (p-values< 0.05) on all three trios tested (Table 3.6). The highest Z-score and smallest p-value was seen in the Car; south west; Mel trio. *A. m. mellifera* proportions (1-α) in the individual RADseq data range from 0.66 to 0.759 across the three trios.

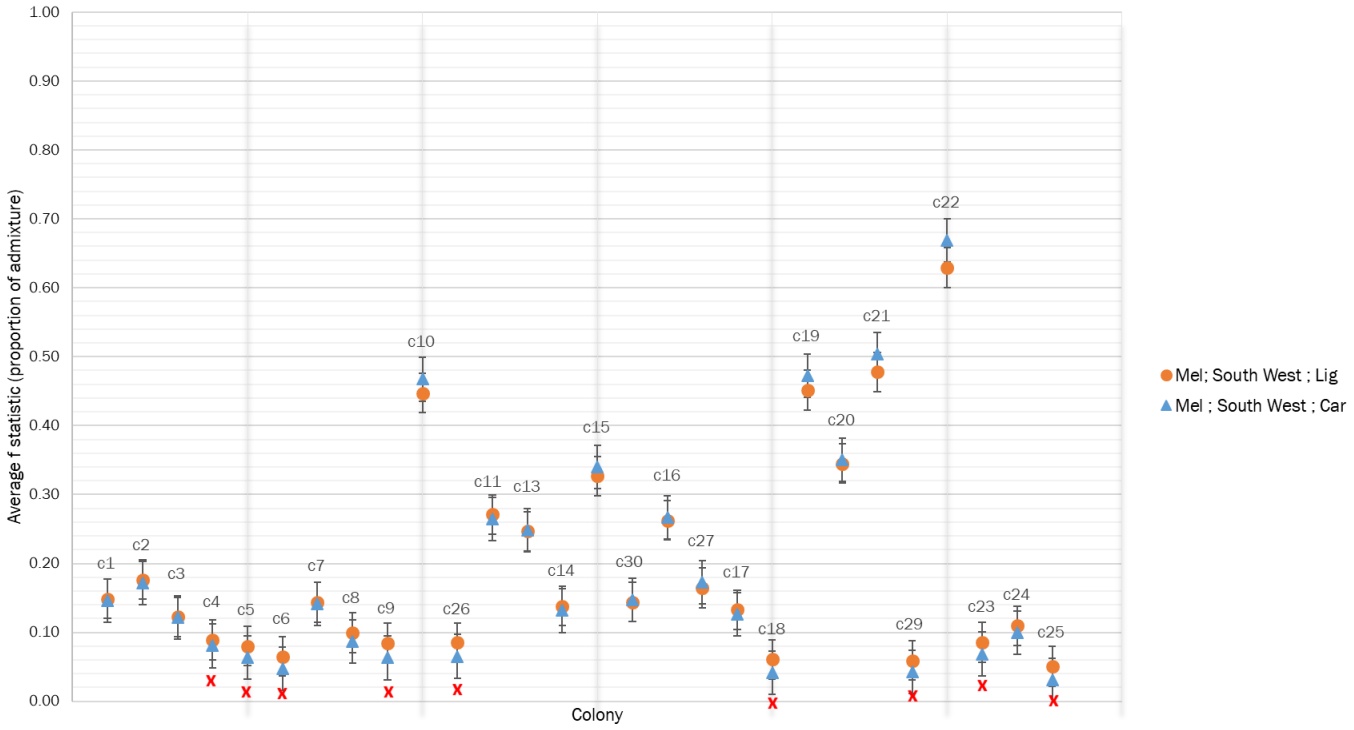


Figure 3.10. Average *f* statistics (proportion of admixture) calculated for pooled colonies. Trios tested for introgression between the south west and two different C-lineage subspecies. Shown here are trios consisting of *A. m. mellifera* ; south west; *A. m. carnica* (blue triangle), and *A. m. mellifera* ; *south west*; *A. m. ligustica (*Orange circle*)*. A red x donates colonies that had a non-significant result. Error bars represent the standard deviation for each south west colonies estimates calculated with every combination of *A. m. mellifera* and *A. m. carnica* or *A. m. ligustica* colony

Table 3.6. Dsuite results for the individual RADseq data examined on a population level. Populations are represented as *A. m. mellifera* (Mel), *A. m. carnica* (Car), *A. m. ligustica* (Lig), south west and C-linage (*A. m. ligustica* and *A. m. carnica* combined in to one population

Trios	D statistic	Z-score	p-value	f4-ratio (α)	1-α
C-lineage; south west; Mel	0.105	3.47167	0.000517	0.292	0.708
Lig; south west; Mel	0.0816	2.63019	0.008534	0.241	0.759
Car; south west; Mel	0.1306	4.374	0.0000122	0.340	0.66

.

3.5 Discussion

This study compared the introgression results from three methods: the C lineage introgession SNP array (Henriques *et al.,* 2018), individual RADseq and pooled colony RADseq, using two statistical approaches. Both ADMIXTURE and the ABBA BABA approach revealed introgression in the *Apis mellifera mellifera* population in South West England. In ADMIXTURE, the pooled colony samples exhibited, on average, higher introgression values than individual RADseq and SNP array results while the individual RADseq and SNP array produced similar Q values. This consistent ADMIXTURE result between the individual genotyping approaches was expected as the SNP array has been rigorously tested to produce results similar to that of whole genome data and sequencing only a single individual restricts the number of patrilines sampled per colony. The ABBA BABA analysis resulted in lower estimates of introgression for the pooled colonies than the colony ADMIXTURE analysis. The overall introgression level for the individual RADseq data estimated using the ABBA BABA method (Table 3.7, proportions of 0.7 *A. m. mellifera* and 0.29 C lineage) was the same as the average ADMIXTURE Q value (Table 3.5, average Q value membership of 0.7 to A. m. mellifera and 0.30 to C-lineage).

Some colonies produced disparate values when examined by the ADMIXTURE and ABBA BABA approach, for example, for colony c26 ADMIXTURE estimated an introgression Q value of 0.32 (SNP array) and 0.29 (pooled colony RAD) while the ABBA BABA approach estimated 0.085 for *A. m. ligustica* introgression and was non-significant for *A. m. carnica* introgression. A similar pattern was seen in colony c9 (ADMIXTURE SNP array Q value 0.22, colony RADseq Q value 0.24, ABBA BABA proportion of 0.085, and non significant for *A. m. carnica* introgression). In both the ABBA BABA and the ADMIXTURE analyses colony C22 was observed to have the

highest introgression level (colony ADMIXTURE Q value of 0.82, f statistic results from *A. m. ligustica* of 0.62 and *A. m. carnica* of 0.70).

ADMIXTURE is designed for unrelated individuals rather than pooled DNA of related individuals as it has been used here, although it has been used on pooled data in other studies (Gmel *et al.,* preprint; Regan *et al.,* 2018; Henriques *et al.,* 2018). Additionally, not all pooled colony samples yielded higher introgression values: in three out of the 17 samples (sample c12, c16 and c10) pooled colony Q values were lower than an individual analysis results. This suggests that results depend on the overall introgression level of the colony and the introgression level of the individual chosen for the analysis rather than the higher introgression being a result of pooled data being placed into ADMIXTURE program. This result is at odds with Henriques *et al* (2018) who found that using the SNP Array on pools resulted in lower ADMIXTURE values than individual samples. However, the Henriques *et al* (2018) result was from the MassARRAY platform using tissue pools and the approach used in this study was different. Firstly, the pooling method was equimolarly pooled DNA (rather than using tissue pools) and the sequencing platform and data processing were different. Regan *et al.,* (2018) compared pool-seq ADMIXTURE results to data set simulated for individuals and found that up to K=3 the results were consistent with those compared to the pooled genotypes, while at K=4 two samples two samples showed an assignment difference and at K=5 one additional sample showed a Q value difference. When ADMIXTURE is used for introgression studies the number and quality of standards used will have a potentially significant effect on the results. The number of standards used here in the RAD-seq individual analysis was 40, while in the pooled analysis 69 standards were used (as many standards that were available were used). This is important as ADMIXTURE uses all the samples input to calculate the clusters and there are no guidelines to the number of standard samples that one

should place into ADMIXTURE to calibrate the analysis. Nevertheless, the SNP array and the individual RAD-seq still resulted in largely similar values even though they used different standards and a different number of samples was placed into the programme.

Clearly, the two statistical approaches are different: ADMIXTURE is a population structure clustering programme while ABBBA BABA tests are based on allele frequency patterns. There have been concerns about the use and over-interpretation of population clustering program coefficients for estimation of introgression values (Anderson and Dunham 2008; Lawson *et al.,* 2018; Kong *et al.,* 2021). These methods are sensitive to the choice of marker, the level of genetic differentiation and the amount of data used in the analysis (Vähä and Primmer 2005; Kalinowski 2011). Additionally, population clustering programmes do not perform hypothesis testing. Although the D statistic has been seen to outperform both ADMIXTURE and STRUCTURE for identifying hybrids (Kong *et al.,* 2021), ABBA BABA approaches can also result in higher error rates in situations with high ILS (Kong *et al.,* 2021).

Overall, the differences in the results from individual and pooled approaches was expected. Pooled sequencing captures more genotypes from the colony and there is an inherent bias in using an individual approach for this reason (i.e. an individual is not representative of a colony). But different statistical approaches also resulted in different introgression estimates. In previous studies an arbitrary cut off for classifying an *A. m. mellifera* individual as introgressed has been used, for example ADMIXTURE Q value >0.05 (Groeneveld *et al.,* 2020), >0.1 (Parejo *et al.,* 2018; Hassett *et al.,* 2018) and discussions about what cut-offs are appropriate or what their biological significance has been begun in this thesis (see Chapter 6) and elsewhere (Ellis *et al.,* 2018). But, if ADMIXTURE results indicate introgression estimates of >0.2 while the ABBA BABA approach classes that colony as not being

introgressed at all, this raises questions about the field largely relying on one approach (clustering programmes) and suggests that sampling one individual my not be a robust enough approach to assess introgression in honey bee populations.

4. Assessing the status of *Apis mellifera mellifera* across the UK and Ireland

4.1 Introduction

A. m. mellifera is the native subspecies in the UK and Ireland. However, the picture for *A. m. mellifera's* distribution across the UK and Ireland remains largely unknown and a more comprehensive investigation of the UK and Ireland is overdue. In Europe there has been evidence of low introgression and 'pure' *A. m. mellifera* in Norway, Sweden, and Finland (Groeneveld *et al.,* 2020) as well as the Netherlands and Switzerland (Soland-Reckeweg *et al.,* 2009; Pinto *et al.,* 2014). Studies of the UK and Ireland have discovered remnant populations of *A. m. mellifera* in Ireland (Hasset *et al.,* 2018; Browne *et al.,* 2020), the Inner Hebrides (Jensen *et al.,* 2004) and the South West England (Ellis *et al.,* 2018) but to-date no wider survey has been carried out.

Monitoring introgression is important as introgression can lead to adaptation or conversely genomic extinction (Allendorf and Luikart., 2009; Epifanio and Philipp, 2000) (see chapter 1 for full discussion). Genomic extinction is a concern in conservation, especially when continual imports introgress with native species. This is the case with *Apis mellifera mellifera*. So far, this thesis has assessed *A. m. mellifera* introgression in a conservation program and examined approaches for measuring introgression at a colony level. But here while assessing introgression the

primary aim is to investigate the UK and Ireland for pockets of remnant *A. m. mellifera*.

The majority of *A. m. mellifera* populations discovered across Europe have been located in small areas usually built up by conservation bee keeping groups (Pinto *et al.,* 2014; Henriques *et al.,* 2018). However, studies in Ireland have revealed that *A. m. mellifera* exists across a wide geographical area in the general population. Hassett *et al.,* (2018) sampled 412 honey bees from 80 sites across Ireland and using both mtDNA and microsatellite markers qualified over 400 samples as *A. m. mellifera* with a 'purity' threshold of >0.95. Browne *et al.,* (2020) examined 76 free-living honey bee colonies across Ireland and found that only 2 samples were below 0.9 purity, concluding that the free living population consisted of mostly of *A. m. mellifera* honey bees. Ireland is therefore an important reservoir of *A. m. mellifera* diversity.

The Inner Hebridean Isle of Colonsay is another stronghold of *A. m. mellifera* in the UK. This population is protected by a Scottish government order and the island is largely under the control of one bee keeper (Andrew Abrahams, personal communication). This population has been included in many wider European studies of *A. m. mellifera* (Jensen *et al.,* 2005; Munoz *et al.,* 2015, Regan *et al.,* 2018; Pinto *et al.,* 2014) and has been observed as consisting of 'pure' *A. m. mellifera*, usually exceeding a threshold > 0.99 'purity' when examined in population structure software (for example, ADMIXTURE or STRUCTURE) (Alexander *et al.,* 2015; Pritchard *et al.,* 2010).

Ellis *et al.* (2018) investigated honey bee lineages in South West England. The study sampled colonies belonging to bee keepers that were part of the B4 conservation group and were suspected of being *A. m. mellifera.* This study found that of 30 colonies sequenced, 4 to 15 were classified as *A. m. mellifera* depending on the purity thresholds chosen.

The question of which thresholds are appropriate for assigning a colony as 'pure' is still unanswered. Answering this requires understanding of how introgression affects an organism. When does introgression result in an interruption of those locally adapted allele combinations? What are the consequence of introgression on *A. m. mellifera* traits? In order to investigate what effect introgression has on particular loci studies turn to investigating patterns of introgression.

Studies of the patterns of introgression in model organisms have observed that introgression signals are weaker in areas of the genome that have low recombination rates or loci with a high density of genes (Sankararaman *et al.,* 2014; Janoušek *et al.,* 2015; Martin and Jiggins 2017). The comparatively high recombination rate of *Apis mellifera* may explain why recombination 'hotspots' (where recombination events cluster in highly localized areas a few kb in length they are referred to as hotspots, (Paigen and Petkov *et al.,* 2010). and introgression patterns have been difficult to detect. Hotspots have been detected in many organisms e.g. mammals, birds, plants, fungi (Singhal *et al.,* 2015; Croll *et al.,* 2015; Latrille *et al.,* 2017) but a lack of hotspots have been detected in honey bees (Wallberg *et al.,* 2015; Stapley *et al.,* 2017). Honey bees have a recombination rate between 19 - 37 cM/Mb (centimorgans per million base pairs) (Beye *et al.,* 2006; Solignac *et al.,* 2007; Liu *et al.,* 2015). This rate of recombination translates to approximately 5 crossover events per chromosome during meiosis while the average from a wide range of taxa is about 1.6 per chromosome (Baker *et al.,* 1976; Beye *et al.,* 2006). In comparison, fungi species have been observed to have extreme crossover rates, estimates of 40 cM/Mb or even greater than 300 cM/Mb (Stukenbrock *et al.,* 2018; Haenel *et al.,* 2018), while mammals have an average recombination rate of 1cM/Mb, which is equivalent to one crossover per chromosome (Dumont *et al.,* 2009). Amongst insects, the highest recombination rates are seen in the Hymenoptera 7.12±5.13 cM/Mb,

followed by Lepidoptera (average recombination rate of 4.47 cM/Mb), Coleoptera (average recombination rate of 2.48 cM/Mb) and Diptera species with the average rate of 1.03 cM/Mb (Wilfert *et al.,* 2007). The highest rates are seen in social Hymenoptera. The bumblebee *Bombus terrestris,* has a cross over rate of 4.4 - 8.7cM/Mb (approximately 1.1 crossover events per chromosome) (Wilfert *et al.,* 2006; Liu *et al.,* 2017). *Bombus terrestis* while social, exists in single mated colonies and lives in annual colonies of a smaller size compared to *Apis mellifera* (Baer and Schmid-Hempel 1999). In general this higher genomic crossover rate is associated with highly social insects. The harvester ant (*Pogonomyrmex rugosus*) has a crossover rate of 14 cM/Mb while the leaf cutter ant (*Acromyrmex echinatior*) has a recombination rate of 6.2 cM/Mb (Sirvio *et al.,* 2006). The effect of this high recombination rate on introgression patterns has been studied in honey bees. Wragg *et al.,* (2018) investigated patterns of admixture across the genome of honey bees in the Reunion Isle. The Reunion Isles native subspecies is *Apis mellifera unicolor,* a member of the A lineage, and the island has seen imports of popular C lineage subspecies. One of the aims of their study was to, examine whether there was any preferential selection of one lineage or another. Examining the ancestry of haplotype blocks along the genome they found that there were a large number of small blocks belonging to either A or C lineage backgrounds. This finding is consistent with repeated recombination over time and the high recombination rates observed in *Apis mellifera*. However, they also observed 15 regions significantly associated with the native *A. m. unicolor* and preferential selection of these regions.

This study investigated putative *A. m. mellifera* across the UK and Ireland using pooled colony whole genome data and examined the pattern of introgression across the genome, to identify whether consistent regions of introgression could be observed.

4.2 Methods

4.2.1 Sample collection

To collect samples from across the UK and Ireland bee keeping groups that specialise in the promotion and care of *A. m. mellifera* or local honey bee breeds were contacted. Societies contacted were the Scottish Native Honey Bee Society, the Native Irish Honey Bee Society, and the Bee Improvement and Bee Breeders Association. The study was also promoted via social media (Twitter and Facebook). For more detailed information on sampled locations see Appendix 2.

Bee keepers were asked about the importation status of their honey bees and if they thought they had *A. m. mellifera* or honey bees similar to *A. m. mellifera*. A total of 113 bee keepers responded to the study's outreach and 81 bee keepers with colonies that met the criteria of no importation chose to continue to participate.

Sample tubes containing 70% ethanol were sent to bee keepers and the bee keepers were asked to collect 40 worker bees per colony.

For comparison with the UK and Ireland data subspecies standards of *A. m. carnica* and *A. m. ligustica,* and *A .m mellifera* were included. *A. m. ligustica* (total n=15) was sampled from Italy (n=5) and Sweden (n=10), *A. m. carnica* (total n=26) from Germany (n=7), Sweden (n=9) and Norway (n=10), while *A. m. mellifera* (total n=28) was sampled in Sweden (n=11), Norway (n=10) and Switzerland (n=7). These samples were obtained from breeding programs (Switzerland and Germany) and experimental apiaries (Sweden, Norway and Italy).

4.2.2 DNA extraction and pooling

Each colony was represented by a pool of 30 worker bees (40 bees were requested from bee keepers for in order to act as spares as some bees did not yield sufficient tissue or DNA).

These methods are the same as the DNA extraction and pooling methods in Chapter 3 section 3.2.3.2. A brief description follows:

For each colony every thorax was weighed and donated an equal amount of tissue, and then an equal amount of DNA to the final pool. Five thoraxes were processed in each extraction. The extractions were performed using an ammonium acetate protocol (Nicholls *et al.,* 2000). Extractions resulted in an average yield of 2110 ng of DNA per extraction.

This created a final pool of 30 workers representing each colony consisting of 1000ng of purified DNA.

Any samples that were supplied without a sufficient number of individuals or did not yield enough DNA to contribute to the final pool were removed from the experiment. A total of 74 samples were processed along with the subspecies standards.

4.2.3 Library preparation and sequencing

All library preparation and whole genome sequencing (WGS) was performed by the Beijing Genomic Institute (BGI) in Hong Kong and sequenced on a BGISEQ-500 producing 100bp paired end reads (Xu *et al.,* 2019).

4.2.4 Bioinformatics

The bioinformatics pipeline was the same the pooled data in chapter 3 sections 3.2.4.2. and 3.2.6.3.

Raw data received had an average of 136 million reads per pooled colony sample with an average of 96% of reads exceeding a phred score of 20. Raw data were quality trimmed and filtered based on quality (Trimmomatic; Bolger *et al.,* 2014). Reads were aligned to the honey bee reference genome (Amel_HAv3.1) and filtered for alignment quality, and unpaired reads were discarded (BWA MEM aligner and Samtools; Li 2013; Li *et al.,* 2009). Then the GATK best practices pipeline was implemented (GATK version 4.1.9.0.; Van der Auwera and O'Connor, 2020). Final filtering for depth and data representation across samples was performed in vcftools (Danecek *et al.,* 2011).

4.2.5 Investigation of population structure and introgression via ADMIXTURE

ADMIXTURE (as described in chapter 3, section 3.3.1) estimates membership proportions (Q) of individuals to clusters (K) that represent ancestral populations (Alexander *et al.,* 2009).

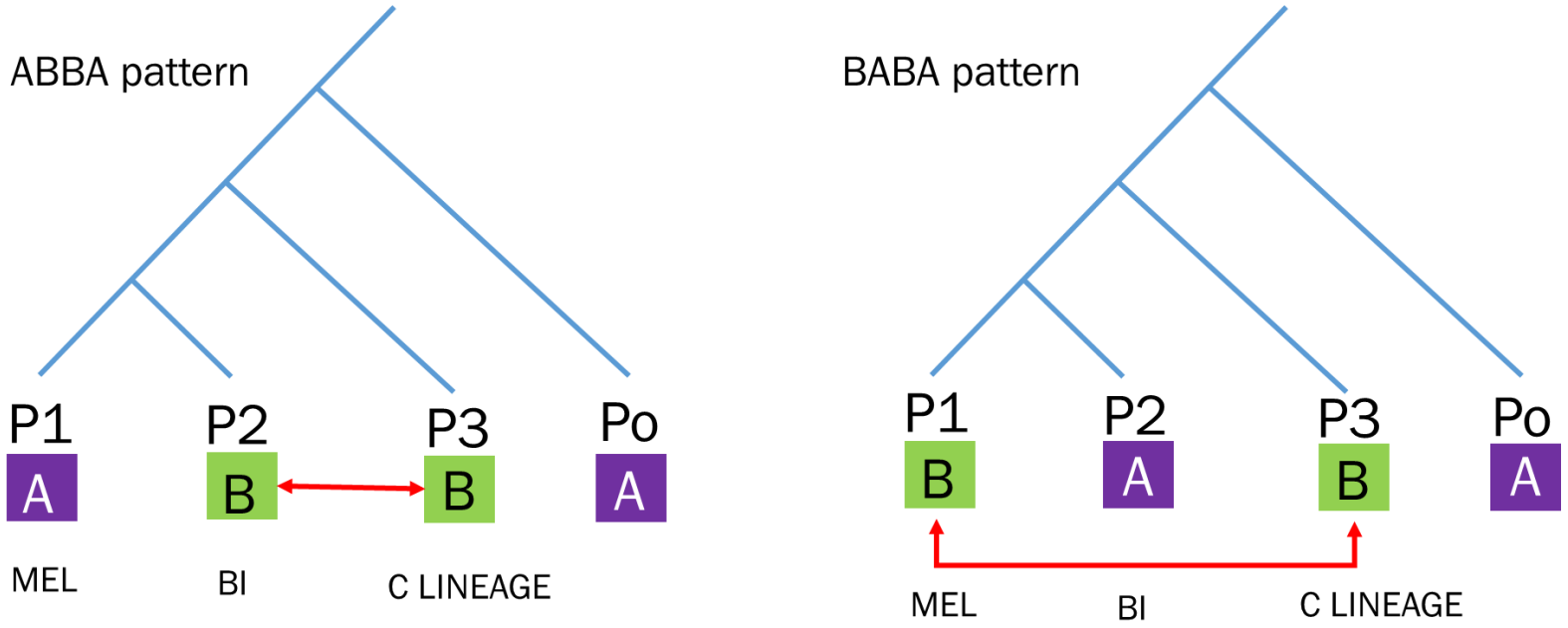
A high density of SNPs were recovered (~9 million). As ADMIXTURE treats sites as independent (i.e. sites not in linkage with one another) it is recommend that data is thinned before it is used in ADMIXTURE (Alexander *et al.,* 2015). Data were arbitrarily thinned to SNPs 1kb apart (Wragg *et al.,* 2018) using the vcftools command–thin. (Danecek *et al.,* 2011).

ADMIXTURE was run for K=1 to K=10 and CV values were viewed for each K. Results were visualised in R using ggplot2, maps and construct (Wickham 2016; Bradburd *et al.,* 2018; R Core Team., 2018). The geographical plotting was performed using approximate locations given by bee keepers to maintain privacy of apiary locations.

4.2.6 Investigation introgression using ABBA BABA statistics.

ABBA BABA statistics (explained in chapter 3, section 3.3.2) investigates introgression using allele frequency patterns. First, the occurrence of introgression was examined for each colony using the D statistic (Figure 4.1), then overall proportion of admixture is estimated using the f statistic (Figure 4.2)(Green *et al* 2010; Reich *et al.,* 2010; Durand *et al.,* 2011; Paterson *et al.,* 2012). As well as overall introgression and admixture proportions, colonies were examined using genomic scans to identify introgressed loci. However, the D statistic is known to perform poorly over smaller genomic regions, which results in overestimated introgression values (values greater than 1), especially when the effective population size is low or regions have low diversity (Martin *et al.,* 2014). To correct for this, a new statistic, $\hat{f}_d$, was proposed (Figure 4.3) for use over smaller regions for detecting introgression between P3 to P2 (Martin *et al.,* 2014). This is related to the f statistic but corrects for inflated values (Figure 4.3). In the $\hat{f}_d$ statistic the denominator is maximised to eliminate values greater than one. When both P2 and P3 have the derived allele frequency, which ever one has the higher frequency is used in the calculation for that site. Specifically, where P1, P2, P3 are the populations being tested and PO is the outgroup. The PD population is dynamic and can represent either P2 or P3 in the denominator calculations depending on which has the higher allele frequency (Figure 4.3) (Martin *et al.,* 2014).

$$D = \frac{\sum(ABBA) - \sum(BABA)}{\sum(ABBA) + \sum(BABA)}$$



$$ABBA = (1 - P1) \times P2 \times P3 \times 1 - Po$$

$$BABA = P1 \times (1 - P2) \times P3 \times 1 - Po$$

Figure 4.1. The calculation of the D statistic for introgression. In our scenario P1 is the *A. m. mellifera* colony (MEL), P2 is the British Isles colony (BI) and P3 is a C lineage colony, which will be represented by either an *A. m. ligustica* or an *A. m. carnica* colony.

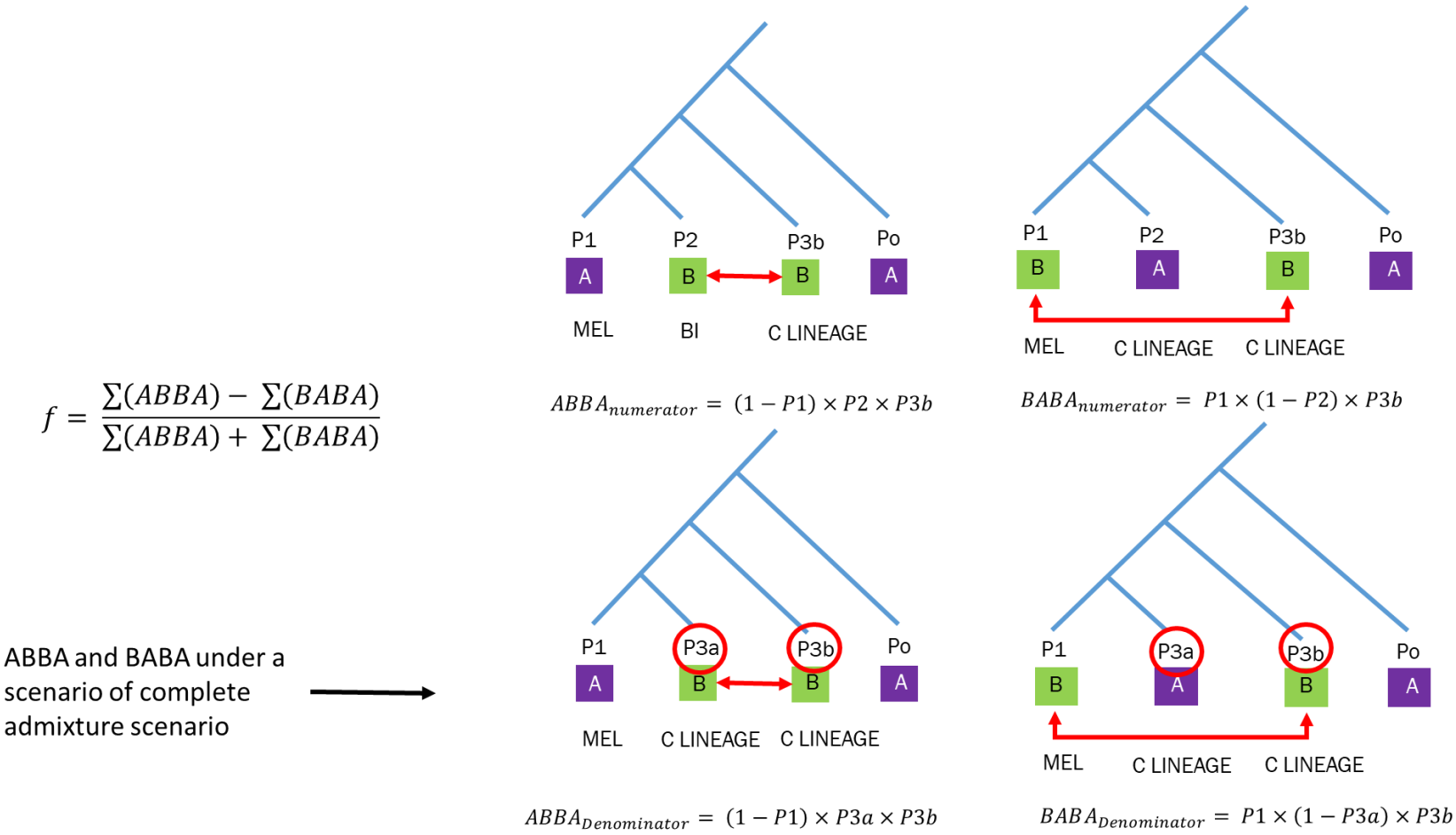
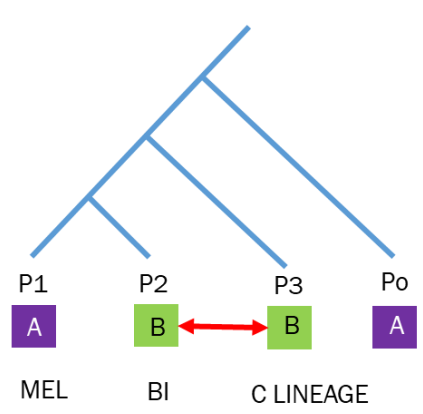
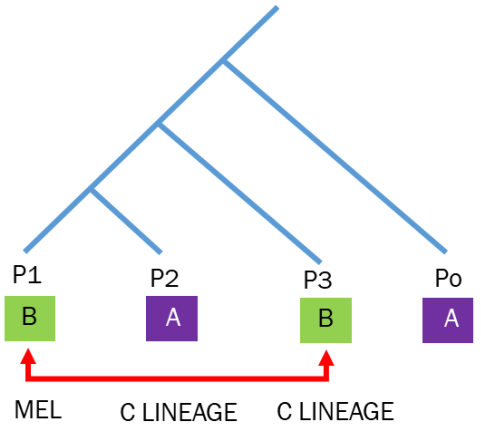


Figure 4.2. The f statistic for calculating the overall admixture proportion for a colony. This is performed by comparing ABBA and BABA patterns between colonies to those patterns under complete admixture. This is done by splitting P3 (in this study the C lineage colony, either *A. m. ligustica* or *A. m. carnica*)

$$\hat{f}_d = \frac{\sum(ABBA - (BABA)}{\sum(ABBA + (BABA)}$$

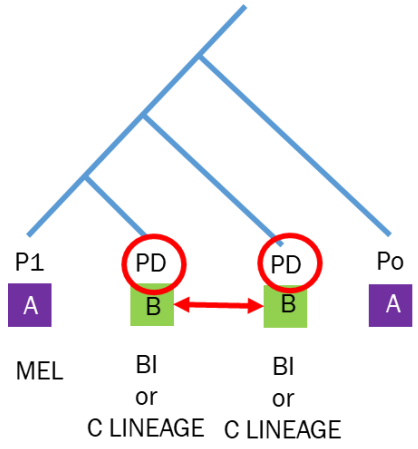


$ABBA_{numerator} = (1 - P1) \times P2 \times P3$

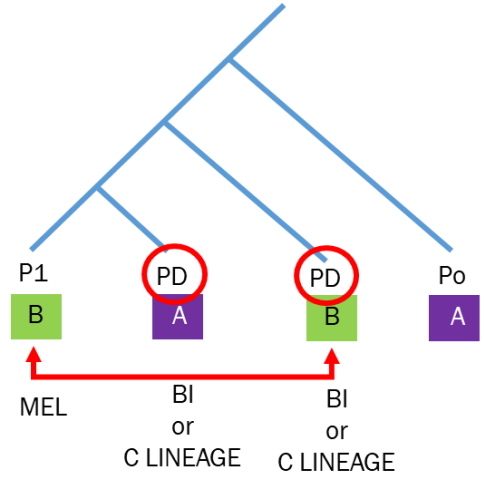


$BABA_{numerator} = P1 \times (1 - P2) \times P3$

PD is dynamic and assigned as either the P2 or P3 population based on which one has the higher allele frequency at that site.



$ABBA_{Denominator} = (1 - P1) \times PD \times PD$



$BABA_{Denominator} = P1 \times (1 - PD) \times PD$

Figure 4.3. The $\hat{f}_d$ statistic is based on the f statistic but used to calculate introgression over smaller genomic regions. Here, instead of P3 being split to calculate the scenario for complete admixture the population with highest allele frequency is used out of P2 or P3.

The $\hat{f}_d$ statistic was applied in overlapping windows across all SNPs. This approach was chosen because windows based on fixed distances of base pairs can contain different quantities of SNPs and the overall pattern of the results can be statistically noisy as windows with fewer SNPs create more variance for all the calculated windows (Malinsky *et al.,* 2021). Genomic scans were performed on window sizes of 100 SNPs.

Windows were considered outliers for a colony if they exceeded the 99th percentile and were detected in a colony regardless of the reference sample used to represent the *A. m. mellifera* (P1 colony) or C lineage (either *A. m. carnica* or *A. m. ligustica,* P3 colony) (Lopes *et al.,* 2016; Barbato *et al.,* 2017; Ravinet *et al.,* 2021). To ascertain whether there were any loci that were consistently introgressed across the UK and Irish samples, the outliers from all colonies were compared.

Alongside this per colony assessment of introgressed loci a whole population approach was performed using Dsuite (Malinsky *et al.,* 2021). In this approach all colonies that are significant for introgression were grouped together and assessed as a population. This approach does not utilise the colony allele frequencies; instead each colonies consensus genotype is used and grouped into a population representing all the introgressed samples from the UK and Ireland.

Colony assessments of ABBA BABA statistics were performed using a custom Python 3 code (Python software foundation available at www.python.org) using Pandas (McKinney, 2010). The population assessments were performed using Dsuite (Malinsky *et al.,* 2021). Before population introgression windows were calculated any colonies that did not show significant introgression were removed.

4.3 Results

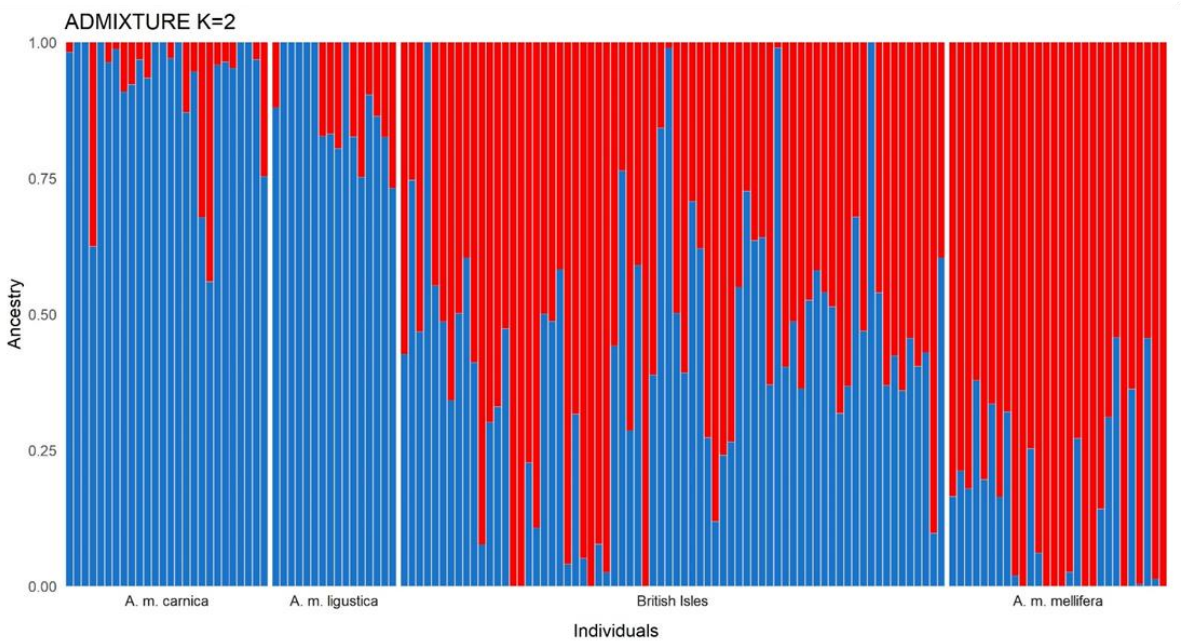
4.3.1 ADMIXTURE

The data set consisted of a total of 140 samples and 208,587 SNPs. Of 74 samples from the UK and Ireland, 70 passed the data filtering process.

The lowest CV value was for where K=3 (CV values, K=1: 0.201, K=2: 0.167, K=3: 0.164, K=4: 0.166, K=5: 0.172, K=6: 0.176, K=7: 0.176, K=8: 0.184, K=9: 0.187). At K=3 the clusters broadly represent *A. m. mellifera* (red), with the blue cluster representing the two C lineages, *A. m. carnica* and *A. m. ligustica*, and a third yellow cluster largely representing the UK and Ireland samples (Figure 4.5). While K=3 was the most likely K value, K=2 delineates the C and M lineages (Figure 4.4) and K=4 begins to define the two C- lineage subspecies (Figure 4.6). Lastly, at K=5 *A. m. carcnia*, *A. m. ligustica* and *A. m. mellifera* are divided into three distinct clusters (Figure 4.7).

Most samples from the UK and Ireland are admixed, particularly those located in south central England (Figures 4.4 to 4.7 and appendix 2). However, there are UK and Irish samples that retain a majority membership to the *A. m. mellifera* cluster (indicated in red throughout Figures 4.4 to 4.7) in all K values presented. These samples are located in the Inner Hebrides, Ireland, Inverness, Jersey, North West England and the Isle of Man.

Figure 4.4. K=2 ADMIXTURE analysis of honey bee colonies from the UK and Ireland along with a map showing the geographical distribution of the samples along with the membership values represented by a pie chart. Red represents the *A. m. mellifera* cluster while blue represents the C lineage cluster



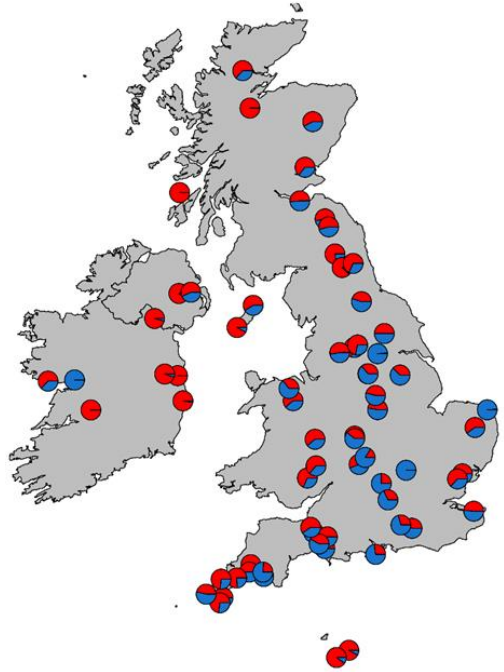


Figure 4.5. K=3 ADMIXTURE analysis of honey bee colonies from the UK and Ireland along with a map showing the geographical distribution of the samples along with the membership values represented by a pie chart. Red represents the *A. m. mellifera* cluster while blue represents the C lineage cluster and the yellow cluster belongs mainly to the British Isles

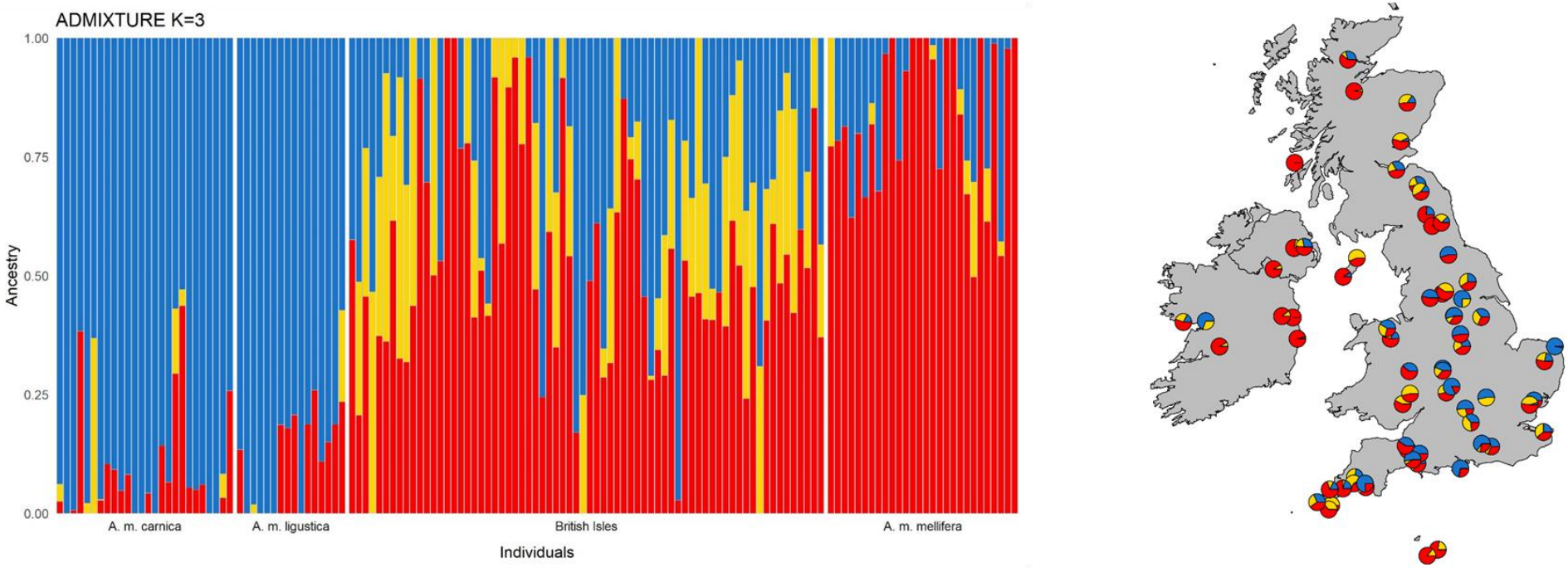
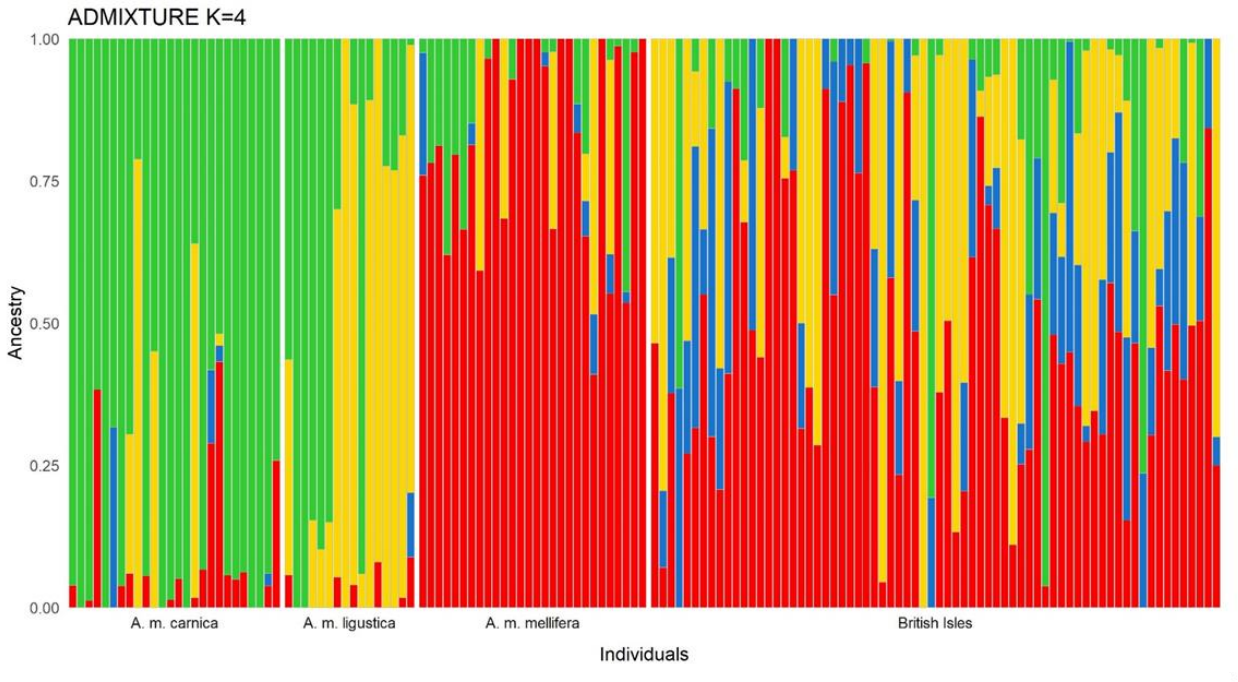


Figure 4.6 K=4 ADMIXTURE analysis of honey bee colonies from the UK and Ireland along with a map showing the geographical distribution of the samples along with the membership values represented by a pie chart. Red represents the *A. m. mellifera* cluster, green and yellow represents the C lineage clusters and the British Isles has a mixture of signals including blue.



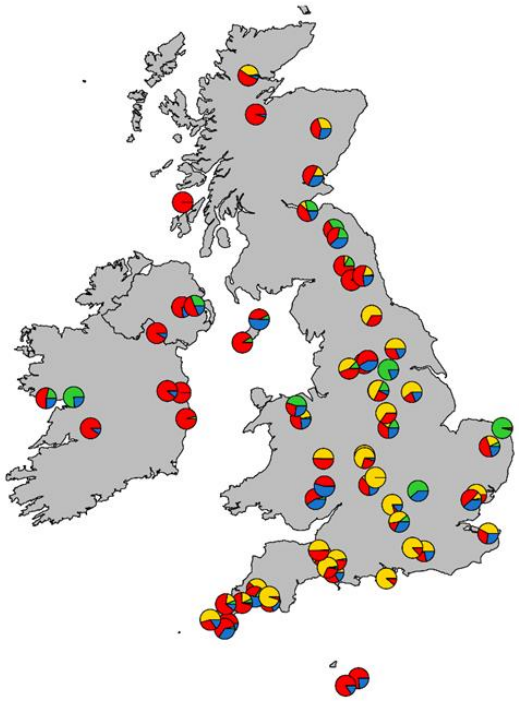
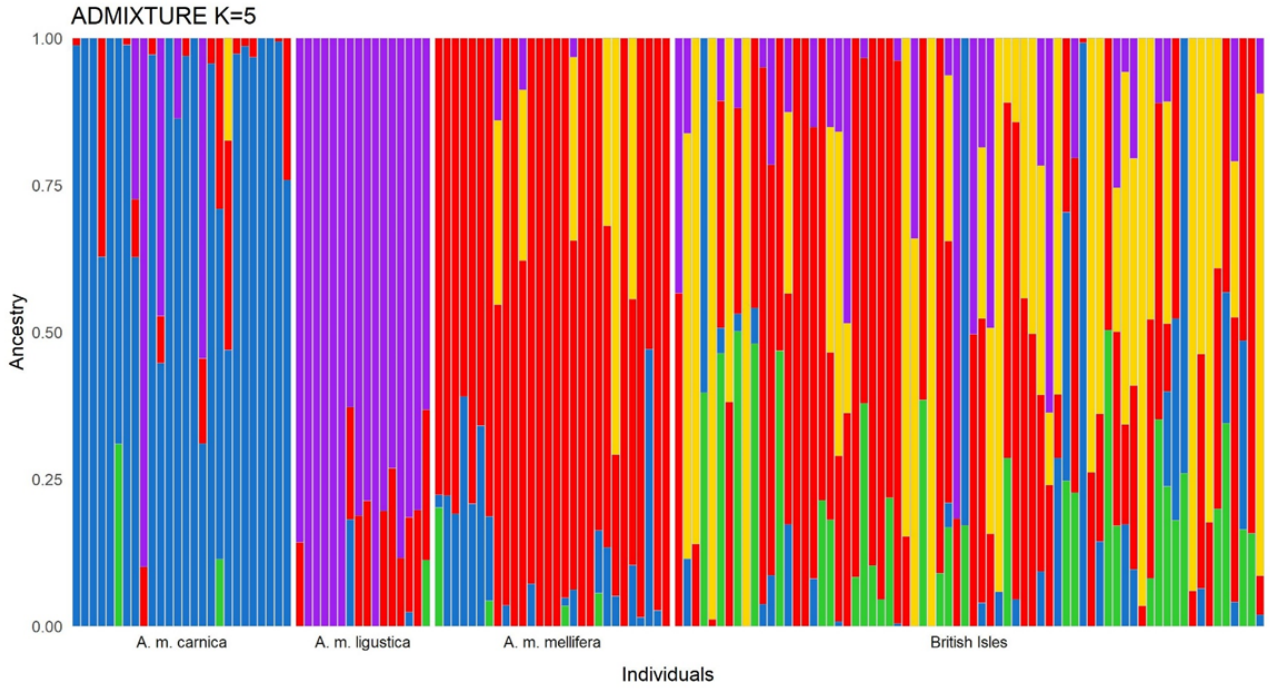
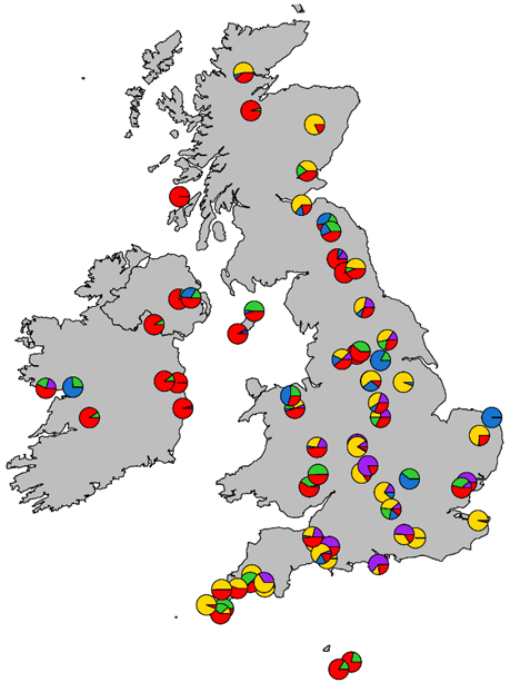


Figure 4.7. K=5 ADMIXTURE of honey bee colonies from the UK and Ireland along with a map showing the geographical distribution of the samples along with the membership values represented by a pie chart. Red represents the *A. m. mellifera* cluster, *A. m. carnica* is represented by a blue cluster, *A. m. ligustica* by a purple cluster and the British Isles has a mixture of signals including yellow





4.3.2 ABBA BABA introgression estimates

A total of 183,943 SNPs were used for the ABBA BABA analysis. A total of 53 samples resulted in significant D statistics (indicating introgression). While 17 samples resulted in non-significant D statistics (indicating no introgression) (Table 4.1). Of these, 12 were not significant for any introgression from C lineage honey bees, while 5 samples were non-significant for *A. m. carnica* but significant for *A. m. ligustica* introgression (Table 4.1).

Table 4.1 Colonies in the UK and Ireland that did not produce a significant positive D value indicating no introgression with the C lineage

UK and Ireland colonies with no introgression

Location	No introgression from
Tamar Valley, England	*A. m. ligustica or A. m. carnica*
Jersey, Channel Isles	*A. m. ligustica or A. m. carnica*
Isle of Man	*A. m. ligustica or A. m. carnica*
Colonsay, Inner Hebrides	*A. m. ligustica or A. m. carnica*
Co Wicklow, Ireland	*A. m. ligustica or A. m. carnica*
Co Monaghan, Ireland	*A. m. ligustica or A. m. carnica*
Jersey, Channel Isles	*A. m. ligustica or A. m. carnica*
Dublin, Ireland	*A. m. ligustica or A. m. carnica*
Belfast, Northern Ireland	*A. m. ligustica or A. m. carnica*
Inverness, Scotland	*A. m. ligustica or A. m. carnica*
Co Tipperary, Ireland	*A. m. ligustica or A. m. carnica*
Co Kildare, Ireland	*A. m. ligustica or A. m. carnica*
Newcastle upon Tyne, England	*A. m carnica*
Rame penninsula, Cornwall	*A. m carnica*
Hereford, England	*A. m carnica*
Colchester, England	*A. m carnica*
Cotswolds, England	*A. m carnica*

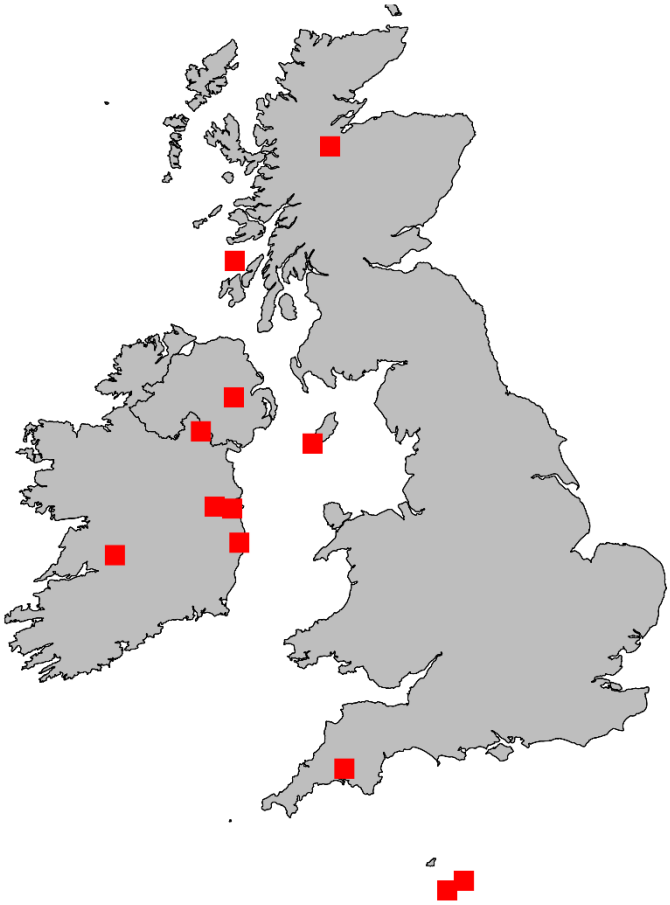


Figure 4.8. Locations of *A. m. mellifera* colonies with no introgression from either *A. m. carnica* or *A. m. ligustica* as calculated by the D statistic

Pure *A. m. mellifera* in the UK and Ireland were observed in Ireland, Northern Ireland, the Inner Hebrides, Northern Scotland, Jersey, The Isle of Man and the South West of England (Figure 4.8).

The windows of introgression derived from the testing of each colony individually then comparing outliers amongst all colonies of the UK and Ireland yielded no consistent windows of introgression. Colonies did have regions that were introgressed and remained outliers regardless of which *A. m. mellifera* sample or C lineage sample they were compared to. However, there were no consistent windows of introgression between colonies, indicating that there was no consistent pattern of outliers amongst the UK and Irish samples. Additionally, there were no loci that were outliers when the

UK and Ireland samples were examined as a population. The $\hat{f}_d$ value of introgression across all chromosomes can be seen to sit at a level of 0.6 (figure 4.9). Additionally there is an absence of low introgression regions in the scan.

In summary, no consistent introgression outliers were discovered amongst the colonies of the UK and Ireland, when tested separately as colonies or when tested as a population.

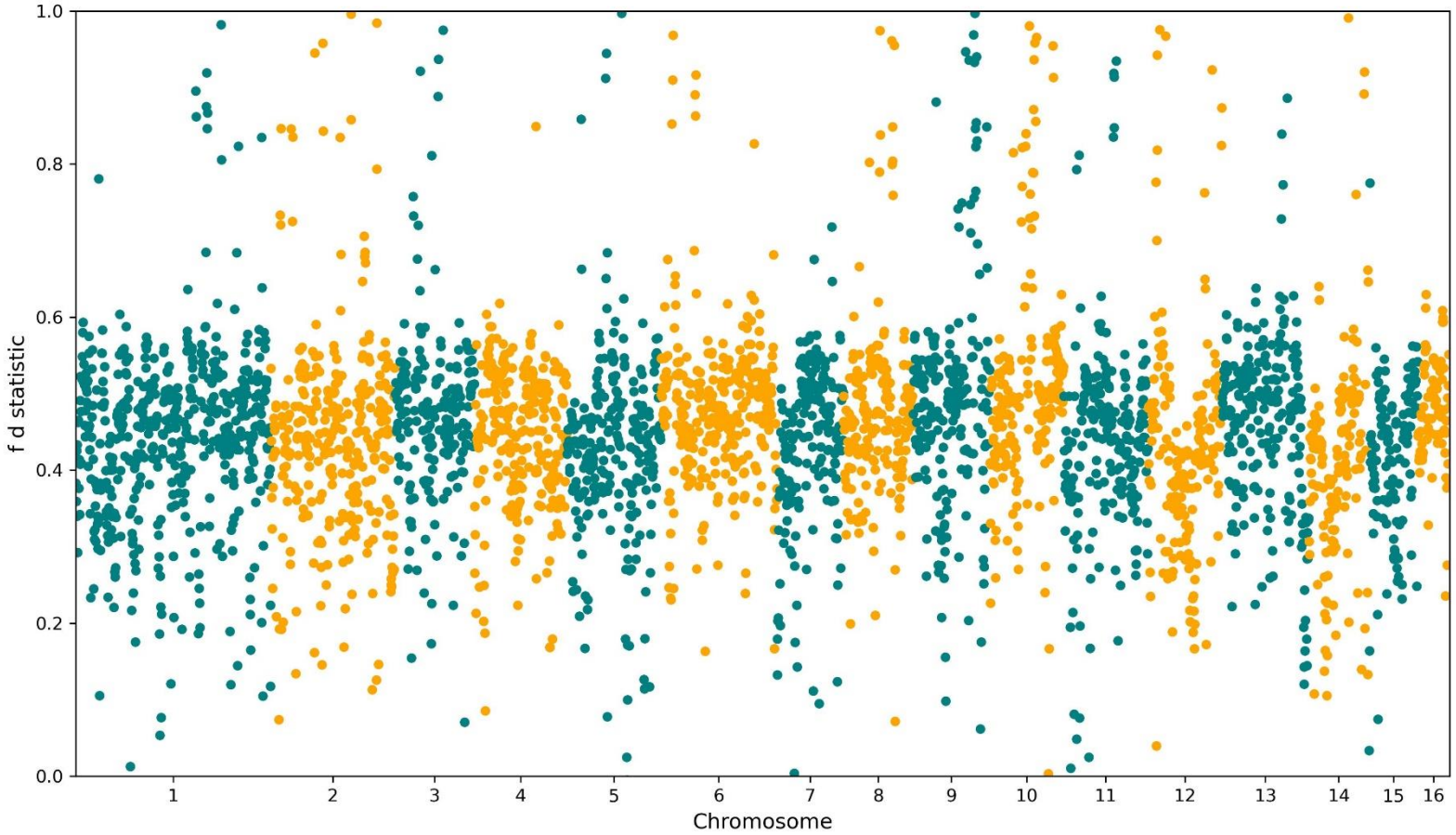


Figure 4.9. $\hat{f}_d$ sliding window analysis of introgression between British Isles population and C lineage standards performed in Dsuite.

4.4 Discussion

This study investigated the UK and Irish population of putative *A. m. mellifera* and examined introgression signals across the genome. While much of the UK and Ireland population is introgressed, new geographical sites of interest to *A. m. mellifera* conservation have been uncovered in Inverness, Jersey and the Isle of Man.

These results are consistent with the findings from previous studies into *A. m. mellifera* population in Ireland (Hasset *et al.,* 2018; Browne *et al.,* 2020), the Inner Hebrides (Jensen *et al.,* 2005; Regan *et al.,* 2018), and South West England (Ellis *et al.,* 2018),

The *A. m. mellifera* colony found here on the Isle of Man is a notable result as the Isle of Man law prohibits the transportation of honey bees on and off the Island (Importation of Bees Brder, 1988; Isle of Man Bees Act, 1989). The purpose of the act was to prevent bee disease from being introduced, particularly as at the time *Varroa destructor*, the parasitic mite, was spreading through Europe (Büchler *et al.,* 2010). The Isle of Man is 30km to 50km off the coast of the UK and the maximum mating distance ever recorded was 15km (Jensen *et al.,* 2005; Arundel *et al.,* 2012). Thus, it is unlikely that honey bees can naturally disperse there from any mainland population. However, only one of two colonies showed no introgression. The other colony showed introgression and can be seen in the ADMXITURE maps (Figures 4.4 to 4.7). This suggests that before the Isle of Man ceased imports foreign subspecies had been bought to the island.

In Jersey, both samples were classified as *A. m. mellifera*. Until now no *Apis mellifera* samples from Jersey have been included in subspecies assignment studies so this is

an important and potentially novel discovery. Despite Jersey being part of the British Isles it is located close to France and invertebrate species in Jersey face an additional threat from the invasive Asian hornet (*Vespa velutina nigrithorax)* (Jones *et al.,* 2020). Honey bees are especially at threat from the Asian hornet as they constitute a large part of their diet (Monceau *et al.,* 2014).

An *A. m. mellifera* colony was also identified in Inverness in the north of Scotland. Regan *et al* (2018) previously sampled this area. Regan *et al.* (2018) sampled 19 locations, 14 in Scotland and 4 in England. Although the primary goal in that study was to characterise the metagenome of the UK honey bee genome. Samples were also examined in ADMIXTURE alongside two Buckfast, two *A. m. carnica* samples and two samples representing *A. m. mellifera* from Colonsay. The use of ADMIXTURE in the study was to identify population structure, not measure introgression. Few standards were used in the study and all standards were from Scotland. In summary, this area has been sampled before but not tested specifically for introgression. Here we have confirmed that this area does contain *A. m. mellifera*.

While these are interesting results, a fine scale analysis to determine the extent of *A. m. mellifera* in these areas would be beneficial.

This study also examined patterns of introgression across the genome in an attempt to discover any consistently introgressed windows between colonies. Various studies have used sliding windows analysis to successfully identified regions of introgression. For example, introgression from red siskins (*Spinus cucullata*) into canaries (*Serinus canaria*) (Lopes *et al.,* 2016), between species of *Heliconiou*s butterflies (Enciso-Romero *et al.,* 2017) and from Asian zebu (*Bos taurus indicus*) into Ethiopian Sheko cattle (Bahbahani *et al.,* 2018). Importantly, regions of introgression have been observed in previous honey bee studies. Wragg *et al.* (2018) examined honey bee populations in the Reunion Isles and identified 15 regions that were preferentially

selected for and significantly associated with the native subspecies and 9 regions associated with introgression from C lineage subspecies. However, Wragg *et al.* (2018) used the program PCAdmix (Brisbin *et al.,* 2012) which requires haplotypes and a genetic map created using phased data. Phased data consists not only of genotype calls but also information on which chromosome contains which allele. Phased data allows the building of haplotype blocks which can be used to estimate ancestry. Phased data requires individual data as a reference and Wragg *et al*. (2018) used drone haplotypes for this purpose. Essentially, the Wragg *et al.* (2018) findings indicated that regions associated with introgression can be detected in honey bees despite the comparatively high recombination rate although the methods used in that study could not be implemented here. In this study, no common windows of introgression were identified across colonies either when windows were derived from each colony separately then compared across samples, or when all colonies were grouped into a population. Patterns of introgression were observed throughout the genome and in no consistent pattern. The introgression level across the genomes was around an $\hat{f}_d$ value of 0.6. If introgression was occurring at specific places, values should be highly variable (from 0 to 1) with peaks of introgression and peaks of non-introgressed regions. This result was not unexpected given the high recombination rate combined with uncontrolled breeding with historical and ongoing imports (Wragg *et al.,* 2018). It is possible that the $\hat{f}_d$ analysis performed in this study is not optimal. The criteria of a window being present across all samples could be too stringent (for example, examining if a window was introgressed in 90% of samples) and there is potential in sliding window analysis for the windows to be too large or too variable in size (regarding the location of SNPs along the genome) and further work may be needed to clarify these issues.

In conclusion, continued imports and high recombination rates in this species have resulted in genome wide introgression across UK and Ireland with no particular pattern of introgressed loci. This study relied on samples being sent in by bee keepers and while this study has good coverage across the UK and Ireland it is not a comprehensive examination of the population and more sampling could further inform the picture for the UK and Ireland. Additionally, the new locations of *A. m. mellifera* found here would benefit from finer scale investigation.

5. Genetic assessment and comparison of subspecies standards

5.1 Introduction

In an attempt to understand subspecies trait differences, the INTERREG bee research project was funded and set up in Sweden and Norway by the University of Skovde with the aim to assess and compare traits of honey bee subspecies. In partnership with the INTERREG project this chapter will genetically assess the colonies used in the experiment. The INTERREG project is potentially the first experiment aimed at disentangling the hardwired subspecies specific differences and environmental effects in European honey bee subspecies.

The INTERREG project established 5 apiaries, 3 in Sweden and 2 in Norway. In Sweden each apiary contained 16 colonies consisting of 3 subspecies and one hybrid (*A. m. mellifera, A. m. ligustica, A. m. carnica*, and the Buckfast hybrid). Additionally, each subspecies was represented by 4 different breeding lines, 3 from Sweden and one imported line (*A. m. ligustica* from Italy, *A. m. carnica* from Austria, and *A. m. mellifera* from the south west of England). The two Norwegian apiaries contained 2 subspecies (*A. m. mellifera* and *A. m. carnica*) and the hybrid Buckfast. In the Norwegian apiaries no breeding lines were imported. This is due to strict importation laws in Norway (perr comms S. Leidenberger 2022) and legal protection for *A. m. mellifera* and *A. m. carnica.* The Norwegian government also runs conservation breeding programmes (Ruottinen *et al.,* 2014; Demant *et al.,* 2019). Therefore, all

the samples at the Norwegian apiaries originated from Norwegian breeding lines. The experiment was set up to collect data on a large number of phenotypic traits as well as a number of environmental parameters over two years from 2019 to 2022 (Norrström *et al.,* 2021.) (Figure 5.1).

Table 5.1 Measurements monitored in the INTERREG project. Monitoring was performed from 2019 to 2022 across 5 apiaries containing three subspecies and a hybrid honey bee

INTERREG project	
Category	Measurements
Health	Varroa counts
	Nosema
Honey	Production (amount)
	Quality (Water, pH)
	Pesticides
	Drugs
	Pollen
	Bacteria
Pollen collection	Pollen species
	Seasonal differences
Behaviour	Swarming
	Foraging
	Fanning
	Defence
	Guarding
	Self-grooming
	Resin-collection
	Start of breeding activities
Weight	Collection of honey
	weight loss during winter
Life history	age at first flight
	winter survival
	survival (juveniles/adults)
	flight acitivty (per day, season)
Environmental	Temperature
	Rain/snow
	Wind speed
	Wind direction

The ‘purity’ of subspecies in this experiment is important, if the study is attempting to investigate behavioural or phenotypic measurements and wishes to differentiate between subspecies, it’s imperative that the samples represent the putative subspecies they have been classified as. While the project has taken steps to identify the subspecies in the project using morphometric and breeding line information, no genetic assessment has been performed on the colonies.

As well genetic analysis to assign subspecies in the INTERREG project, this data set allows for the study of the genetic differences between subspecies. Previous chapters have concentrated on the introgression levels in population and breeding programmes using the subspecies as standards to calibrate the other populations, but questions remain around how genomic differences lead to phenotypic differences, if any exist, between these subspecies. The genetic data set resulting from the subspecies in the INTERREG experiment allows for a preliminary investigation into this question.

Studies have consistently observed that C lineage and *Apis mellifera mellifera* subspecies group into separate clusters in population structure analysis (Soland-Reckeweg *et al.,* 2009; Oleksa *et al.,* 2011; Pinto *et al.,* 2014; Muñoz *et al.,* 2015; Ellis *et al.,* 2018; Henriques *et al.,* 2018; Browne *et al.,* 2020). Evaluating the genomic location and effect of the SNPs that contribute to population structure can be the first step in uncovering the genetic basis of difference between groups (Cavallo and Martin, 2005; Miller *et al.,* 2020). The location, inside or outside of a coding region, and effect of a SNP, synonymous or non-synonymous, can inform about the possible biological impact of the SNPs driving the differentiation between groups (Jonbart and Ahmed, 2011; Filippi *et al.,* 2015; Quigley *et al.,* 2020).

Alterations in a base pair can result in different outcomes. For example, a non-synonymous change is where the resulting amino acid produced is different, while synonymous changes are referred to as silent, resulting in no change to the subsequent amino acid (Futuyma and Kirkpatrick 2017). These synonymous changes are not, however, always truly silent. Synonymous changes have been observed to have influences on gene expression, cellular function and protein folding (Hunt *et al.,* 2009; Hunt *et al.,* 2014). Therefore synonymous changes might not be totally silent (Goymer 2007; Zhou et al., 2012; Lui *et al.,* 2021).

This chapter aims to assess the subspecies involved in the INTERREG project. Samples will be examined for subspecies assignment using a clustering program ADMIXTURE, then the genetic variation among individuals will be investigated using a PCA and finally the biological roles of SNPs that contribute to the subspecies differentiation via a DAPC and SNP annotation programme.

5.2 Methods

5.2.1 Sample collection

A total of 51 colonies were sequenced from the Sweden and Norway experiment (putative subspecies assignment of *A. m. mellifera* (n=21), *A. m. carnica* (n=19), *A. m. ligustica* (n=11)). For comparison additional European standards were included, *A. m. mellifera* from Switzerland (n=7), *A. m. carnica* from Germany (n=7) and *A. m. ligustica* from Italy (n=5), resulting in a total of 70 colony samples. Colonies were collected by collaborating researcher institutions and placed into 70% ethanol. This chapters data was processed and sequenced along with the samples examined in chapter 4.

5.2.2 DNA extraction

DNA extraction was performed using the pooled samples method (see chapter 3 section 3.2.3.2). In summary, each colony sample of 30 worker bees was extracted in six batches (5 thoraxes per batch) using an ammonium acetate protocol (Nicholls *et al.,* 2000). These batched extractions produced an average yield of 2110 ng of DNA per extraction. Each colony was equimolarly pooled into 1000ng pool representing 30 worker bees before sending for sequencing.

5.2.3 Library prep and sequencing

All library prep was performed by BGI in Hong Kong on the pooled colony samples (Xu *et al.,* 2019).

5.2.4 Bioinformatics

The bioinformatics pipeline here is the same as described in chapter 3 section 3.2.4.2. As a brief overview:

The raw data received from BGI had an average of 136 million reads per sample with an average of 96% of reads exceeding a phred score of 20. Reads were filtered for quality and trimmed (Trimmomatic, Bolger *et al.,* 2014), paired reads were aligned to the genome and filtered for mapping quality and arranged in order (BWA MEM aligner version 0.7.17 and Samtools) (Li, 2013; Li *et al.,* 2019). After read group editing (PICARD) the GATK best practices pipeline was followed (Van der Auwera *et al.,* 2013; Van der Auwera and O'Connor, 2020). Sites and samples were then filtered to reduce missing data (vcftools) (Danecek *et al.,* 2011). Finally, to avoid overstratification in subsequent analysis (Wragg *et al.,* 2018) arbitrary thinning was applied by removing any SNPs closer than 1kb to one another using the '–thin' command in vcftools (Danecek *et al.,* 2011).

5.2.5 ADMIXTURE

ADMIXTURE (Alexander, *et al.,* 2009, also used in chapter 3 section 3.3.1), is a clustering algorithm that estimates ancestral groups from SNP data. The programme estimates ancestry membership proportions (Q values) of samples to ancestral clusters (K) and uses a cross validation (CV) procedure to estimate the number of clusters that best explains the data.

ADMIXTURE analysis was performed on plink formatted bed files (Purcell *et al.,* 2007). These were produced using the vcftools command '–plink' (Danecek, *et al.,* 2011) and then the plink command '–make-bed' (version 1.07). To assess the most likely number of clusters CV values were viewed for K1 to K10 at default settings (ADMIXTURE version Linux 1.3.0). ADMIXTURE graphs were produced in R (R core team., 2018) using ggplot 2 (version 3.3.5) (Wickham, 2016).

5.2.6 Principal component analysis

A principal component analysis (PCA) is a multivariate approach that can be used to assess the distribution of genetic variation among samples, which is made up of within group variation and between group variation. A PCA on genetic data uses synthetic variables as linear combinations of alleles (Jombart *et al.,* 2009) and each axis is independent of the next (no correlation between them). Effectively, the principal component (PC) axes represent variation in allele frequencies and the spatial position of the sample is based on a comparison of the total genetic variation amongst all the samples. PCAs are not model based and so no pre-analysis assumptions (such as number of clusters) need to be made.

The WGS pooled colony filtered vcf file resulting from the bioinformatics pipeline was entered into R to perform the PCA. Specifically, the vcf file was imported to R (R core

team., 2018) as a genlight object using vcfR library (version 1.12.0) (Knaus and Grünwald., 2017). The PCA was performed using the adegenet (version 2.1.5) command glPca (Jombart *et al.,* 2010). Results were visualised using ggplot2 (version 3.3.5) (Wickham, 2016).

5.2.7 Discriminant Analysis of Principal Components

A Discriminant Analysis of Principal Components (DAPC) places emphasis on between group variation and minimises the importance of within group variation (Jombart *et al.,* 2010). Discriminant functions are synthetic variables made from combinations of (here) alleles that explain the variance between groups. However, in order to examine the difference between groups, genetic clusters (K) are defined before the analysis to maximise the variation between groups (Jombart and Collins, 2015). To infer the most appropriate number of clusters (K) the R package adegenet (version 2.1.5) was used. This employs a clustering algorithm which allows for the comparison of Baysian Information Criterion (BIC) values associated with k values. The most appropriate k value is often indicated by an 'elbow' in the curve (Jombart and Collins, 2015). As well as the most appropriate k value, the optimum number of PCs to calculate discriminant functions is also required. This is a trade-off between too few PCs, where important information is lost and the results uninformative, and too many PCs, where over-fitting of the data can be the result. Over fitting can produce misleading results, specifically a large separation between groups. To inform the most useful number of PCs adegenet implements the a-score test. The a-score is the result from repeatedly performing the DAPC analysis using random groups and computing a score for each group. The a-score is the difference between observed discrimination (the proportion of successful reassignment) and the random discrimination (values obtained using random groups) (Jombart and Collins, 2015).

Before this analysis was performed any samples with an introgression Q value of >0.05 (Groeneveld *et al.,* 2020; Vaha and Primmer 2005) indicating introgression at an arbitrary threshold were removed. This threshold has been used in previous studies (Hassett *et al.,* 2018; Groeneveld *et al.,* 2020). This was to help evaluate the difference between *A. m. mellifera* and C lineage (*A. m. carnica* and *A. m. ligustica*) honey bees.

The vcf file resulting from the bioinformatic pipeline was imported in to R to perform the DAPC. Specifically, the vcf file was imported as a genlight object using the vcfR library (version 1.12.0) (Knaus and Grünwald., 2017). To assess the most likely number of groups (K) the adegenet command find.clusters (version 2.1.5) was used while retaining all PCs (Jombart and Collins, 2015). The Bayesian Information Criterion (BIC) values were evaluated to determine the K value that best explains the data. The output from find.clusters was used to perform an initial DAPC on the data while retaining all PCs. This initial DAPC was examined for a-score using the optim.a.score command in adegenet (Jombart and Collins, 2015). This calculated a-scores to inform the optimum number of PCs. After viewing the a-score results an optimised DAPC was performed with the most appropriate number of PCs and cluster (k). The DAPC results were then visualised in R using the scatter command from ade4 (version 1.7-18) (Dray and Dufour, 2007).

5.2.8 Annotation of SNPs

To examine SNPs that contribute to the discriminant function the var.contr slot was obtained from the optimised DAPC. The var.contr slot contains the variable contributions to each discriminant factor in a DAPC object. The top 1% of contributing SNPs were output in a list and matched to the chromosome and position information using a custom python code. A vcf file was filtered to only contain these top

contributing SNPs using the vcftools –positions command. The resulting vcf file was then annotated using SnpEff (version 5.0e) and filtered using SnpSift (Cingolani *et al.,* 2012). First, the fasta file and the GTF annotation file (annotation release 104) for the reference genome Amel_HAv3.1 were downloaded and added to the SnpEff database using the build command. The vcf of the top contributing SNPs were then annotated and a summary statistic output file and gene count summary file were produced. Next, the SnpSift filter command was used to obtain only SNPs that were in coding regions. The coding region gene symbols were then examined for gene ontology terms by placing the list of gene symbols into the Database for Annotation, Visualization and Integrated Discovery (DAVID) (v2022q1) (Huang *et al.,* 2009; Sherman *et al.,* 2022) and gene symbols were double checked by manually inputting into the NCBI gene search. DAVID tools were used to examine gene ontology further and attempt to gain some insight into the types of biological differences between the subspecies. The tools used were the gene functional classification tool, the functional annotation clustering tool and the gene ontology biological processes tool. The gene functional tool groups genes that are functionally similar together while the functional annotation tool groups annotation terms together. The biological processes tool groups gene ontology terms into similar biological processes. All DAVID tools were run using default settings.

5.3 Results

The full data set used in the ADMIXTURE analysis and the PCA consisted of all standards represented by 70 colonies and 208,587 SNPs. The second data set represented low introgression colonies and was used in the DAPC analysis, this consisted of 48 colonies and 208, 587 SNPs.

5.3.1 ADMIXTURE

The most likely number of clusters was K=2, while K=3 also presented a similar CV value and is also appropriate for describing the data (CV values, K=1: 0.20, K=2: 0.15, K=3: 0.159, K=4: 0.162, K=5: 0.171, K=6: 0.177, K=7: 0.181, K=8: 0.194, K=9: 0.201, K=10: 0.219).

The *A. m. mellifera* samples with the highest introgression values were Norway H Mel 147 with Q value membership to *A. m. ligustica*, while Norway D Mel 155 and Switzerland Mel 17 had Q value membership to *A. m. carnica* (Figure 5.1, Figure 5.2 and Table 5.2). There were 11 *A. m. mellifera* samples that showed no introgression, all from Norway and Sweden, while all *A. m. mellifera* samples from Switzerland showed some degree of Q value membership to a cluster other than *A. m. mellifera*.

In K=2 (Figure 5.1), one cluster represents both C-lineage honey bees, *A. m. ligustica* and *A. m. carnica*, and one cluster represents *A. m. mellifera*. In the K=3 analysis the three subspecies separate out into the three clusters (Figure 5.2). In both K=2 and K=3 some introgression was found across all subspecies.

At K=3 five *A. m. carnica* colonies showed membership to the *A. m. ligustica* cluster that was not detected at K=2. This is because at the K=2 analysis the C-lineage

subspecies are clustered together. For example the *A. m. carnica* sample, Sweden NA 29, at K=2 clustered with C-lineage and had an introgression Q value of 0.091 (Figure 5.2 and Table 5.2,). But at K=3 when the C-lineage are split across two clusters Sweden NA 29 showed a Q value membership of 0.89 to *A. m. ligustica*. A similar pattern is seen with Sweden NA 39 (k=2; 0.07 and k=3; 0.48) and Sweden U 63 (k=2; 0.14 and k=3; 0.55) which showed membership to *A. m. ligustica* (Figure 5.2, Figure 5.2 and Table 5.2).

There were 4 *A. m. carnica* samples, Norway D Car 139, Germany Car 4, Norway D Car 138 and Norway H Car 163, that had membership values to the *A. m. mellifera* cluster (Table 5.2). The *A. m. carnica* sample with the largest *A. m. mellifera* membership was Norway D 139 with an *A. m. mellifera* Q value membership of 0.45 at K=2 and 0.42 at K=3 (Table 5.2).

A. m. ligustica from Italy had the least introgression with only one sample present with any membership values to any other cluster. The *A. m. ligustica* sample with the most introgression at K=3 was Sweden U Lig 22, while at K=2 Sweden NA Lig 43 showed the highest introgression. Sweden U Lig 22 was the only *A. m. ligustica* sample that showed membership to the *A. m. carnica* at K=3.

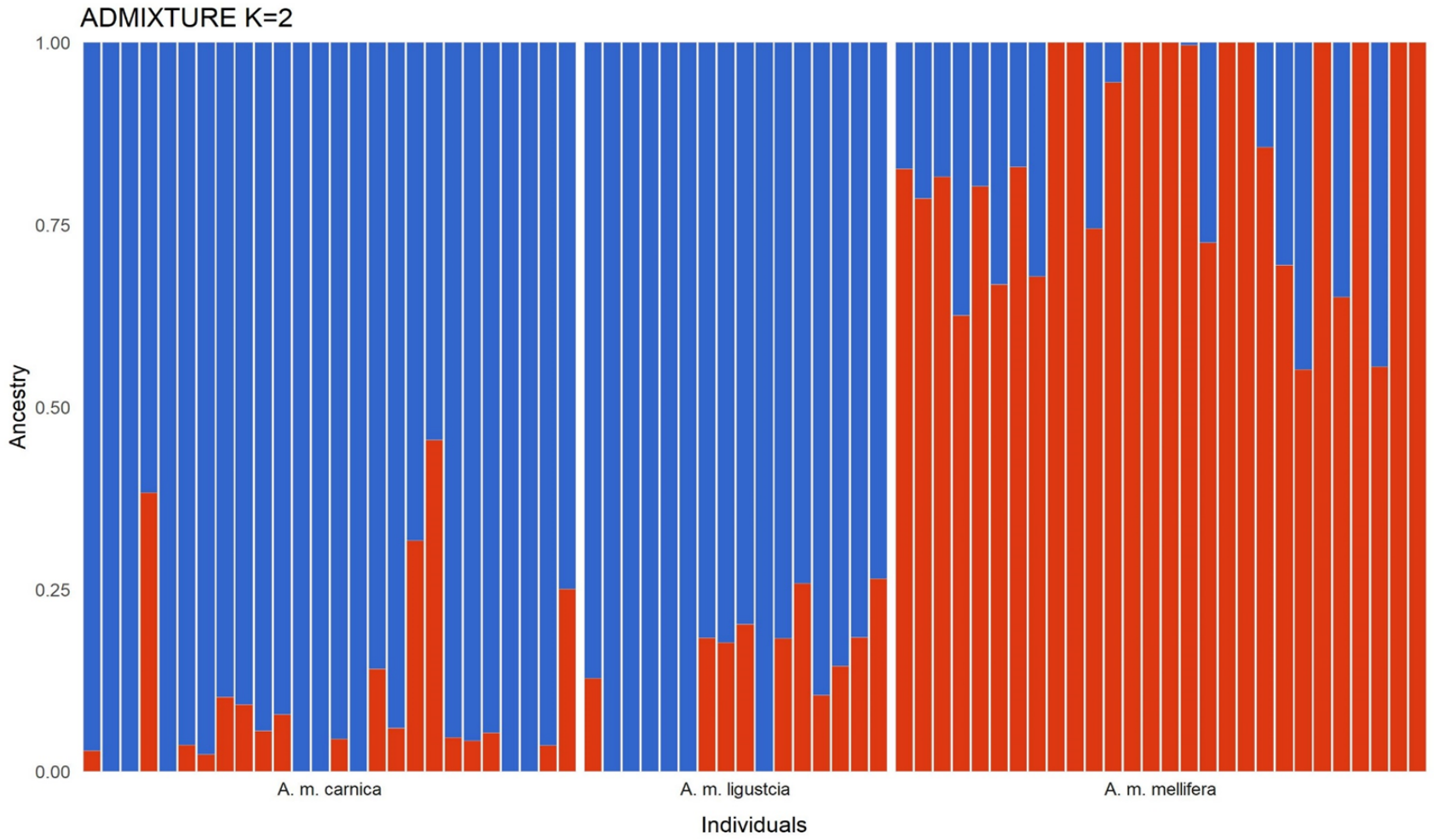


Figure 5.1 ADMIXTURE analysis. K=2 of 208,587SNPs across 70 colonies. The red cluster representing the *A. m. mellifera* and the blue cluster representing the C lineage honey bees

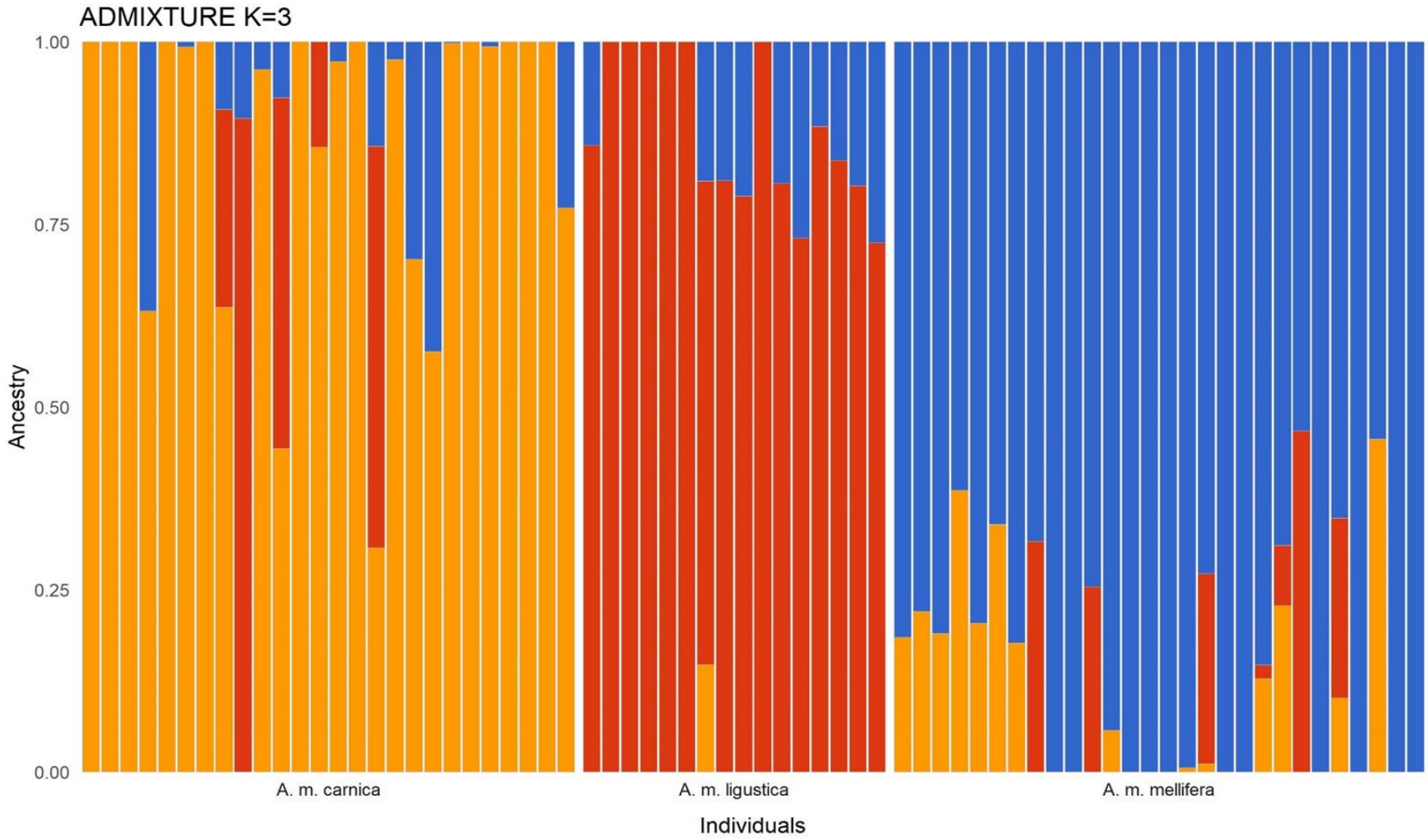


Figure 5.2. ADMIXTURE analysis K=3 of 208,587SNPs across 70 colonies. Here, the blue cluster represents *A. m. mellifera*, the red cluster represents *A. m. ligustica* and the yellow cluster represents *A. m. carnica*

Table 5.2. ADMIXTURE membership Q values for K=2 and K=3 estimated using 208,587 SNPs. Results are presented in the same order as the bar plots in Figure 5.1 and Figure 5.2

ADMIXTURE						
Colony		K=2		K=3		
Location and ID	Putative subspecies	C-lineage	*A. m. mellifera*	*A. m. mellifera*	*A. m. ligustica*	*A. m. carnica*
Germany Car 1	*A. m. carnica*	0.971	0.029	0.000	0.000	1.000
Germany Car 2	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Germany Car 3	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Germany Car 4	*A. m. carnica*	0.618	0.382	0.369	0.000	0.631
Germany Car 5	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Germany Car 6	*A. m. carnica*	0.964	0.036	0.007	0.000	0.993
Germany Car 7	*A. m. carnica*	0.977	0.023	0.000	0.000	1.000
Sweden U Car 23	*A. m. carnica*	0.898	0.102	0.093	0.270	0.637
Sweden NA Car 29	*A. m. carnica*	0.908	0.092	0.105	0.895	0.000
Sweden NA Car 35	*A. m. carnica*	0.944	0.056	0.038	0.000	0.962
Sweden NA Car 39	*A. m. carnica*	0.921	0.079	0.076	0.480	0.443
Sweden J Car 46	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Sweden J Car 48	*A. m. carnica*	1.000	0.000	0.000	0.144	0.856
Sweden U Car 60	*A. m. carnica*	0.955	0.045	0.027	0.000	0.973
Sweden U Car 62	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Sweden U Car 63	*A. m. carnica*	0.859	0.141	0.143	0.550	0.307
Norway H Car 137	*A. m. carnica*	0.940	0.060	0.024	0.000	0.975
Norway D Car 138	*A. m. carnica*	0.683	0.317	0.298	0.000	0.702
Norway D Car 139	*A. m. carnica*	0.545	0.455	0.424	0.000	0.576
Norway D Car 142	*A. m. carnica*	0.954	0.046	0.002	0.000	0.998
Norway H Car 143	*A. m. carnica*	0.958	0.042	0.000	0.000	1.000
Norway D Car 151	*A. m. carnica*	0.947	0.053	0.006	0.000	0.994
Norway H Car 156	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Norway H Car 157	*A. m. carnica*	1.000	0.000	0.000	0.000	1.000
Norway D Car 158	*A. m. carnica*	0.965	0.035	0.000	0.000	1.000
Norway H Car 163	*A. m. carnica*	0.750	0.250	0.227	0.000	0.773
Italy Lig 8	*A. m. ligustica*	0.872	0.128	0.141	0.859	0.000
Italy Lig 9	*A. m. ligustica*	1.000	0.000	0.000	1.000	0.000
Italy Lig 10	*A. m. ligustica*	1.000	0.000	0.000	1.000	0.000
Italy Lig 12	*A. m. ligustica*	1.000	0.000	0.000	1.000	0.000
Italy Lig 13	*A. m. ligustica*	1.000	0.000	0.000	1.000	0.000
Sweden U Lig 21	*A. m. ligustica*	1.000	0.000	0.000	1.000	0.000
Sweden U Lig 22	*A. m. ligustica*	0.817	0.183	0.191	0.662	0.147
Sweden U Lig 26	*A. m. ligustica*	0.823	0.177	0.190	0.810	0.000
Sweden U Lig 27	*A. m. ligustica*	0.798	0.202	0.211	0.789	0.000
Sweden NA Lig 34	*A. m. ligustica*	1.000	0.000	0.000	1.000	0.000
Sweden NA Lig 41	*A. m. ligustica*	0.818	0.182	0.194	0.806	0.000
Sweden NA Lig 43	*A. m. ligustica*	0.742	0.258	0.269	0.731	0.000
Sweden J Lig 45	*A. m. ligustica*	0.896	0.104	0.116	0.884	0.000

Sweden J Lig 50	*A. m. ligustica*	0.855	0.145	0.162	0.838	0.000
Sweden J Lig 54	*A. m. ligustica*	0.816	0.184	0.197	0.803	0.000
Sweden J Lig 58	*A. m. ligustica*	0.736	0.264	0.275	0.725	0.000
Switzerland Mel 14	*A. m. mellifera*	0.173	0.827	0.815	0.000	0.185
Switzerland Mel 15	*A. m. mellifera*	0.214	0.786	0.780	0.000	0.220
Switzerland Mel 16	*A. m. mellifera*	0.184	0.816	0.810	0.000	0.190
Switzerland Mel 17	*A. m. mellifera*	0.374	0.626	0.614	0.000	0.386
Switzerland Mel 18	*A. m. mellifera*	0.197	0.803	0.796	0.000	0.204
Switzerland Mel 19	*A. m. mellifera*	0.332	0.668	0.661	0.000	0.339
Switzerland Mel 20	*A. m. mellifera*	0.170	0.830	0.823	0.000	0.177
Sweden U Mel 24	*A. m. mellifera*	0.321	0.679	0.684	0.316	0.000
Sweden U Mel 28	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Sweden U Mel 30	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Sweden NA Mel 37	*A. m. mellifera*	0.255	0.745	0.746	0.254	0.000
Sweden NA Mel 38	*A. m. mellifera*	0.055	0.945	0.943	0.000	0.057
Sweden U Mel 40	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Sweden NA Mel 42	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Sweden J Mel 47	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Sweden J Mel 52	*A. m. mellifera*	0.003	0.997	0.994	0.000	0.006
Sweden J Mel 53	*A. m. mellifera*	0.274	0.726	0.728	0.260	0.011
Sweden J Mel 55	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Norway D Mel 136	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Norway H Mel 140	*A. m. mellifera*	0.144	0.856	0.854	0.019	0.128
Norway H Mel 144	*A. m. mellifera*	0.306	0.694	0.689	0.083	0.227
Norway H Mel 147	*A. m. mellifera*	0.448	0.552	0.533	0.467	0.000
Norway D Mel 150	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Norway D Mel 152	*A. m. mellifera*	0.349	0.651	0.653	0.246	0.102
Norway H Mel 153	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Norway D Mel 155	*A. m. mellifera*	0.445	0.555	0.544	0.000	0.456
Norway H Mel 162	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000
Norway D Mel 164	*A. m. mellifera*	0.000	1.000	1.000	0.000	0.000

5.3.2 Principal component analysis

The first two PCs explained 45.8% of the variance, with PC1 explaining the majority (43.1%). PC1 on the x axis separates the *A. m. mellifera* subspecies from the two C-lineage subspecies, while PC2 on the y axis separates the *A. m. carnica* and *A. m.*

ligustica samples (Figure 5.3). The samples generally group with their putative subspecies with a few exceptions. The *A. m. carcnica* samples higher on the PC1 axis towards the *A. m. mellifera* group are Norway D Car 139, Germany Car 4, Norway D Car 138 and Norway H Car 163 (Figure 5.3) and the Sweden NA car 29 sample is paced closer to the *A. m. ligustica* group. The *A. m. ligustica* sample that places highest on the PC2 axis is Sweden U Lig 22, while the *A. m. ligutsica* samples that is placed highest on the PC1 axis is Sweden J Lig 58 (Figure 5.3). The *A. m. mellifera* samples lowest on the PC1 axis are Norway D Mell 155, Norway H Mel 147 and Switzerland Mel 17. These are the same *A. m. mellifera* samples that had high introgression Q values in the ADMIXTURE analysis.

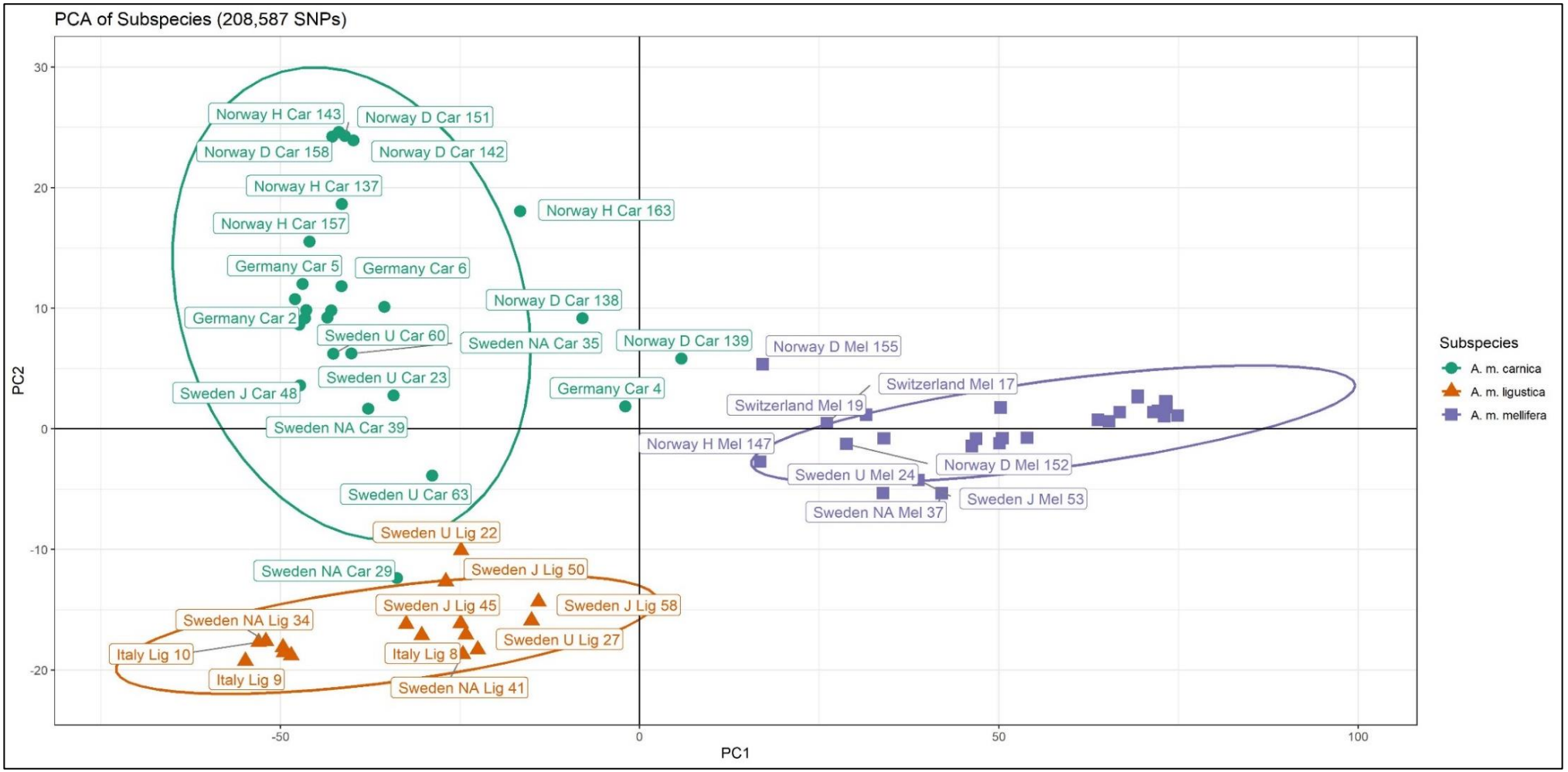


Figure 5.3. Principal component analysis from Adegenet. PC1 explains 43.1% of the variance while PC2 explains 2.7% of the variation in the data set. Here putative subspecies is represented by colours and shapes. *A. m. carnica* is represented by green circles, *A. m. ligustica* by orange triangles and *A. m. mellifera* by purple squares. Samples are circled with a group ellipse with a confidence level of 0.95.

5.3.3 Discriminant Analysis of Principal Components and genetic variants contributing to subspecies discrimination

The most likely number of clusters explained by the data was K=2 (Figure 5.4) and one principle component was most useful for describing the data. The samples that were grouped into each genetic cluster were consistent with the putative lineage of all colonies (Figure 5.5). The patterns of SNP contributions to the discriminant function were distributed evenly across the genome, although sections of high contributions of alleles can be seen on chromosome 11 (Figure 5.6). This section of high contributions of SNPs ran on chromosome 11 ran from base pair 3,793,613 to 8,241,540 and this region has 1017 genes within it (exon data viewed via the NCBI genomic data viewer for the Amel_HAv3.1).

Of the top 1% of contributing SNPs examined in SnpEff the majority of the SNPs and effects were classed as 'modifier' meaning they are non-coding with no evidence of an impact or impacts are hard to predict. In there were 27 SNPs estimated by SnpEff to have a 'moderate' impact and 124 SNPs predicted as having a 'low' impact (Figure 5.3). All moderate changes were classified as non-synonymous changes. Low effects varied from synonymous changes, downstream gene variants, upstream gene variants, intron variants or changes to start or stop codons.

Of the 151 Low and Moderate SNPs 139 were identified by DAVID (v2022q1). The gene functional classification resulted in 3 groups with the highest enrichment score for group 1 (enrichment score 1.37) (Figure 5.4). Group 1 contained the genes acitivating signal cointegretor 1 coomplex subunit3 (LOC552814), helicase domino (LOC413341), Lymphoid-specific helicase (LOC726235), Helicase SKI2W (LOC413690), ATP-dependant DNA helicase Q4 (LOC410301) (Figure 5.4).

In the Functional Annotation clustering analysis and biological processes clustering were not significant for any groups and had false discovery rates of 1.

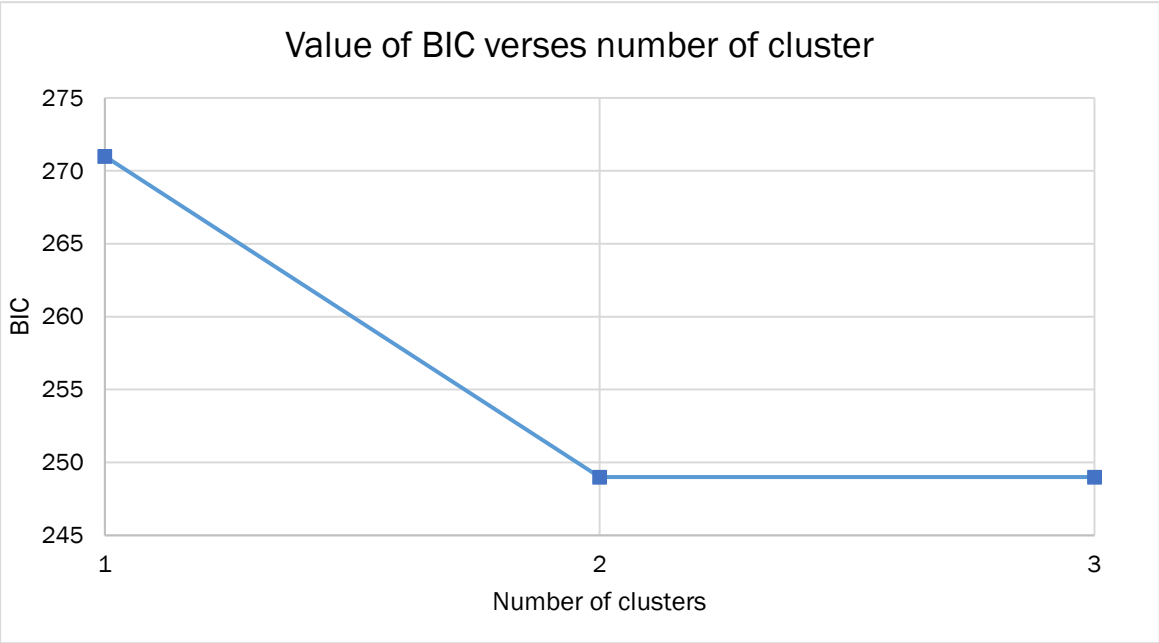


Figure 5.4 Results from Adegenet find.clusters command performed with all principal components. Number of clusters (K) on x axis and Bayesian Information Criterion (BIC) values on the y axis

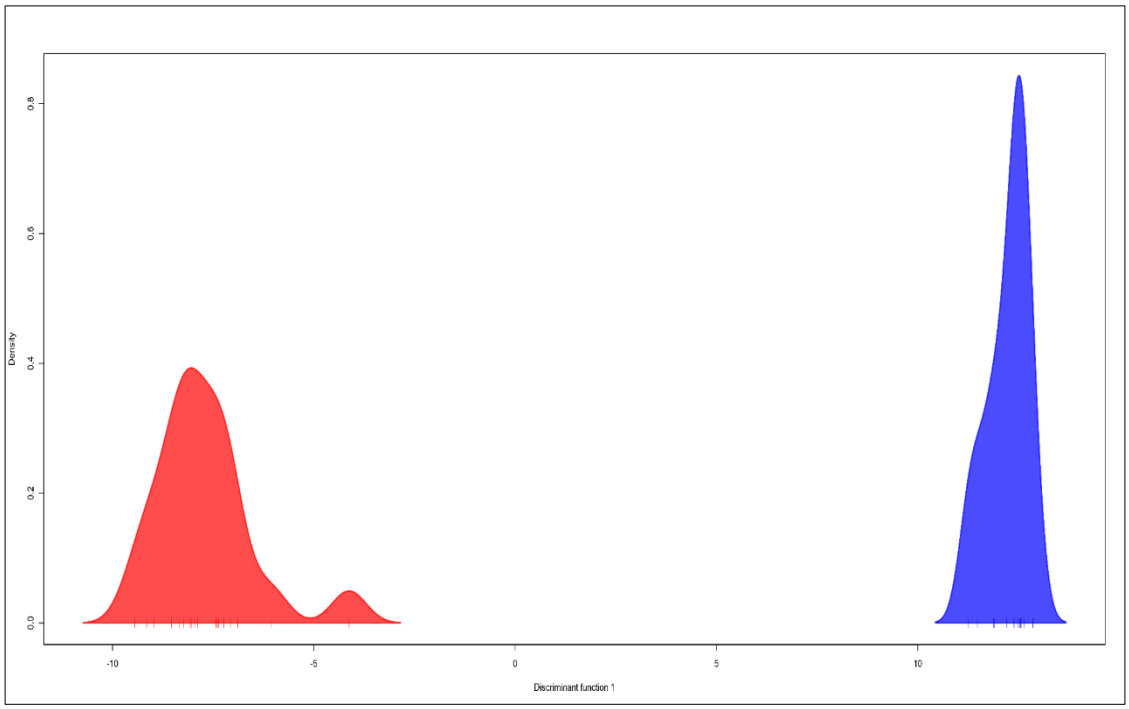


Figure 5.5. A discriminant analysis of principal components performed on low introgression colonies using 208, 587 SNPs using one discriminant factor using two genetic clusters. The x axis in the discriminant function and y represents the density of samples at the coordinates. C-lineage samples are represented by the red group and *A. m. mellifera* samples are represented by the blue group

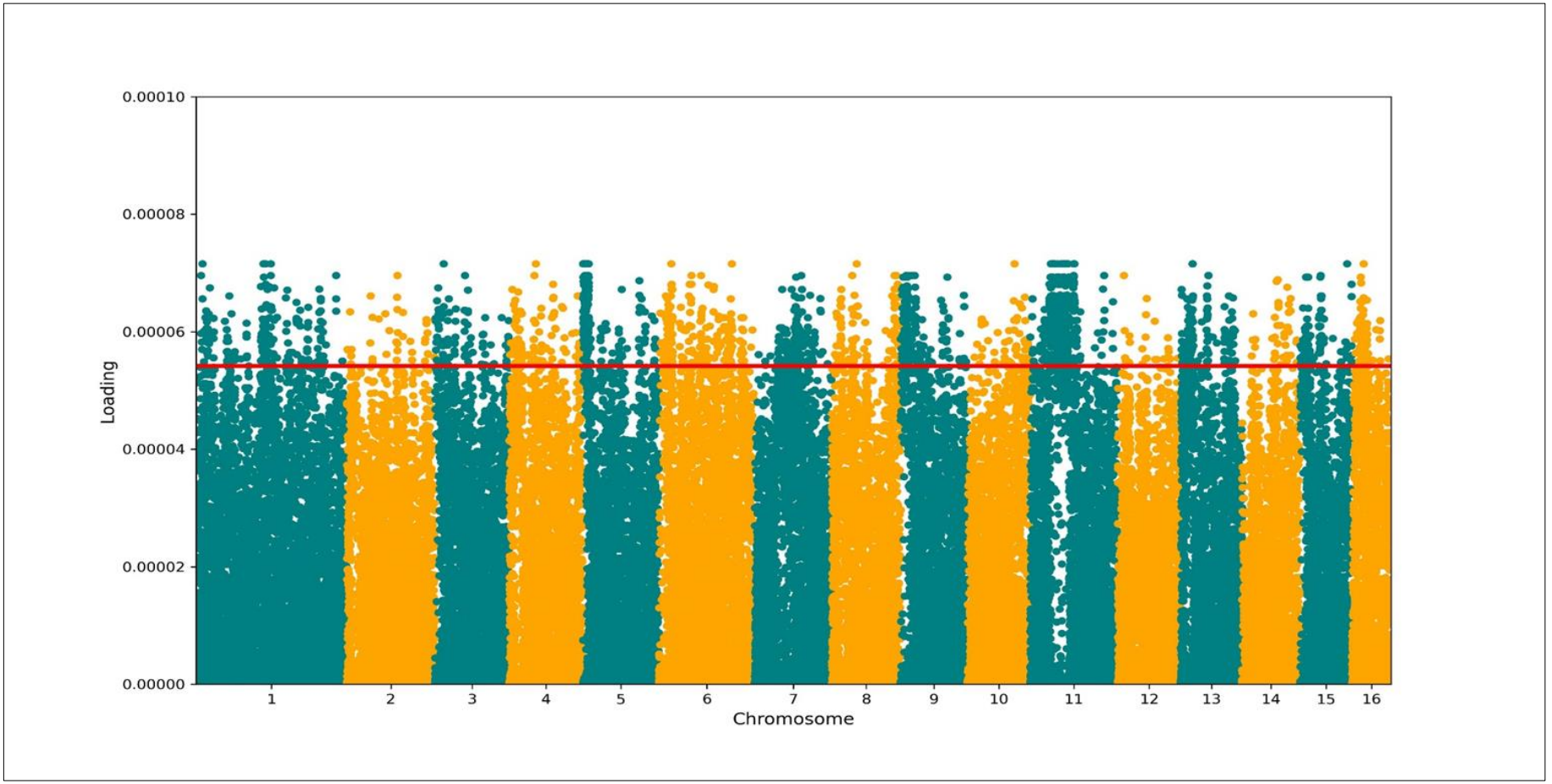


Figure 5.6 Genome wide visualisation of the contributions of all 208, 587 SNPs to the structure seen in the Discriminant Analysis of principal components. The x axis represents the position of each SNP on the chromosomes and the y-axis the loading contribution for each SNP. The red line indicates the 99th percentile.

Table 5.3 SnpEff summary of SNP impacts from the top 1% of SNPs contributing to the discrimination between C lineage and *A. m. mellifera* honey bees

SnpEff summary		
SNPs causing effects	Predicted impact	Number of effects
124	Low	354
27	Moderate	97
1935	Modifier	8526

Table 5.4 Gene functional classification analysis performed in the DAVID online interface from official gene IDs of the top 1% of functional SNPs derived from a DAPC and annotated by SnpEff

Gene Functional Classification Result	
Official Gene Symbol	Gene Name
Gene Group 1	Enrichment Score: 1.37
LOC552814	activating signal cointegrator 1 complex subunit 3
LOC413341	helicase domino
LOC726235	lymphoid-specific helicase
LOC413690	helicase SKI2W
LOC410301	ATP-dependent DNA helicase Q4
Gene Group 2	Enrichment Score: 0.39
LOC725959	serine/arginine repetitive matrix protein 2
LOC102654858	zinc finger and BTB domain-containing protein 41
LOC107964024	hybrid signal transduction histidine kinase L-like
LOC409268	uncharacterized LOC409268
LOC410301	ATP-dependent DNA helicase Q4
LOC725248	dedicator of cytokinesis protein 1
LOC102654286	trichohyalin-like
LOC551016	midasin
LOC410204	titin
LOC727639	uncharacterized LOC727639
LOC410953	MLX-interacting protein
LOC412408	nuclear export mediator factor NEMF homolog
LOC724622	DNA-binding protein RFX7
LOC410237	uncharacterized LOC410237
LOC409665	LIM domain and actin-binding protein 1
LOC551558	trichohyalin
LOC551883	voltage-dependent T-type calcium channel subunit alpha-1G
LOC727052	mucin-17-like
LOC412665	zinc finger protein Pegasus
LOC727236	uncharacterized protein CG5098
LOC412617	lipoma-preferred partner homolog
LOC411535	rho GTPase-activating protein 44
LOC410688	RNA-binding protein 5
LOC725681	uncharacterized

Gene Group 3	Enrichment Score: 0.184
LOC100578352	ionotropic receptor 75a-like
LOC409931	protein FAM69C
LOC724831	uncharacterized LOC724831
LOC100577501	nuclear pore membrane glycoprotein 210
LOC726562	phosphoinositide 3-kinase regulatory subunit 4
LOC552002	uncharacterized
LOC551485	uncharacterized
LOC724680	uncharacterized
LOC410719	syntaxin-17
LOC727639	uncharacterized
LOC727592	uncharacterized
LOC412925	D-galactonate transporter
LOC408646	monocarboxylate transporter 12
LOC414051	ATP-binding cassette sub-family G member 1
LOC551618	prostatic acid phosphatase
LOC412215	chitin synthase chs-2
Abscam	Dscam family member AbsCAM
LOC412401	lipophorin receptor
LOC412536	leukocyte elastase inhibitor
LOC552313	sterol O-acyltransferase 1
LOC551137	proton-associated sugar transporter A
LOC412391	transient receptor potential cation channel, subfamily M, member 3
LOC410134	blood vessel epicardial substance
LOC726958	tumor necrosis factor receptor superfamily member wengen
LOC412011	muscarinic acetylcholine receptor gar-2
LOC551883	voltage-dependent T-type calcium channel subunit alpha-1G
LOC100577135	protein amnionless

5.4 Discussion

This chapter assessed subspecies for a project aiming to monitor subspecies traits differences using genetic assignment methods and examined SNPs responsible of subspecies differentiation.

Some of the subspecies standards used in the INTERREG project are introgressed. This is an important factor for the INTERREG project and any future work will be able to now account for this when analysing the traits data resulting from these colonies. Nordstrom *et al.,* (2021) published a study examining subspecies differences in winter weight loss and over winter consumption of stores using the Swedish colonies. The study found only a marginal difference between subspecies. The results in this chapter show that within the Sweden apiary experiment they have two *A. m. carnica* colonies that have introgression Q values of >0.4 and three *A. m. mellifera* samples with introgression Q values >0.2. This could be an important factor in their results and highlights the need for genetic testing in studies where comparing traits between distinct groups is the goal.

The ADMIXTURE and PCA analysis showed corresponding results regarding introgressed samples. The *A. m. mellifera* samples that were identified as having C – lineage introgression, Norway D Mell 155, Norway H Mel 147 and Switzerland Mel 17 were all seen to be lower down the PC1 axis. In contrast, the *A. m. carnica* samples that presented with *A. m. mellifera* introgression in the ADMIXTURE analysis, Norway D Car 139, Germany Car 4, Norway D Car 138 and Norway H Car 163, were further up the PC1 axis towards the *A. m. mellifera* group in the PCA.

The *A. m. mellifera* introgression in the *A. m. carnica* samples is an interesting result as 3 out of the 4 were from Norway where *A. m. mellifera* has some protection and

there are government breeding programmes for *A. m. carcnia* and *A. m. mellifera*. This could be of concern to Norway's conservation programme. Groeneveld *et al.* (2020) examined Nordic honey bee breeding stocks using SNP data and discovered that often honey bee breeders falsely categorise colonies as introgressed and introgressed colonies as pure when using wing morphometry. This result could be the result of miscategorising colonies within breeding programmes.

The GO analysis results from the SNPs that contribute to the differentiation between *A. m. mellifera* and C lineage honey bees did not yield significant result in clustering terms or functions. This could be for a number of reasons. Firstly, the quality of the annotation even though the honey bee is a relatively well study organism, protein annotations of specific genes still contain many hypothetical protein annotation and predicted a protein to be 'like' another based on amino acid sequence. Moreover, many gene annotations were uncharacterised and are therefore not informative in a gene ontology clustering analyses. Secondly, these types of GO analysis are often employed in differential gene expression data (Sherman *et al.,* 2022; Huang *et al.,* 2009), or using many more SNPs, while here it was used on a limited number of SNP to obtain some biological insight into the SNP that had the greatest effect on population structure analysis. Nonetheless, even though these results are not significant, they do merit some discussion.

The gene functional class in the most enriched group that differentiate the subspecies contained helicases and ATP dependant helicases.

Helicases are proteins that separate nucleic acid strands and are involved in replication and translation (Gorbalenya and Koonin 1993). Helicases have previously been shown to be enriched in eukaryotes in response to cold stress (Ronges *et al.,* 2012; Lipaeva *et al.,* 2021). ATPase associated with a diverse range of activities but is considered important in cell division and has been seen to be involved in in extreme

cold tolerance in fish (Soyano and Mushirobira, 2018) and, salt and cold stress in plants (Jou *et al.,* 2006; Wang *et al.,* 2012).

It is interesting that these proteins all be implicated in types of stress response as *A. m. mellifera* is often stated by beekeepers to be a more "hardy" bee less prone to suffer in bad weather, more likely to survive harsher winters. It is important to state again that these were not significantly enriched GO terms or functions and that many of these are broad terms implicated in a wide range of functions and the best way to establish if *Apis mellifera mellifera* does possess traits such as cold tolerance is to perform an experiment measuring those traits.

Experiments that measure traits are the best way to directly assess species or subspecies trait differences, yet if those species or subspecies are difficult to tell apart, genetic evaluation is a powerful tool to assure accurate experimental set up. Once the INTERREG project has processed and analysed the phenotypic data there will be further analysis of this genetic data which will provide insight into the functional genetic regions that underlay subspecies phenotypic trait differences.

6. General Discussion

6.1 Introduction

The integrity and identity of species and/or populations is a complex and important issue in evolutionary ecology (Abbott *et al.,* 2016). The integrity of species, subspecies or populations can be eroded via hybridisation or introgression (Ortiz-Barrientos and Baack 2014; Goulet *et al.,* 2017). This can occur when previously separate populations come into contact. If one of these populations is locally adapted and has unique combinations of alleles, the mating between the two groups will result in the loss of those co-adapted allele complexes (Rhymer and Simberloff, 1996; Lawson, *et al.,* 2017). Effectively, there may be a loss of the allele combinations that made these populations or groups different. Conversely, the maintenance of species requires gene flow to distribute genetic diversity among populations (Gailing and Curtu 2014; Larson *et al.,* 2014). In this respect gene flow can potentially lead to adaptation via the augmented genetic diversity it provides (Hendrick 2013; Barbato *et al.,* 2020). The subject of the species or population integrity and identity is important when it comes to the management of genetic resources. However, when genetic integrity is eroded, at what level is local adaptation broken down? Does the integrity of a subspecies, species or population matter if hybridisation and introgression increase diversity and subsequently adaptive potential?

This thesis considers these issues via the investigation of the honey bee subspecies *Apis mellifera mellifera* across the UK and Ireland.

The aims in this thesis were to:

(1) Assess the effectiveness of an *A. m. mellifera* breeding program in the south west of England.

(2) Compare three different approaches for methods for measuring introgression.

(3) Investigate the current extent of *A. m. mellifera* across UK and Ireland outside of specific conservation and breeding programs using whole genome data.

(4) In collaboration with researchers in Scandinavia running a phenotypic monitoring experiment, investigate subspecies purity and the underlying genetics differences between the subspecies using whole genome data

These aims were achieved by:

(1) Assessing temporal changes in introgression levels in apiaries involved in a breeding and conservation program using microsatellite markers in the south west of England (Chapter 2).
(2) Comparing introgression results from individual and colony level approaches. Introgression in *A. m. mellifera* was assessed using the SNP array and a reduced representation genome method (RADseq). SNP Array and RADseq genotyping was performed on an individual honey bee and RADseq genotyping was performed on pooled colonies (Chapter 3).
(3) Putative *A. m. mellifera* colonies across the UK and Ireland were examined using whole genome data (Chapter 4).
(4) Subspecies assignment of colonies involved in a phenotypic monitoring project were examined using whole genome data as well as an examination of subspecies differences (Chapter 5).

However, broader questions were posed in the general introduction of this thesis about the interaction of local adaptation of introgression and hybridisation. During the course of this thesis attempts were made to assess the local adaptation of *A. m. mellifera*. To comprehensively assess local adaptation, genotypic, phenotypic and local environmental information from a large number of individuals is required (Sork *et al.,* 2013; Garcia-Navas *et al.,* 2014). There were several attempts to gather this information from across the UK, unfortunately, none were successful. However, the collaboration with INTERREG bee research project (chapter 5) will allow for an examination of local adaptation in *A. m. mellifera* but the timings of the INTERREG analysis were out of the control of this project.

6.2 Key findings

The key findings of this thesis are:

(1) Chapter 2 indicates that bee keepers can have a positive impact in reducing introgression when working together as part of a conservation and breeding programme and using isolated mating locations.

(2) Chapter 3 suggest that individual introgression estimates can be considerably different to colony level introgression estimates.

(3) Chapter 4 revealed that much of the UK and Ireland is introgressed but a few pockets of *A. m. mellifera* remain.

(4) Chapter 5 highlighted the need for the assessment of colonies involved in a phenotypic monitoring experiment and implicated SNPs involved in subspecies differentiation.

6.3 Implications of key findings

The next sections discuss the implications of the key findings to the wider bee keeping, apicultural research and policy making communities.

6.3.1 Conservation and breeding

The results from chapter 2, showing that introgression can be reduced though honey bee conservation programs, are in line with previous studies. It has been observed in other conservation programs that the use of isolated mating areas, where selected breeding colonies are mated without the presence of C lineage honey bees, results in lower introgression levels than those bred in unprotected areas (Pinto *et al.,* 2014). There has been particular success using isolated mating apiaries on the Rame Peninsula in Cornwall. The Rame Peninsula population has seen little introgression in this study and previously (Ellis *et al.,* 2018). However, there are challenges involved with getting bee keepers to combine their efforts. Bee keeping groups can be fractious when opinions differ on how to proceed. For example, in the South West of England there are three *A. m. mellifera* conservation groups, The Bee Improvement Program for Cornwall (BIPCo), the B4 Project and the Cornwall Bee Improvement and Bee Breeders Group (CBIBBG). These groups have some members in common and have largely the same goal, to breed and preserve *A. m. mellifera*. However, while these groups have engaged in financial collaboration to pay for genetic testing (B4 personal communication); there has been no protected mating apiary set up for the whole group to use (though some individuals do have access to isolated land for breeding) and there has been no formal queen rearing program. This lack apparent of cooperation between bee keepers impedes conservation goals. The lack of queen rearing is a major factor in UK bee keeping (personal communication B4 project), not only does it limit breeding programmes, but it is necessitating the continued

importation of honey bees in the UK. In recent years there has been an increase in the popularity of bee keeping and importation continues as the demand for queens out stretches the supply in the UK. Lodesani and Costa (2003) studied queen rearing programmes across Europe and found that Austria had 275 registered queen breeders, Germany 270 queen breeders, Italy had 24 registered queen breeders most of which are large scale professional breeders. However, the UK has no large scale queen rearing and at this time consists of approximately 10 bee keepers producing 2000 queens a year. Queen production in the UK is not regulated or centrally organised and the only monitoring is regarding queen importation (Lodesani and Costa 2003). Queen rearing should be a goal of conservation and breeding programmes. Moreover, these conservation and breeding programs require an initial assessment of stock. However, the genetic methods used to assess that stock may require review (chapter 3).

6.3.2 Genetic approaches

Chapter 3 indicates that individual introgression estimates can be considerably different to introgression estimates conducted using colony level allele frequencies (chapter 3) and that the genetic methods being used to assess honey bees requires assessment. The SNP array, was designed for use by conservation and breeding programs, government agencies, and research facilities (Pinto *et al.,* 2015; Munoz *et al.,* 2015; Henriques *et al.,* 2018). Additionally, the SNP array was rigorously tested and designed to produce ADMIXTURE results close to that of whole genome for an individual honey bee (Henriques *et al.,* 2018; Alexander *et al.,* 2015). Which was further confirmed in chapter 3, where the individual RADseq and the SNP array results were highly correlated. While this thesis was underway examining differences between colony and individual introgression estimates (chapter 3), the SNP array has become commercially available to bee keepers (via companies such as Apigenix and

beebytes). The cost of SNP array ranges from approximately £49 to €120 per sample. One of these services uses up to 15 drones to assess each colony. Importantly, this approach will inform the bee keeper about the queen's genotype (drones develop from unfertilized eggs); and this method is effectively a maternity test. Testing drones that are laid by the queen does not carry information about patrilines present within the colony. Thus if the queen is *A. m. mellifera* but has mated with multiple C lineage drones, this testing will not reveal that. In this case, test results will inform the bee keeper about the alleles that colony will spread in to the surrounding population via the drones, but not inform a breeding program about which colonies to breed from. Specifically, it does not collect information about the introgression of subsequent queens reared from that colony. Bee keepers can make a difference if they work together but the genetic testing must be informative for the questions they are asking (chapter 3). There is a significant difference between asking "is this queen *A. m. mellifera*?" or "is this colony *A. m. mellifera*?" and it is important that beekeepers and researchers a like understand and acknowledge this difference. Knowing which colonies to breed from is important, again, because there are a lack of queens available generally but this especially true of *A. m. mellifera*.

6.3.3 The effects of importation on UK honey bee populations

Chapter 4 indicated that much of the British and Irish honey bee population is introgressed (chapter 4) and only a few *A. m. mellifera* colonies remain. The findings from this chapter were in line with previous studies indicating that Ireland (Hassett *et al.,* 2018; Browne *et al.,* 2020) the inner Hebrides (Jensen *et al.,* 2005; Munoz *et al.,* 2015, Regan *et al.,* 2018; Pinto *et al.,* 2014) and the South West of England (Ellis *et al.,* 2018) contain *A. m. mellifera* populations. Honey bees, unlike fully domesticated livestock, breed openly with the population surrounding them and this largely introgressed population is the result of repeated importation. The UK imports on

average 15,000 foreign honey bee queens a year (Beebase, 2021, BeeBase 2022). Considering this widespread loss of genetic integrity across the UK and Ireland honey bee population, it could be argued that there is no reason to cease imports as there nothing to save. Furthermore, some bee keepers are in favour of importation so why not let them continue the practice (Government Petition, 2021; PA media, 2021). There are two reasons that importation is still detrimental even after much of the genetic integrity has been eroded. Firstly, high level of ongoing gene flow can disrupt local adaptation (Kawecki and Ebert, 2004; Blanquart, *et al.,* 2013) and secondly, the threat posed by disease and pests.

Gene flow disrupts local adaptation unless there is a strong selection pressure (Yeaman and Otto, 2011; Tigano and Friesen, 2016). For example, fish stocking (where fish are raised in a hatchery and released into the wild) was originally viewed as a way to prevent population declines caused by over fishing, habitat loss and disease (Waples and Hendry 2008). However, this repeated introduction of farmed fish has been seen to hinder not aid the recovery of wild populations (Araki *et al.,* 2008). Wild populations are different to their farmed counterparts behaviourally, morphologically and genetically as a result of the different environments they experience (Hindar *et al.,* 1991; Heath *et al.,* 2003; Huntingford 2004). When these two different groups interbreed it has resulted in a reduction in genetic differentiation between the two groups (Araguas *et al.,* 2004; Hansen *et al.,* 2006; Karaiskou *et al.,* 2009). Additionally, it can lead to a loss of local adaptation in the wild population via the loss of allele combinations that have been selected for by the wild as farmed fish are introduced repeatedly (Araki *et al.,* 2008; Hansen *et al.,* 2009; Bourret *et al.,* 2011). Decades of science have shown that while stocking increases the overall numbers, the long term effects of stocking are detrimental to wild populations and does not create a self-sustaining population (Claussen and Philipp 2022).

The repeated importation of honey bees has parallels with fish stocking. There is a perception that honey bee importation will prevent population declines. A petition circulated during Brexit by bee keepers that support imports stated "You have the chance to stop the losses and allow imports to continue" (Government Petition, 2021). Additionally, honey bees across Europe are shaped by a different environments. Chapter 5 indicated regions of the genome that contribute to the differences between subspecies. This finding is in line with previous studies that have observed morphological and genetic differences between *A. m. ligustica* and *A. m. carnica* and *A. m. mellifera* (Ruttner 1988; Franck *et al.,* 2000; Dall'Olio, *et al.,* 2007; Rortais, *et al.,* 2011; Pinto, *et al*., 2014). The continued influx of alleles from other subspecies prevents and interrupts the build-up of allele combinations for local adaptation.

Additionally, importation increases the threat posed by disease and pests. The UK has a history of importing taxa that have led to a number of ecological disasters. For example, in the 1960s Dutch Elm Disease (*Ophiostoma ulmi*) was imported to the UK (Brasier 1979). The beetle borne fungus is responsible for the extensive loss of elm habitat via the death of approximately 30 million elm trees in the UK (Harwood *et al.,* 2011; Potter *et al.,* 2011). Approximately 40 years later as Ash dieback (*Chalara fraxinea*) spread throughout Europe the UK continued to import Ash (*Fraxinus* species) saplings from Germany and the Netherlands (Woodward and Boa 2013; Enderle *et al.,* 2019). Ash dieback arrived in the UK in 2012 and it is estimated that it could result in the deaths of approximately 80% of Ash trees (Mitchell *et al.,* 2016; Corker *et al.,* 2019). Ash has been described as a keystone species (Pautasso *et al.,* 2013), a species that maintains the organisation, stability, and function of their community and has a disproportionally large impact on their ecosystem (Kotliar 2000; Delibes-Mateos *et al.,* 2011). As a result, there are concerns about an

extinction cascade caused by the large scale losses (Hultberg *et al.,* 2020). Effort to stop the spread of these diseases by the Forestry Commission, The UK Ministry of Agriculture, Fisheries and Food, and local Councils has been described by scientists as 'too little too late' (Tomlinson and Potter 2010) as in both cases these threats were known about and trade continued regardless. Smith *et al* (2007) examined invertebrate plant pests bought to the UK between 1970 and 2004, a total of 164 were recorded and 114 of those were the result of human mediated transport (not a natural migration). Additionally, this list was seen to be conservative as it didn't include poorly studies groups for example, Nematoda or Acari. Recent arrivals also include, the oak processionary moth (*Thaumetopoea processionea)* which feeds on oak trees and again this insect was spreading across Europe (Groenen *et al.,* 2012) when it was introduced to the UK via imported of oak trees in 2006 (Townsend 2008). The Asian longhorn beetle (*Anoplophora glabripennis*) a pest of deciduous trees, particularly sycamore (*Acer pseudoplatanus*), was also imported from China into the UK in 2012 (Straw *et al.,* 2015).

The continued importation of honey bees brings with it the risk of unwanted species being transported into the UK. Currently there is a new threat posed by *Aethina tumida*, the small hive beetle (Sabella *et al.,* 2022). This parasitic beetle has been seen to cause fermentation in the honey via deposited faecal matter and feeds on honey, pollen and larvae (Ellis and Hepburn 2006). This parasite has spread to the USA, Canada, Australia, and Europe (Animal Health Australia 2003; Dubuc 2013; Numa-Vergel *et al.,* 2021). In Europe it is currently present in Portugal and Italy (Cepero *et al.,* 2014; Palmeri *et al.,* 2015). Italy is one of the main locations that the UK imports from (BeeBase 2021; BeeBase 2022). A response from the UK government regarding concerns about small hive beetle was that "This invasive pest has only been detected in one part of Europe, namely southern Italy, and exports of

bees from the affected region into either Great Britain or Northern Ireland are not permitted." (Government response, 2021). Importantly, *V. destructor* was bought to the UK from Europe in 1992 (De Jong *et al.,* 1982; Büchler *et al.,* 2010). Its arrival in the UK, most likely from the importation of infested colonies, occurred while there was a ban on the importation from affected regions (The National Bee Unit Managing *Varroa*, 2020). The regulations around importation of species are poor, as is the United Kingdom's track record of preventing ecologically devastating incidents.

There is a growing movement to limit importations into *A. m. mellifera* populations. The Isle of Colonsay is currently the only legally protected *A. m. mellifera* population in UK or Ireland. However, in Ireland where a substantial population of *A. m. mellifera* resides (Hasset *et al.,* 2018; Browne *et al.,* 2020), steps towards banning importation have begun. A bill to ban importation into Ireland has been introduced and debated by the Irish senate (Protection of Native Irish Honey Bee Bill, 2021). Limiting importation in Ireland will safeguard important genetic resources, while limiting importation in the UK will encourage local adaptation and prevent the spread of disease and pests.

6.3.4 Informing introgression thresholds

Another question raised by this thesis is when does *A. m. mellifera* stops being *A. m. mellifera*. At what threshold of introgression are traits disrupted? Given that honey bees have a high recombination rate it could be hypothesised that the breaking up of co-adapted alleles would happen more rapidly than in species with lower recombination rates. Studies have observed that introgression rates are affected by recombination rates (Janoušek *et al.,* 2015; Martin *et al.,* 2019). But what does that mean in terms of introgression proportions in honey bees? If an *A. m. mellifera* colony has an introgression proportion of 0.2, does it have any *A. m. mellifera* traits?

Answering this question requires large scale phenotypic monitoring paired with genetic sampling. While chapter 5 beings to assess the colonies in the INTRREG project, future work from this collaboration will begin to answer questions surrounding introgression thresholds in the future.

6.3.5 The conservation of honey bees

Finally there is a topic that is relevant to this body of work, and that is whether conservation of honey bees is necessary. Recently, there has been tense debate about honey bees and their impact on wild pollinators (Geldmann and González-Varo 2018; Kleijin *et al.,* 2018; Saunders *et al.,* 2018) and given that there is in excess of 260,000 colonies in the UK (National Bee Unit Hive count, 2022) is the conservation of this species important? Studies have shown that high densities of honey bee colonies can generate foraging competition between wild pollinators and honey bees (Mallinger *et al.,* 2017; Geslin *et al.,* 2017; Ropars *et al.,* 2019). The general pattern found by studies is that high colony densities do have negative impacts on wild bees. However, this competition has been observed to cause a reduction of foraging success for both wild pollinators and honey bees (Henry and Rodet 2018). Essentially, there are no winners in high density bee keeping and it seems logical that there is a carrying capacity for pollinators in an area and that limiting the honey bee density in areas could be necessary (Alaux *et al.,* 2019; Henry and Rodet 2020). However, the definition of high density for different landscapes is difficult to clarify. Attempts have been made, and estimates of 3.1 colonies per km^2 to 3.5 colonies per km^2, were suggested from experiments using rosemary and tyme plants in scrubland in the Garraf Natural Park in Spain (Torné-Noguera *et al.,* 2016). A study on agricultural land in Germany suggested colony densities of 1 colony per 1.5-2km^2 (Steffan-Dewenter and Tscharntke 2000). There has been suggestion that an 'Apiary Influence Range' could be used to help inform colony densities (Henry and Rodet

2020) and aid land managers in assessing the ideal number of colonies per km^2 (Henry and Rodet 2020). The Apiary Influence Range is the distance at which the competition induced by an apiary is lessened beyond a certain distance, and has been suggested to be 0.6 – 1.1km. However, this experiment was based in France and estimates were performed assuming using 30-50 colonies per apiary (Henry and Rodet 2020), which might not be applicable to the UK where the majority of bee keepers are hobbyists and the average number of colonies per apiary is 4 (National Bee Unit Husbandry Survey, 2017). Overall, these estimates have a long way to go in terms of building up well informed guidance for bee keepers and land owners. Regardless, some conservationists call for the complete ban of honey bees in protected areas (Geldmann and González-Varo 2018).

This debate has seeped into the main stream with multiple articles, for example, 'Honey bees are voracious; is it time to put the brakes on the boom in bee keeping' (Turner, 2021), or 'Conservation of honey bees may lead to food shortages' (Heap, 2022). As this debate gathers pace it is already bringing consequences for bee keepers. Specifically, those many bee keepers who do not own the land they house their bees on can be vulnerable to land management policies dictated by land owners and in some cases this has left bee keepers struggling to find suitable locations for apiaries (Durant, 2019).

Honey bees do pose a risk to other pollinators where they have been introduced outside of their native range not just via foraging competition but by replacement (Russo 2016; Ollerton 2017). This is especially true in Asia where *Apis mellifera* has become more popular than the native *Apis cerana* (Theisen-Jones and Bienefeld 2016), and in Mexico where honey bees are preferred to the native *Melipona beechii* due to their high productivity (Quezada-Euán 2018).

Tensions between wild bee conservationists and *Apis mellifera* conservationists are largely unhelpful. Focusing on only the negative impacts of honey bees could have counterproductive consequence for crop production (Kleijin *et al.,* 2018; Potts *et al.,* 2016) as honey bees an integral part of pollinating services in the UK and Ireland (POST, 2010; Breeze *et al.,* 2011;) and globally (Gallai *et al.,* 2009). The best approach to reconciling this debate is to use an inclusive approach where stakeholders are well informed (Kleijin *et al.,* 2018). Importantly wild pollinators does not necessarily exclude honey bees, the extent of wild honey bee colonies in the UK remains largely unknown. It is important to note that there are a variety of the issues facing wild pollinators some of which can be addressed while data are collected on appropriate honey bee colony densities. Impacts such as, on the loss of nesting sites, loss of abundance and diversity in floral resources, long term exposure to pesticides, pollution, introduction of parasites and pest and climate change (Potts *et al.,* 2016; Hallmann *et al.,* 2017; Rhodes 2018).

6.4 Conclusion and recommendations

The goal of this thesis was to assess honey bee subspecies diversity and adaptation in their natural range in the UK (Requier *et al.,* 2019). The recommendations from this work are to prioritise limiting importation, promoting the use of local subspecies and to manage honey bees alongside wild pollinators to the benefit of both groups.

6.4.1 Applied recommendations

There are several ways in which the situation for *A. m. mellifera* and honey bees more generally could be improved in the UK.

(1) Queen rearing programs.

Queen rearing would not only improve breeding and conservation programs for *A. m. mellifera* but more broadly bee keepers would not have to rely on imports to fulfil demand generated by the popularity of bee keeping.

(2) Isolated mating sites and protected areas for *A. m. mellifera*.

In *A. m. mellifera* breeding programs and refuges; isolated mating apiaries and protected areas would benefit breeding and conservation programs and reduce introgression from other subspecies.

(3) The restriction of imports

The restriction of imports will encourage local adaptation in the UK and reduce the risk of disease and pests entering the country.

(4) Bee keeper synergy

It would be highly beneficial for bee keeping groups to work together to achieve goals such as queen rearing, shared isolated mating apiary sites, and *A. m. mellifera* reservations.

6.4.2 Research recommendations

The following are suggestions for future research into *A. m. mellifera:*

(1) Reassess genetic testing with a view to informing breeding programs

It is important that genetic testing informs bee keepers about breeding colonies. Colony pooling is the most accurate way to assess a colony. Research should use this method when assessing breeding colonies. If commercial companies assess one bee or only sample drones they need to be clear in their communication to bee keepers about what the information tells them.

(2) Investigate the natural range of *A. m. mellifera* for remnant populations.

There are many areas still left unexamined and the wider European picture for *A. m. mellifera* is still largely unknown. Until 2018 there was no empirical evidence that Ireland had a large population of *A. m. mellifera.* Therefore, it is not unfeasible that other substantial locations of *A. m. mellifera* exist. The discovery of new locations is required so that protection can be put in place to prevent further loss of genetic resources.

(3) Investigate local adaptation.

Empirical data is still lacking about the adaptation of *A. m. mellifera*. While there is evidence of morphological and genetic differences, confirmation of their differentiating traits are still needed. Future work resulting from this project in collaboration with the INTERREG project is a promising starting point.

(4) Inform purity thresholds for *A. m. mellifera*

Research to inform 'purity' thresholds is lacking. Work is needed to inform the level of introgression at which traits are lost. This work could help inform breeding colony

choice. The collaboration with the INTERREG project may be the first steps towards informing these thresholds.

7. References

Abbott, R.J., Barton, N.H. and Good, J.M. 2016, Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, 25: 2325-2332.

Adam, B. 1968. "Isle of Wight" or Acariñe Disease: its Historical and Practical Aspects', *Bee World*. 49(1), 6-18.

Aizen, M.A. and Harder, L.D., 2009. The global stock of domesticated honey bees is growing slower than agricultural demand for pollination. *Current Biology*, 19(11), 915-918.

Alaux, C., Le Conte, Y. and Decourtye, A., 2019. Pitting wild bees against managed honey bees in their native range, a losing strategy for the conservation of honey bee biodiversity. *Frontiers in Ecology and Evolution*, (7), 60.

Alburaki, M., Moulin, S., Legout, H., Alburaki, A. and Garnery, L., 2011. Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq. *Apidologie*, *42*(5), 628-641.

Alcala, N., Streit, D., Goudet, J. and Vuilleumier, S., 2013. Peak and persistent excess of genetic diversity following an abrupt migration increase. *Genetics*, 193(3): 953-971.

Alexander, D.H., Novembre, J. and Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9): 1655-1664.

Alexander, D.H., Shringarpure, S.S., Novembre, J. and Lange, K., 2015. Admixture 1.3 software manual. Los Angeles: UCLA Human *Genetics Software Distribution*.

Alexander, M., Ho, S.Y., Molak, M., Barnett, R., Carlborg, Ö., Dorshorst, B., Honaker, C., Besnier, F., Wahlberg, P., Dobney, K. and Siegel, P., 2015. Mitogenomic analysis of a 50-generation chicken pedigree reveals a rapid rate of mitochondrial evolution and evidence for paternal mtDNA inheritance. *Biology Letters,* 11(10).

Allendorf, F.W., Leary, R.F., Spruell, P. and Wenburg, J.K., 2001. The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, 16(11): 613-622.

Alliance of Genome Resources Consortium, 2019. The alliance of genome resources: building a modern data ecosystem for model organism databases. *Genetics*, 213(4):1189-1196.

Amkreutz, L. and Spithoven, M., 2019. Hunting Beneath the Waves: Bone and Antler Points from the North Sea Doggerland Off the Dutch Coast. *Wachholtz Verlag*. 31-44

Anderson, E. 1949. Introgressive Hybridization. New York, NY: John Wiley and Sons. 109

Anderson, E.C. and Dunham, K.K., 2008. The influence of family groups on inferences made with the program Structure. *Molecular Ecology Resources*, 8(6), 1219-1229.

Animal Health Australia. 2003. Small Hive Beetle National Management Plan. Animal Health Australia Council Ltd. 17, Deakin, Australia. Available at: https://honeybee.org.au/pdf/SHB_FINAL_Mgt_Plan_AHA%20_31-10-03_.pdf Last Accessed 27th July 2022

Araguas, R.M., Sanz, N., Pla, C. and Garcia-Marin, J.L., 2004. Breakdown of the brown trout evolutionary history due to hybridization between native and cultivated fish. *Journal of Fish Biology*, 65:.28-37.

Araki, H., Berejikian, B.A., Ford, M.J. and Blouin, M.S., 2008. Fitness of hatchery-reared salmonids in the wild. *Evolutionary Applications*, 1(2), 342-355.

Arias, M. C. and Sheppard, W. S. 1996 Molecular Phylogenetics of Honey Bee Subspecies (*Apis mellifera* L.) Inferred from Mitochondrial DNA Sequence, *Molecular Phylogenetics and Evolution*, 5(3), 557–566.

Arundel, J., Oldroyd, B.P. and Winter, S., 2012. Modelling honey bee queen mating as a measure of feral colony density. *Ecological Modelling*, 247, 48-57.

Avalos, A., Fang, M., Pan, H., Ramirez Lluch, A., Lipka, A.E., Zhao, S.D., Giray, T., Robinson, G.E., Zhang, G. and Hudson, M.E., 2020. Genomic regions influencing aggressive behavior in honey bees are defined by colony allele frequencies. *Proceedings of the National Academy of Sciences*, 117(29), 17135-17141.

Bachtrog, D., Thornton, K., Clark, A. and Andolfatto, P., 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution*, 60(2), 292-302.

Bahbahani, H., Afana, A. and Wragg, D., 2018. Genomic signatures of adaptive introgression and environmental adaptation in the Sheko cattle of southwest Ethiopia. *PloS one*, 13(8).

Bailey, L. 1964 The “Isle of Wight Disease”: The Origin and Significance of the Myth, *Bee World. Taylor & Francis*, 45(1), 32–37

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): 3376.

Baiser, B., Olden, J.D., Record, S., Lockwood, J.L. and McKinney, M.L., 2012. Pattern and process of biotic homogenization in the New Pangaea. *Proceedings of the Royal Society B: Biological Sciences*, 279(1748): 4772-4777.

Baker, R.A. and Chmielewski, W., 2003. How old are bees? A look at the fossil record. *Journal of Apicultural Science*, 47(1), 79-86.

Ballard, J.W.O. and Whitlock, M.C., 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), 729-744.

Barbanera, F., Guerrini, M., Khan, A.A., Panayides, P., Hadjigerou, P., Sokos, C., Gombobaatar, S., Samadi, S., Khan, B.Y., Tofanelli, S. and Paoli, G., 2009. Human-mediated introgression of exotic chukar (*Alectoris chukar, Galliformes*) genes from East Asia into native Mediterranean partridges. *Biological Invasions*, 11(2), 333-348.

Barbanera, F., Negro, J.J., Di Giuseppe, G., Bertoncini, F., Cappelli, F. and Dini, F., 2005. Analysis of the genetic structure of red-legged partridge (*Alectoris rufa, Galliformes*) populations by means of mitochondrial DNA and RAPD markers: a study from central Italy. *Biological Conservation*, 122(2), 275-287.

Barbato, M., Hailer, F., Orozco-terWengel, P., Kijas, J., Mereu, P., Cabras, P., Mazza, R., Pirastru, M. and Bruford, M.W., 2017. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Scientific Reports*, 7(1), 1-13.

Barbato, M., Hailer, F., Upadhyay, M., Del Corvo, M., Colli, L., Negrini, R., Kim, E.S., Crooijmans, R.P., Sonstegard, T. and Ajmone-Marsan, P., 2020. Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Scientific Reports*, 10(1), 1-11.

Barilani, M., Bernard-Laurent, A., Mucci, N., Tabarroni, C., Kark, S., Garrido, J.A.P. and Randi, E., 2007. Hybridisation with introduced chukars (*Alectoris chukar*) threatens the gene pool integrity of native rock (*A. graeca*) and red-legged (*A. rufa*) partridge populations. *Biological Conservation*, 137(1), 57-69.

Barrett, R.D. and Schluter, D., 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38-44.

Barton, N.H. and Hewitt, G.M., 1989. Adaptation, speciation and hybrid zones. *Nature*, 341(6242), 497-503.

BBKA press office, 2021, Importing Bees "against best practice for our environment". Available at: https://www.bbka.org.uk/news/importing-bees-against-best-practice-for-our-environment Last accessed: 24th July 2022

BeeBase 2021. The National Bee Unit Third Country Import Report 2021. Available at: https://nationalbeeunit.com/public/BeeDiseases/importReport.cfm?year=2021 last Accessed 19th July 2022.

BeeBase 2022. The National Bee Unit Third Country Import Report 2022. Available at: https://nationalbeeunit.com/public/BeeDiseases/importReport.cfm?year=2022 last Accessed 19th July 2022

Beeler, R.B., Sharples, M.T. and Tripp, E.A., 2020. Introgression among three western North American bilberries (*Vaccinium* section *Myrtillus*). *Systematic Botany*, 45(3), 576-584.

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

Blanquart, F., Kaltz, O., Nuismer, S.L. and Gandon, S., 2013. A practical guide to measuring local adaptation. *Ecology Letters*, 16(9), 1195-1205.

Bogenhagen, D.F., Wang, Y., Shen, E.L. and Kobayashi, R., 2003. Protein components of mitochondrial DNA nucleoids in higher eukaryotes. *Molecular & Cellular Proteomics*, 2(11), 1205-1216.

Bohmanova, J., Misztal, I. and Cole, J.B., 2007. Temperature-humidity indices as indicators of milk production losses due to heat stress. *Journal of Dairy Science*, 90(4), 1947-1956.

Bolger, A. M., Lohse, M. and Usadel, B. 2014 Trimmomatic: A flexible trimmer for Illumina sequence data, *Bioinformatics*, 30(15), 2114–2120.

Bolstad, G.H., Hindar, K., Robertsen, G., Jonsson, B., Sægrov, H., Diserud, O.H., Fiske, P., Jensen, A.J., Urdal, K., Næsje, T.F. and Barlaup, B.T., 2017. Gene flow from domesticated escapes alters the life history of wild Atlantic salmon. *Nature Ecology & Evolution*, 1(5), 1-5.

Bonnet, T., Leblois, R., Rousset, F. and Crochet, P.A., 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, 71(9), 2140-2158.

Bourret, V., O'reilly, P.T., Carr, J.W., Berg, P.R. and Bernatchez, L., 2011. Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity*, 106(3), .500-510.

Bradburd, G.S., Coop, G.M. and Ralph, P.L., 2018. Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1), 33-52.

Branstetter, M.G., Ješovnik, A., Sosa-Calvo, J., Lloyd, M.W., Faircloth, B.C., Brady, S.G. and Schultz, T.R., 2017. Dry habitats were crucibles of domestication in the evolution of agriculture in ants. *Proceedings of the Royal Society B: Biological Sciences,* 284(1852), 20170095.

Brasier, C.M., 1979. Dual origin of recent Dutch elm disease outbreaks in Europe. Nature, 281(5726), 78-80.

Breeze, T.D., Bailey, A.P., Balcombe, K.G. and Potts, S.G., 2011. Pollination services in the UK: how important are honeybees?. *Agriculture, Ecosystems & Environment*, 142(3-4), 137-143.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G. and Bustamante, C.D., 2012. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Human biology*, 84(4), 343.

Brockmann, A., Dietz, D., Spaethe, J. and Tautz, J., 2006. Beyond 9-ODA: sex pheromone communication in the European honey bee *Apis mellifera* L. *Journal of Chemical Ecology*, 32(3), 657-667.

Brodschneider, R. and Crailsheim, K., 2010. Nutrition and health in honey bees. *Apidologie*, 41(3), 278-294.

Browne, K.A., Hassett, J., Geary, M., Moore, E., Henriques, D., Soland-Reckeweg, G., Ferrari, R., Mac Loughlin, E., O'Brien, E., O'Driscoll, S. and Young, P., 2020. Investigation of free-living honey bee colonies in Ireland. *Journal of Apicultural Research*, 60(2), 229-240.

Büchler, R., Berg, S. and Le Conte, Y., 2010. Breeding for resistance to *Varroa destructor* in Europe. *Apidologie*, 41(3), 393-408.

Burgarella, C., Barnaud, A., Kane, N.A., Jankowski, F., Scarcelli, N., Billot, C., Vigouroux, Y. and Berthouly-Salazar, C., 2019. Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Frontiers in Plant Science*, 10, 4.

Bush, W.S. and Moore, J.H., 2012. Genome-wide association studies. *PLoS Computer Biology*, 8(12)

Cassel-Lundhagen, A., Kaňuch, P., Low, M. and Berggren, Å., 2011. Limited gene flow may enhance adaptation to local optima in isolated populations of the Roesel's bush cricket (*Metrioptera roeselii*). *Journal of Evolutionary Biology*, 24(2), 381-390.

Catchen, J.M., Hohenlohe, P.A., Bernatchez, L., Funk, W.C., Andrews, K.R. and Allendorf, F.W., 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, 17(3), 362-365.

Cavallo, A. and Martin, A.C., 2005. Mapping SNPs to protein sequence and structure data. *Bioinformatics*, 21(8), 1443-1450.

Cepero, A., Higes, M., Martínez-Salvador, A., Meana, A. and Martín-Hernández, R., 2014. A two year national surveillance for *Aethina tumida* reflects its absence in Spain. *BMC Research Notes*, 7(1), 1-4.

Cerdá, X. and Retana, J., 2000. Alternative strategies by thermophilic ants to cope with extreme heat: individual versus colony level traits. Oikos, 89(1), 155-163.

Chávez-Galarza, J., Garnery, L., Henriques, D., Neves, C.J., Loucif-Ayad, W., Jonhston, J.S. and Pinto, M.A., 2017. Mitochondrial DNA variation of *Apis mellifera iberiensis*: further insights from a large-scale study using sequence data of the tRNAleu-cox2 intergenic region. *Apidologie*, *48*(4), 533-544.

Chávez-Galarza, J., Henriques, D., Johnston, J.S., Azevedo, J.C., Patton, J.C., Muñoz, I., De la Rúa, P. and Pinto, M.A., 2013. Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, 22(23), 5890-5907.

Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., Liu, S., Lu, H., Tian, S., Li, R. and Shi, W., 2016. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. *Molecular Biology and Evolution*, 33(5), 1337-1348..

Chen, Y., Hehr, C.L., Atkinson-Leadbeater, K., Hocking, J.C. and McFarlane, S., 2007. Targeting of retinal axons requires the metalloproteinase ADAM10. *Journal of Neuroscience*, 27(31), 8448-8456.

Christie, M.R., Marine, M.L., Fox, S.E., French, R.A. and Blouin, M.S., 2016. A single generation of domestication heritably alters the expression of hundreds of genes. *Nature Communications*, 7(1), 1-6.

Chutimanitsakun, Y., Nipper, R.W., Cuesta-Marcos, A., Cistué, L., Corey, A., Filichkina, T., Johnson, E.A. and Hayes, P.M., 2011. Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*, 12(1), 1-13.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80-92.

Claussen, J.E. and Philipp, D.P., Assessing the role of supplementation stocking: A perspective. *Fisheries Management and Ecology*. 00. 1-9

Coker, T.L., Rozsypálek, J., Edwards, A., Harwood, T.P., Butfoy, L. and Buggs, R.J., 2019. Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants, People, Planet.* 1(1), pp.48-58.

Coles, B.J., 1998, January. Doggerland: a speculative survey. *Proceedings of the Prehistoric Society*, 64, 45-81.

Cooper, B.A., 1986. The honeybees of the British Isles. *British Isles Bee Breeders' Association*. 158

Cornuet, J.M., Garnery, L. and Solignac, M., 1991. Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. *Genetics*, *128*(2), 393-403.

Corona, M., Estrada, E. and Zurita, M., 1999. Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee *Apis mellifera*. *Journal of Experimental Biology*, 202(8), 929-938.

Coyle, F.A. and McGarity, A.C., 1991. Two new species of *Nesticus* spiders from the southern Appalachians (*Araneae, Nesticidae*). *Journal of Arachnology*, 161-168.

Crane, E., 2005. The rock art of honey hunters. *Bee world*, 86(1), 11-13.

Crozier, R.H., Crozier, Y.C. and Mackinlay, A.G., 1989. The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Molecular Biology and Evolution*, *6*(4), 399-411.

Culliney, T.W., 1983. Origin and evolutionary history of the honey bees *Apis*. *Bee World*, 64(1), 29-38.

Cunliffe, B., 2002. The extraordinary voyage of Pytheas the Greek: The man who discovered Britain. Bloomsbury Publishing USA.

Currat, M., Ruedi, M., Petit, R.J. and Excoffier, L., 2008. The hidden side of invasions: massive introgression by local genes. *Evolution: International Journal of Organic Evolution*, 62(8), 1908-1920.

Cushman, J.H., Lawton, J.H. and Manly, B.F., 1993. Latitudinal patterns in European ant assemblages: variation in species richness and body size. *Oecologia,* 95(1), 30-37.

Cushman, J.H., Lawton, J.H. and Manly, B.F., 1993. Latitudinal patterns in European ant assemblages: variation in species richness and body size. *Oecologia*, 95(1), 30-37.

Dainat, B., Vanengelsdorp, D. and Neumann, P., 2012. Colony collapse disorder in Europe. *Environmental Microbiology Reports*, 4(1), 123-125.

Dall'Olio, R., Marino, A., Lodesani, M. and Moritz, R.F., 2007. Genetic characterization of Italian honeybees, *Apis mellifera ligustica*, based on microsatellite DNA polymorphisms. *Apidologie*, *38*(2), 207-217.

Dams, L.R., 1978. Bees and honey-hunting scenes in the Mesolithic rock art of eastern Spain. *Bee World*, 59(2), 45-53.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.

Darras, H. and Aron, S., 2015. Introgression of mitochondrial DNA among lineages in a hybridogenetic ant. *Biology Letters*, 11(2)

Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D., Davey, J.W., Whibley, A., Nadeau, N.J., Zimin, A.V., Hughes, D.S., Ferguson, L.C., Martin, S.H. and Salazar, C., 2012.

Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405), 94.

Davey, J.W. and Blaxter, M.L., 2010. RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9(5-6), 416-423.

Dávila, S.G., Cuenca, O.T., Gil, M.G., Toledano-Díaz, O.A., Castaño, C., Santiago-Moreno, J. and Campo, J.L., 2019. Anti-predator behaviour in pure and hybrid red-legged partridges. *Poultry Science*, 98(8), 3085-3092.

de Haro, L., Labadie, M., Chanseau, P., Cabot, C., Blanc-Brisset, I., Penouil, F. and for Toxicovigilance, N.C.C., 2010. Medical consequences of the Asian black hornet (*Vespa velutina*) invasion in Southwestern France. *Toxicon,* 55(2-3), 650-652.

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I. and Serrano, J., 2009. Biodiversity, conservation and current threats to European honeybees. *Apidologie, 40*(3), 263-284.

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I. and Serrano, J., 2009. Biodiversity, conservation and current threats to European honey bees. *Apidologie*, 40(3), 263-284.

De la Rúa, P., Jaffé, R., Muñoz, I., Serrano, J., Moritz, R.F. and Kraus, F.B., 2013. Conserving genetic diversity in the honeybee: Comments on Harpur *et al.*(2012).

DEFRA and National Bee Unit (2018) Guidance On Importing Bees Into England, 2015(May 2015). Available at: https://nationalbeeunit.com/downloadDocument.cfm?id=977#:~:text=Imports%20of%20honey%20bees%20are,than%2020%20attendant%20worker%20bees . Last Accessed: 21st July 2022

DeGrandi-Hoffman, G., Chen, Y., Huang, E. and Huang, M.H., 2010. The effect of diet on protein concentration, hypopharyngeal gland development and virus load in worker honey bees (*Apis mellifera* L.). *Journal of insect physiology*, *56*(9), 1184-1191.

Delibes-Mateos, M., Smith, A.T., Slobodchikoff, C.N. and Swenson, J.E., 2011. The paradox of keystone species persecuted as pests: a call for the conservation of abundant small mammals in their native range. *Biological Conservation*, *144*(5), 1335-1346.

Demant, T., Lind Madsen, T., Kauko, L., Ruottinen, L., Dahle, B., Kirkerud, L. A., Laupstad Vatland, A., Rødsdalen, T.-E., Arvidsson, I. & Groeneveld, L. F., 2019. Second plan of action for the conservation of the Nordic brown bee. 2019:02. NordGen Publication Series. Available at: https://www.norden.org/en/publication/second-plan-action-conservation-nordic-brown-bee Last Accessed: 21st of July 2022

Dempewolf, H., Hodgins, K.A., Rummell, S.E., Ellstrand, N.C. and Rieseberg, L.H., 2012. Reproductive isolation during domestication. *The Plant Cell*, 24(7), 2710-2717.

Des Marais, D.L., Hernandez, K.M. and Juenger, T.E., 2013. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. *Annual Review Ecology Evolution Systematics*, 44(1), 5-29.

Dietemann, V., Nazzi, F., Martin, S.J., Anderson, D.L., Locke, B., Delaplane, K.S., Wauquiez, Q., Tannahill, C., Frey, E., Ziegelmann, B. and Rosenkranz, P., 2013. Standard methods for *varroa* research. *Journal of apicultural research*, 52(1), 1-54.

Díez-del-Molino, D., Carmona-Catot, G., Araguas, R.M., Vidal, O., Sanz, N., García-Berthou, E. and García-Marín, J.L., 2013. Gene flow and maintenance of genetic diversity in invasive mosquitofish (*Gambusia holbrooki*). *PLoS One*, 8(12), 82501.

Dray, S. and Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software,* 22, 1-20.

Dubuc, M. 2013. Small hive beetle infestation (*Aethina tumida*), Canada. OIE report, available at .http://www.oie.int/wahis_2/public/wahid.php/Reviewreport/Review?page_refer= MapFullEventReport&reportid=14742 Last accessed 18th July 2022.

Dumont, B.L. and Payseur, B.A., 2008. Evolution of the genomic rate of recombination in mammals. *International Journal of Organic Evolution*, 62(2), 276-294.

Eckert, C.D., Winston, M.L. and Ydenberg, R.C., 1994. The relationship between population size, amount of brood, and individual foraging behaviour in the honey bee, *Apis mellifera* L. *Oecologia*, 97(2), 248-255.

Eckholm, B.J., Huang, M.H., Anderson, K.E., Mott, B.M. and DeGrandi-Hoffman, G., 2015. Honey bee (*Apis mellifera*) intracolonial genetic diversity influences worker nutritional status. *Apidologie*, *46*(2), 150-163.

Edmands, S., 2007. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, 16(3), 463-475.

Elekonich, M.M. and Roberts, S.P., 2005. Honey bees as a model for understanding mechanisms of life history transitions. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 141(4), 362-371.

El-Kassaby, Y.A., 1992. Domestication and genetic diversity—should we be concerned?. *The Forestry Chronicle*, 68(6), 687-700.

Ellis, J.A. and Ong, B., 2017. The MassARRAY® system for targeted SNP genotyping. In Genotyping, *Humana Press*, 77-94.

Ellis, J.S., Soland-Reckeweg, G., Buswell, V.G., Huml, J.V., Brown, A. and Knight, M.E., 2018. Introgression in native populations of *Apis mellifera mellifera* L: implications for conservation. *Journal of Insect Conservation*, *22*(3), 377-390.

Enciso-Romero, J., Pardo-Díaz, C., Martin, S.H., Arias, C.F., Linares, M., McMillan, W.O., Jiggins, C.D. and Salazar, C., 2017. Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, 26(19), 5160-5172.

Enderle, R., Stenlid, J. and Vasaitis, R., 2019. An overview of ash (*Fraxinus* spp.) and the ash dieback disease in Europe. *CAB Review*, 14, 1-12.

Epifanio, J. and Philipp, D., 2000. Simulating the extinction of parental lineages from introgressive hybridization: the effects of fitness, initial proportions of parental taxa, and mate choice. *Reviews in Fish Biology and Fisheries*, 10(3), 339-354.

Erler, S., Denner, A., Bobiş, O., Forsgren, E. and Moritz, R.F., 2014. Diversity of honey stores and their impact on pathogenic bacteria of the honey bee, *Apis mellifera*. *Ecology and Evolution*, 4(20), 3960-3967.

Estoup, A., Solignac, M. and Cornuet, J.M., 1994. Precise assessment of the number of patrilines and of genetic relatedness in honey bee colonies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 258(1351), 1-7.

Excoffier, L., Laval, G. and Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1.

Fambrough, D., Pan, D., Rubin, G.M. and Goodman, C.S., 1996. The cell surface metalloprotease/disintegrin Kuzbanian is required for axonal extension in *Drosophila*. *Proceedings of the National Academy of Sciences*, *93*(23), 13233-13238.

Fantham, H.B. and Porter, A., 1912. The morphology and life history of *Nosema apis* and the significance of its various stages in the so-called 'Isle of Wight' disease in bees *(Microsporidiosis*). *Annals of Tropical Medicine & Parasitology*, *6*(2), 163-195.

Fedotov, S.A., Besedina, N.G., Bragina, J.V., Danilenkova, L.V., Kamysheva, E.A. and Kamyshev, N.G., 2019. Overexpression of isoform B of Dgp-1 gene enhances locomotor activity in senescent *Drosophila* males and under heat stress. *Journal of Comparative Physiology A*, 205(6), 897-910.

Filippi, C.V., Aguirre, N., Rivas, J.G., Zubrzycki, J., Puebla, A., Cordes, D., Moreno, M.V., Fusari, C.M., Alvarez, D., Heinz, R.A. and Hopp, H.E., 2015. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biology*, 15(1), 1-12.

Finnegan, A.K. and Stevens, J.R., 2008. Assessing the long-term genetic impact of historical stocking events on contemporary populations of Atlantic salmon, Salmo salar. *Fisheries Management and Ecology*, 15(4), 315-326.

Fledel-Alon, A., Wilson, D.J., Broman, K., Wen, X., Ober, C., Coop, G. and Przeworski, M., 2009. Broad-scale recombination patterns underlying proper disjunction in humans. *PLoS Genetics*, 5(9)

Flint-Garcia, S.A., 2013. Genetics and consequences of crop domestication. *Journal of Agricultural and Food Chemistry*, 61(35), 8267-8276.

Foll, M. and Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977-993.

Fontaine, K.M., Cooley, J.R. and Simon, C., 2007. Evidence for paternal leakage in hybrid periodical cicadas (*Hemiptera: Magicicada* spp.). *PloS One*, 2(9), 892.

Food and Agriculture Organisation. 2007. The State of the World's Animal Genetic Resources for Food and Agriculture. FAO, Rome. Available at: http://www.fao.org/3/a1260e/a1260e.pdf Last Accessed 21st September 2021

Forsgren, E., Budge, G.E., Charrière, J.D. and Hornitzky, M.A., 2013. Standard methods for European foulbrood research. *Journal of Apicultural Research*, 52(1), 1-14.

Franck, P., Garnery, L., Celebrano, G., Solignac, M. and Cornuet, J.M., 2000. Hybrid origins of honeybees from Italy *(Apis mellifera ligustica*) and Sicily (*A. m. sicula*). *Molecular Ecology*, 9(7), 907-921.

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B.P., Hepburn, H.R., Solignac, M. and Cornuet, J.M., 2001. Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity*, *86*(4), 420-430.

Frankel, O.H., 1970. Genetic dangers in the green revolution. *World Agric*, 19(3), 9-13.

Frankham, R., 2010. Where are we in conservation genetics and where do we need to go? *Conservation genetics*, 11(2), 661-663.

Frankham, R., Ballou, J.D., Eldridge, M.D., Lacy, R.C., Ralls, K., Dudash, M.R. and Fenster, C.B., 2011. Predicting the probability of outbreeding depression. *Conservation Biology*, 25(3), 465-475.

Fries, I., 1993. *Nosema apis*—a parasite in the honey bee colony. *Bee World*, *74*(1), 5-19.

Gailing, O. and Curtu, A.L., 2014. Interspecific gene flow and maintenance of species integrity in oaks. *Annals of Forest Research*, 57(1), 5-18.

Gallai, N., Salles, J.M., Settele, J. and Vaissière, B.E., 2009. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics,* 68(3), 810-821.

Gao, S.B., Mo, L.D., Zhang, L.H., Zhang, J.L., Wu, J.B., Wang, J.L., Zhao, N.X. and Gao, Y.B., 2018. Phenotypic plasticity vs. local adaptation in quantitative traits differences of *Stipa grandis* in semi-arid steppe, China. *Scientific Reports*, 8(1), 1-8.

García-Navas, V., Ferrer, E.S., Sanz, J.J. and Ortego, J., 2014. The role of immigration and local adaptation on fine-scale genotypic and phenotypic population divergence in a less mobile passerine. *Journal of Evolutionary Biology*, 27(8), 1590-1603.

Garlake, P.S., 1995. The hunter's vision: the prehistoric art of Zimbabwe. *American Antiquity*, 63(2), 360-361

Garnery, L., Franck, P., Baudry, E., Vautrin, D., Cornuet, J.M. and Solignac, M., 1998. Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*) II. Microsatellite loci. *Genetics Selection Evolution*, 30(Supplement), 49-S74.

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhué, C. and Estoup, A., 2013. Estimation of population allele

frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology*, 22(14), 3766-3779.

Gautier, M., Vitalis, R., Flori, L. and Estoup, A., 2022. f-Statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolfstat. *Molecular Ecology Resources*, 22(4), 1394-1416.

Geldmann, J. and González-Varo, J.P., 2018. Conserving honey bees does not help wildlife. *Science*, 359(6374), 392-393.

Genersch, E., 2010. Honey bee pathology: current threats to honey bees and beekeeping. *Applied Microbiology and Biotechnology*, 87(1), 87-97.

Gertsch, W.J., 1984. The spider family *Nesticidae* (*Araneae*) in North America, Central America, and the West Indies. University of Texas Libraries.

Geslin, B., Gauzens, B., Baude, M., Dajoz, I., Fontaine, C., Henry, M., Ropars, L., Rollin, O., Thébault, E. and Vereecken, N.J., 2017. Massively introduced managed species and their consequences for plant–pollinator interactions. *Advances in Ecological Research*, 57, 147-199

Gill, R.J., Ramos-Rodriguez, O. and Raine, N.E., 2012. Combined pesticide exposure severely affects individual-and colony-level traits in bees. *Nature*, 491(7422), 105-108.

Girardot, F., Monnier, V. and Tricoire, H., 2004. Genome wide analysis of common and specific stress responses in adult *Drosophila melanogaster*. *BMC Genomics*, 5(1), 1-16.

Glémin, S. and Ronfort, J., 2013. Adaptation and maladaptation in selfing and outcrossing species: new mutations versus standing variation. *International Journal of Organic Evolution*, 67(1), 225-240.

Gmel, A., Guichard, M., Dainat, B., Williams, G., Eynard, S., Vignal, A. and Neuditschko, M., 2021. Runs of homozygosity derived from pool-seq data reveal fine-scale population structures in Western honey bees (*Apis mellifera*). Preprint.

González-Varo, J.P. and Geldmann, J., 2018. Response—"Bee conservation: key role of managed bees" and "Bee conservation: inclusive solutions". *Science*, 360(6387), 390-390.

Good, J.M. and Sullivan, J., 2001. Phylogeography of the red-tailed chipmunk (Tamias ruficaudus), a northern Rocky Mountain endemic. *Molecular Ecology*, 10(11), 2683-2695.

Good, J.M., Demboski, J.R., Nagorsen, D.W. and Sullivan, J., 2003. Phylogeography and introgressive hybridization: chipmunks (genus *Tamias*) in the northern Rocky Mountains. *Evolution*, 57(8), 1900-1916.

Good, J.M., Vanderpool, D., Keeble, S. and Bi, K., 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, 69(8), 1961-1972.

Gorbalenya, A.E. and Koonin, E.V., 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Current Opinion in Structural Biology, 3(3), 419-429.

Gordon, D.H., 1958. The pre-historic background of Indian culture.

Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices, *Journal of Heredity*, 86(6), 485-486

Goulet, B.E., Roda, F. and Hopkins, R., 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiology*, 173(1), 65-78.

Goulson, D., 2003. Effects of introduced bees on native ecosystems. Annual Review of Ecology, Evolution, and Systematics, 1-26.

Goulson, D., 2019. The insect apocalypse, and why it matters. *Current Biology*, 29(19), 967-971.

Government Petition, 2021 Reverse the ban on honeybee imports. Stop the honeybee decline! Available at https://petition.parliament.uk/petitions/573228 Last accessed 24th July 2022

Government response, 2021. Response to: Stop the importation of honey bees into GB from the EU via Northern Ireland. Available at: https://petition.parliament.uk/petitions/577603 Last Accessed: 24th July 2022.

Goymer, P., 2007. Synonymous mutations break their silence. *Nature Reviews Genetics*, 8(2), 92-92.

Gramates, L.S., Agapite, J., Attrill, H., Calvi, B.R., Crosby, M.A., Dos Santos, G., Goodman, J.L., Goutte-Gattat, D., Jenkins, V.K., Kaufman, T. and Larkin, A., 2022. FlyBase: A guided tour of highlighted features. *Genetics,* 220(4)

Greenlaw, J.S., 1993. Behavioral and morphological diversification in Sharp-tailed Sparrows (*Ammodramus caudacutus*) of the Atlantic Coast. *The Auk*, 110(2), 286-293.

Groenen, F. and Meurisse, N., 2012. Historical distribution of the oak processionary moth *Thaumetopoea processionea* in Europe suggests recolonization instead of expansion. *Agricultural and forest entomology*, 14(2), 147-155.

Groeneveld, L.F., Kirkerud, L.A., Dahle, B., Sunding, M., Flobakk, M., Kjos, M., Henriques, D., Pinto, M.A. and Berg, P., 2020. Conservation of the dark bee (*Apis*

mellifera mellifera): Estimating C-lineage introgression in Nordic breeding stocks. *Acta Agriculturae Scandinavica, Section A—Animal Science*, 69(3), 157-168.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E.J.A.G. and Weigend, S., 2010. Genetic diversity in farm animals–a review. *Animal genetics*, 41, 6-31.

Grozinger, C.M. and Zayed, A., 2020. Improving bee health through genomics. *Nature Reviews Genetics*, 21(5), 277-291.

Grupe, A.C. and Quandt, C.A., 2020. A growing pandemic: A review of *Nosema* parasites in globally distributed domesticated and native bees. *PLoS Pathogens*, 16(6), 1008580.

Guichard, M., Dainat, B., Eynard, S., Vignal, A., Servin, B., Beestrong Consortium and Neuditschko, M., 2021. Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences. *Animal Genetics*. 52(4), 472-481

Haasl, R.J. and Payseur, B.A., 2012. Microsatellites as targets of natural selection. *Molecular Biology and Evolution*, 30(2), 285-298.

Haenel, Q., Laurentino, T.G., Roesti, M. and Berner, D., 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology*, 27(11), 2477-2497.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T. and Goulson, D., 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PloS One, 12(10)

Hanna, Z.R., Dumbacher, J.P., Bowie, R.C., Henderson, J.B. and Wall, J.D., 2018. Whole-genome analysis of introgression between the spotted owl and barred owl (*Strix occidentalis* and *Strix varia*, respectively; Aves*: Strigidae*) in western North America. *G3: Genes, Genomes, Genetics*, 8(12), 3945-3952.

Hansen, M.M., Bekkevold, D., Jensen, L.F., Mensberg, K.L.D. and Nielsen, E.E., 2006. Genetic restoration of a stocked brown trout *Salmo trutta* population using microsatellite DNA analysis of historical and contemporary samples. *Journal of Applied Ecology*, 43(4), 669-679.

Hansen, M.M., Fraser, D.J., Meier, K. and MENSBERG, K.L.D., 2009. Sixty years of anthropogenic pressure: A spatio-temporal genetic analysis of brown trout populations subject to stocking and population declines. *Molecular Ecology*, *18*(12), 2549-2562.

Harpur, B.A., Kent, C.F., Molodtsova, D., Lebon, J.M., Alqarni, A.S., Owayss, A.A. and Zayed, A., 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences*, 111(7), 2614-2619.

Harpur, B.A., Kent, C.F., Molodtsova, D., Lebon, J.M., Alqarni, A.S., Owayss, A.A. and Zayed, A., 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences*, 111(7), 2614-2619.

Harpur, B.A., Minaei, S., Kent, C.F. and Zayed, A., 2012. Management increases genetic diversity of honey bees via admixture. *Molecular Ecology*, 21(18), 4414-4421.

Hart, A.G., Bot, A. and Brown, M.J., 2002. A colony-level response to disease control in a leaf-cutting ant. *Naturwissenschaften*, 89(6), 275-277.

Harwood, T.D., Tomlinson, I., Potter, C.A. and Knight, J.D., 2011. Dutch elm disease revisited: past, present and future management in Great Britain. *Plant Pathology*, 60(3), 545-555.

Hassett, J., Browne, K.A., McCormack, G.P., Moore, E., Society, N.I.H.B., Soland, G. and Geary, M., 2018. A significant pure population of the dark European honey bee (*Apis mellifera mellifera*) remains in Ireland. *Journal of Apicultural Research*, 57(3), 337-350.

Heap, E-J., 2022. 'Conservation of honey bees may lead to food shortages'. The Telegraph, London, 23 May,

Heath, D.D., Heath, J.W., Bryden, C.A., Johnson, R.M. and Fox, C.W., 2003. Rapid evolution of egg size in captive salmon. *Science*, 299(5613), 1738-1740.

Hedrick, P.W., 2010. Cattle ancestry in bison: explanations for higher mtDNA than autosomal ancestry. *Molecular Ecology*, 19(16), 3328-3335.

Hedrick, P.W., 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular ecology*, 22(18), 4606-4618.

Henriques, D., Browne, K.A., Barnett, M.W., Parejo, M., Kryger, P., Freeman, T.C., Muñoz, I., Garnery, L., Highet, F., Jonhston, J.S. and McCormack, G.P., 2018. High sample throughput genotyping for estimating C-lineage introgression in the dark honeybee: an accurate and cost-effective SNP-based tool. *Scientific Reports*, 8(1), 1-14.

Henriques, D., Wallberg, A., Chávez-Galarza, J., Johnston, J.S., Webster, M.T. and Pinto, M.A., 2018. Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula. *Scientific Reports*, *8*(1), 1-14.

Henry, M. and Rodet, G., 2018. Controlling the impact of the managed honeybee on wild bees in protected areas. *Scientific Reports*, 8(1), 1-10.

Hermisson, J. and Pennings, P.S., 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4), 2335-2352.

Hernandez, J., Meints, B. and Hayes, P., 2020. Introgression breeding in barley: Perspectives and case studies. *Frontiers in Plant Science*, 11, 761.

Hindar, K., Ryman, N. and Utter, F., 1991. Genetic effects of cultured fish on natural fish populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(5), 945-957.

Hoekstra, H.E. and Coyne, J.A., 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution: International Journal of Organic Evolution*, 61(5), 995-1016.

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)

Homola, J.J., Loftin, C.S., Cammen, K.M., Helbing, C.C., Birol, I., Schultz, T.F. and Kinnison, M.T., 2019. Replicated landscape genomics identifies evidence of local adaptation to urbanization in wood frogs. *Journal of Heredity*, 110(6), 707-719.

Howard-McCombe, J., Ward, D., Kitchener, A.C., Lawson, D., Senn, H.V. and Beaumont, M., 2021. On the use of genome-wide data to model and date the time of anthropogenic hybridisation: an example from the Scottish wildcat. *Molecular Ecology*. 30(15), 3688-3702

Huang, D.W., Sherman, B.T. and Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44-57.

Hultberg, T., Sandström, J., Felton, A., Öhman, K., Rönnberg, J., Witzell, J. and Cleary, M., 2020. Ash dieback risks an extinction cascade. *Biological Conservation*, *244*, 108516.

Hunt, R., Sauna, Z.E., Ambudkar, S.V., Gottesman, M.M. and Kimchi-Sarfaty, C., 2009. Silent (synonymous) SNPs: should we care about them? Single nucleotide polymorphisms, *Methods in Molecular Biology*, 23-39.

Hunt, R.C., Simhadri, V.L., Iandoli, M., Sauna, Z.E. and Kimchi-Sarfaty, C., 2014. Exposing synonymous mutations. *Trends in Genetics*, 30(7), 308-321.

Huntingford, F.A., 2004. Implications of domestication and rearing conditions for the behaviour of cultivated fishes. *Journal of Fish Biology*, *65*, 122-142.

Hutchinson, J.M., Schlitt, B. and Reise, H., 2021. One town's invasion by the pest slug Arion vulgaris (*Gastropoda: Arionidae*): microsatellites reveal little introgression from Arion ater and limited gene flow between infraspecific races in both species. *Biological Journal of the Linnean Society*, 134(4), 835-850.

Ibrahim, A., Reuter, G.S. and Spivak, M., 2007. Field trial of honey bee colonies bred for mechanisms of resistance against *Varroa destructor*. Apidologie, 38(1), pp.67-76.

Ilyasov, R.A., Poskryakov, A.V., Petukhov, A.V. and Nikolenko, A.G., 2016. New approach to the mitotype classification in black honeybee *Apis mellifera mellifera* and Iberian honeybee *Apis mellifera iberiensis*. *Russian journal of genetics*, 52(3), 281-291.

Importation of bees order, Isle Of Man Department of Agriculture, Fisheries and Forestry, 1988. Available at https://www.iombeekeepers.com/wp-content/uploads/2015/08/Importation-of-Bees-Order-1988.pdf Last accessed, 14th of July 2022

Inbar, S., Cohen, P., Yahav, T. and Privman, E., 2020. Comparative study of population genomic approaches for mapping colony-level traits. *PLoS Computational Biology*, 16(3), 1007653.

Isle of Man Bees Act, Isle Of Man Department of Agriculture, Fisheries and Forestry, 1989.. Available at: https://www.iombeekeepers.com/wp-content/uploads/2015/08/Bees-Act-1989.pdf Last Accessed: 27th of July 2022

Jacobs, A. and Elmer, K.R., 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Molecular Ecology*, 30(20), 4955-4969.

Jahn, G.C., Beardsley, J.W. and González-Hernández, H., 2003. A review of the association of ants with mealybug wilt disease of pineapple. *Proceedings of Hawaiian Entomology Society*, 36, 9-28

Janoušek, V., Munclinger, P., Wang, L., Teeter, K.C. and Tucker, P.K., 2015. Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, 32(5), 1208-1220.

Jensen, A.B., Palmer, K.A., Boomsma, J.J. and Pedersen, B.V., 2005. Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, 14(1), 93-106.

Jensen, A.B., Palmer, K.A., Chaline, N., Raine, N.E., Tofilski, A., Martin, S.J., Pedersen, B.V., Boomsma, J.J. and Ratnieks, F.L., 2005. Quantifying honey bee mating range and isolation in semi-isolated valleys by DNA microsatellite paternity analysis. *Conservation Genetics*, 6(4), 527-537.

Jiggins, C.D., 2017. The ecology and evolution of *Heliconius* butterflies. Oxford University Press.

Johnson, R., Clair, B.S. and Lipow, S., 2001. Genetic conservation in applied tree breeding programs. *Proceedings ITTO conference on in situ and ex situ conservation of commercial tropical trees:* 215-230.

Jombart, T. and Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070-3071.

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

Jombart, T., Pontier, D. and Dufour, A.B., 2009. Genetic markers in the playground of multivariate analysis. *Heredity*, *102*(4), 330-341.

Jones, E.P., Conyers, C., Tomkies, V., Semmence, N., Fouracre, D., Wakefield, M. and Stainton, K., 2020. Managing incursions of *Vespa velutina nigrithorax* in the UK: An emerging threat to apiculture. *Scientific Reports*, 10(1), 1-8.

Jou, Y., Chiang, C.P., Jauh, G.Y. and Yen, H.E., 2006. Functional characterization of ice plant SKD1, an AAA-type ATPase associated with the endoplasmic reticulum-Golgi network, and its role in adaptation to salt stress. *Plant Physiology*, 141(1), 135-146.

Jug, T., Berrebi, P. and Snoj, A., 2005. Distribution of non-native trout in Slovenia and their introgression with native trout populations as observed through microsatellite DNA analysis. *Biological Conservation*, 123(3), 381-388.

Kalinowski, S. T., Manlove, K. R. and Taper, M. L. 2007. ONCOR-manual. Available at: https://www.montana.edu/kalinowski/software/documents/ONCOR_Manual_21Oct2007.pdf Last Accessed 21th July 2022

Kalinowski, S.T., Wagner, A.P. and Taper, M.L., 2006. ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6(2), 576-579.

Karaiskou, N., Triantafyllidis, A., Katsares, V., Abatzopoulos, T.J. and Triantaphyllidis, C., 2009. Microsatellite variability of wild and farmed populations of *Sparus aurata*. *Journal of Fish Biology*, 74(8), 1816-1825.

Karlsson, S., Diserud, O.H., Fiske, P., Hindar, K. and Handling editor: W. Stewart Grant, 2016. Widespread genetic introgression of escaped farmed Atlantic salmon in wild salmon populations. *ICES Journal of Marine Science*, 73(10), 2488-2498.

Kauhausen-Keller, D. and Keller, R., 1994. Morphometrical control of pure race breeding in the honeybee (*Apis mellifera* L). *Apidologie*, *25*(2), 133-143.

Kawecki, T.J. and Ebert, D., 2004. Conceptual issues in local adaptation. *Ecology letters*, 7(12), 1225-1241.

Keeling, M.J., Franklin, D.N., Datta, S., Brown, M.A. and Budge, G.E., 2017. Predicting the spread of the Asian hornet (*Vespa velutina*) following its incursion into Great Britain. *Scientific Reports*, 7(1), 1-7.

Kent, C.F. and Zayed, A., 2013. Evolution of recombination and genome structure in eusocial insects. *Communicative & Integrative Biology*, 6(2).

Kim, B.Y., Wei, X., Fitz-Gibbon, S., Lohmueller, K.E., Ortego, J., Gugger, P.F. and Sork, V.L., 2018. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus sect. Quercus: Fagaceae*). *Molecular Ecology*, 27(22), 4556-4571.

Kim, E.S. and Rothschild, M.F., 2014. Genomic adaptation of admixed dairy cattle in East Africa. *Frontiers in Genetics*, 5, 443.

Kim, E.S., Elbeltagy, A.R., Aboul-Naga, A.M., Rischkowsky, B., Sayre, B., Mwacharo, J.M. and Rothschild, M.F., 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116(3), 255-264.

Kleijn, D., Biesmeijer, K., Dupont, Y.L., Nielsen, A., Potts, S.G. and Settele, J., 2018. Bee conservation: inclusive solutions. *Science*, 360(6387), 389-390.

Knaus, B.J. and Grünwald, N.J., 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology resources, 17(1), 44-53.

Koeniger, N., Koeniger, G., Gries, M. and Tingek, S., 2005. Drone competition at drone congregation areas in four Apis species. *Apidologie*, 36(2), 211-221.

Kofler, R., Pandey, R.V. and Schlötterer, C., 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24), 3435-3436.

Kong, S. and Kubatko, L.S., 2021. Comparative performance of popular methods for hybrid detection using genomic data. *Systematic Biology*, 70(5), 891-907.

Korb, S.K., Fric, Z.F. and Bartoňová, A., 2016. Phylogeography of *Koramius charltonius* (Gray, 1853) (*Lepidoptera*: *Papilionidae*): a case of too many poorly circumscribed subspecies. *Nota lepidopterologica*, 39(2), 169.

Kotliar, N.B., 2000. Application of the new keystone-species concept to prairie dogs: how well does it work?, *Conservation Biology*, 14(6), 1715-1721.

Kritsky, G., 2015. The tears of re: beekeeping in ancient Egypt. Oxford University Press

Kritsky, G., 2017. The quest for the perfect hive: ancient Mediterranean origins. Beekeeping in the Mediterranean from antiquity to the resent. *Athens: Agricultural Organization*, 50-6.

LaFramboise, T., 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Research, 37(13), 4181-4193.

Larson, E.L., White, T.A., Ross, C.L. and Harrison, R.G., 2014. Gene flow and the maintenance of species boundaries. *Molecular Ecology*, *23*(7), 1668-1678.

Laugen, A.T., Laurila, A., Räsänen, K. and Merilä, J., 2003. Latitudinal countergradient variation in the common frog (*Rana temporaria*) development rates–evidence for local adaptation. *Journal of Evolutionary Biology*, 16(5), 996-1005.

Laurent, S., Pfeifer, S.P., Settles, M.L., Hunter, S.S., Hardwick, K.M., Ormond, L., Sousa, V.C., Jensen, J.D. and Rosenblum, E.B., 2016. The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. *Molecular Ecology*, 25(1), 306-323.

Lawson, D.J., Van Dorp, L. and Falush, D., 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 1-11.

Lawson, L.P., Fessl, B., Vargas, F.H., Farrington, H.L., Cunninghame, H.F., Mueller, J.C., Nemeth, E., Sevilla, P.C. and Petren, K., 2017. Slow motion extinction:

inbreeding, introgression, and loss in the critically endangered mangrove finch (*Camarhynchus heliobates*). *Conservation Genetics*, 18(1), 159-170.

Le Corre, V., Siol, M., Vigouroux, Y., Tenaillon, M.I. and Délye, C., 2020. Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *Proceedings of the National Academy of Sciences*, 117(41), 25618-25627.

Leipe, D.D., Koonin, E.V. and Aravind, L., 2004. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. Journal of Molecular Biology, 343(1), 1-28.

Lemopoulos, A., Prokkola, J.M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M.L., Koskiniemi, J. and Vainikka, A., 2019. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106-2120.

Lewis, S.L. and Maslin, M.A., 2015. Defining the anthropocene. *Nature*, *519*(7542), 171-180.

Lewontin RC. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*. 49(1):49-67.

Lewontin, R.C. and Kojima, K.I., 1960. The evolutionary dynamics of complex polymorphisms. *Evolution*, 458-472.

Leza, M., Herrera, C., Marques, A., Roca, P., Sastre-Serra, J. and Pons, D.G., 2019. The impact of the invasive species *Vespa velutina* on honeybees: A new approach based on oxidative stress. *Science of The Total Environment*, 689, 709-715.

Li, H. and Durbin, R., 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, M., Jin, L., Ma, J., Tian, S., Li, R. and Li, X., 2016. Detecting mitochondrial signatures of selection in wild Tibetan pigs and domesticated pigs. *Mitochondrial DNA Part A*, 27(1), 747-752.

Li, M., Li, C. and Guan, W., 2008. Evaluation of coverage variation of SNP chips for genome-wide association studies. European Journal of Human Genetics, 16(5), 635-643.

Li, W-H., 1997. Molecular Evolution *Genetics Research*, 70(3), 267-271.

Li, Z., Yang, F., Xuan, Y., Xi, R. and Zhao, R., 2019. Pelota-interacting G protein Hbs1 is required for spermatogenesis in *Drosophila. Scientific Reports*, 9(1), 1-14.

Lima, T.G., Burton, R.S. and Willett, C.S., 2019. Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod *Tigriopus californicus*. *Evolution,* 73(3), 609-620.

Linnen, C.R., Kingsley, E.P., Jensen, J.D. and Hoekstra, H.E., 2009. On the origin and spread of an adaptive allele in deer mice. *Science*, 325(5944), 1095-1098.

Lipaeva, P., Vereshchagina, K., Drozdova, P., Jakob, L., Kondrateva, E., Lucassen, M., Bedulina, D., Timofeyev, M., Stadler, P. and Luckenbach, T., 2021. Different ways to play it cool: Transcriptomic analysis sheds light on different activity patterns of three amphipod species under long-term cold exposure. *Molecular Ecology*, 30(22), 5735-5751.

Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., Zhang, J., Feng, J., Kaltenboeck, L., Kucuktas, H. and Liu, Z., 2014. Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes*, 7(1), 1-12.

Liu, Y., Yang, Q. and Zhao, F., 2021. Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annual Review of Biochemistry*, 90, 375.

Lodesani, M. and Costa, C., 2003. Bee breeding and genetics in Europe. *Bee World*, 84(2), 69-85.

Lopes, R.J., Johnson, J.D., Toomey, M.B., Ferreira, M.S., Araujo, P.M., Melo-Ferreira, J., Andersson, L., Hill, G.E., Corbo, J.C. and Carneiro, M., 2016. Genetic basis for red coloration in birds. *Current Biology*, 26(11), 1427-1434.

Lopes, R.J., Johnson, J.D., Toomey, M.B., Ferreira, M.S., Araujo, P.M., Melo-Ferreira, J., Andersson, L., Hill, G.E., Corbo, J.C. and Carneiro, M., 2016. Genetic basis for red coloration in birds. *Current Biology*, 26(11), 1427-1434.

Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F. and Storfer, A., 2017. Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*. 17, 142-152

Mace, G.M. and Purvis, A., 2008. Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology*, 17(1), 9-19.

Malinsky, M., Matschiner, M. and Svardal, H., 2021. Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584-595.

Mallet, J., 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution, 20(5), 229-237.

Manel, S., Berthier, P. and Luikart, G., 2002. Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation biology*, *16*(3), 650-659.

Manier, M.K., 2004. Geographic variation in the long-nosed snake, *Rhinocheilus lecontei (Colubridae*): beyond the subspecies debate. *Biological Journal of the Linnean Society*, 83(1), 65-85.

Mao, X. and Rossiter, S.J., 2020. Genome-wide data reveal discordant mitonuclear introgression in the intermediate horseshoe bat (*Rhinolophus affinis*). *Molecular Phylogenetics and Evolution*, 150,

Martin, S.H. and Jiggins, C.D., 2017. Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development, 47, 69-74.

Martin, S.H., Davey, J.W., Salazar, C. and Jiggins, C.D., 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS biology*, 17(2)

Matheson A., Buchmann S.L., O'Toole C., Westrich P., Williams I.H. 1996. The conservation of bees, *Academic Press*, Harcourt Brace, London.

Mathpal, Y., 1984. Newly discovered rock paintings in central India showing honey collection. *Bee world*, 65(3), 121-126.

Mattila, H.R. and Seeley, T.D., 2007. Genetic diversity in honey bee colonies enhances productivity and fitness. *Science*, 317(5836), 362-364.

Maul, V. and Hähnle, A., 1994. Morphometric studies with pure bred stock of *Apis mellifera carnica* Pollmann from Hessen. *Apidologie*, 25(2), 119-132.

Mayr, E., 1982. Of what use are subspecies? *The Auk*, 99(3), 593-595.

Mayr, E., 2000. The biological species concept. Species concepts and phylogenetic theory: a debate. Columbia University Press, New York, 17-29.

McGowan, P. and Madge, S., 2010. Pheasants, Partridges & Grouse: Including buttonquails, sandgrouse and allies. Bloomsbury Publishing.

McGuire, J.A., Linkem, C.W., Koo, M.S., Hutchison, D.W., Lappin, A.K., Orange, D.I., Lemos-Espinal, J., Riddle, B.R. and Jaeger, J.R., 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of *crotaphytid* lizards. *Evolution: International Journal of Organic Evolution*, 61(12), 2879-2897.

McKinney, G.J., Larson, W.A., Seeb, L.W. and Seeb, J.E., 2017. RAD seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al.(2016). Molecular Ecology Resources, 17(3), pp.356-361.

McKinney, W., 2010, Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference. 445(1), 51-56.

McLean-Rodríguez, F.D., Camacho-Villa, T.C., Almekinders, C.J., Pè, M.E., Dell'Acqua, M. and Costich, D.E., 2019. The abandonment of maize landraces over the last 50 years in Morelos, Mexico: a tracing study using a multi-level perspective. *Agriculture and Human Values,* 36(4), 651-668.

Meixner, M.D., Pinto, M.A., Bouga, M., Kryger, P., Ivanova, E. and Fuchs, S., 2013. Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *Journal of Apicultural Research*, 52(4), 1-28.

Meyer, R.S. and Purugganan, M.D., 2013. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, 14(12), 840-852

Michel, A.P., Rull, J., Aluja, M. and Feder, J.L., 2007. The genetic structure of hawthorn-infesting Rhagoletis pomonella populations in Mexico: implications for sympatric host race formation. *Molecular Ecology*, 16(14), 2867-2878.

Mikkelsen, T., Hillier, L., Eichler, E., Zody, M., Jaffe, D., Yang, S.P., Enard, W., Hellmann, I., Lindblad-Toh, K., Altheide, T. and Archidiacono, N., 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055), 69-87.

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

Milligan, B.G., 2003. Maximum-likelihood estimation of relatedness. *Genetics*, 163(3), 1153-1167.

Mitchell, R.J., Hewison, R.L., Hester, A.J., Broome, A. and Kirby, K.J., 2016. Potential impacts of the loss of *Fraxinus excelsior* (Oleaceae) due to ash dieback on woodland vegetation in Great Britain. *New Journal of Botany*, 6(1), 2-15.

Monceau, K., Bonnard, O. and Thiéry, D., 2014. *Vespa velutina:* a new invasive predator of honeybees in Europe. *Journal of Pest Science*, 87(1), 1-16.

Moore, A.J., Brodie III, E.D. and Wolf, J.B., 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution*, 51(5), 1352-1362.

Mora-Márquez, F., García-Olivares, V., Emerson, B.C. and López de Heredia, U., 2017. ddradseqtools: a software package for in silico simulation and testing of double-digest RAD seq experiments. *Molecular Ecology Resources*, 17(2), 230-246.

Morfin, N., Given, K., Evans, M., Guzman-Novoa, E. and Hunt, G.J., 2020. Grooming behavior and gene expression of the Indiana “mite-biter” honey bee stock. *Apidologie*, 51(2), 267-275.

Morin, P.A., Martien, K.K., Archer, F.I., Cipriano, F., Steel, D., Jackson, J. and Taylor, B.L., 2010. Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. *Journal of Heredity*, 101(1), 1-10.

Mueller, U.G., Gerardo, N.M., Aanen, D.K., Six, D.L. and Schultz, T.R., 2005. The evolution of agriculture in insects. *Annual. Review of Ecology. Evolution. And Systematics.* 36, 563-595.

Muir, A.P., Biek, R., Thomas, R. and Mable, B.K., 2014. Local adaptation with high gene flow: temperature parameters drive adaptation to altitude in the common frog (*Rana temporaria*). *Molecular Ecology*, 23(3), 561-574.

Muñoz, I. and De la Rúa, P., 2021. Wide genetic diversity in Old World honey bees threaten by introgression. *Apidologie,* 52(1), 200-217.

Muñoz, I., Dall’Olio, R., Lodesani, M. and De la Rúa, P., 2009. Population genetic structure of coastal Croatian honeybees (*Apis mellifera carnica*). *Apidologie*, 40(6), 617-626.

Munoz, I., Henriques, D., Jara, L., Johnston, J.S., Chávez-Galarza, J., De La Rua, P. and Pinto, M.A., 2017. SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*). *Molecular Ecology Resources*, 17(4), 783-795.

Muñoz, I., Henriques, D., Johnston, J.S., Chávez-Galarza, J., Kryger, P. and Pinto, M.A., 2015. Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*). *PloS one*, 10(4).

Muñoz, I., Madrid-Jiménez, M.J. and De la Rúa, P., 2012. Temporal genetic analysis of an introgressed island honey bee population (Tenerife, Canary Islands, Spain). *Journal of Apicultural Research*, 51(1), 144-146.

Nadeau, N.J., Ruiz, M., Salazar, P., Counterman, B., Medina, J.A., Ortiz-Zuazaga, H., Morrison, A., McMillan, W.O., Jiggins, C.D. and Papa, R., 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome research*, 24(8), 1316-1333.

National Bee Unit Hive Count 2022, February Update 2022 - 2021 Hive Count. Available at: https://nationalbeeunit.com/public/News/news.cfm Last Accessed 19th July 2022.

National Bee Unit Husbandry Survey, 2017. The National Bee Unit. Available at: https://nationalbeeunit.com/index.cfm?pageid=203 Last Accessed 21st of July 2022.

Natola, L., Curtis, A., Hudon, J. and Burg, T.M., 2021. Introgression between *Sphyrapicus nuchalis* and *S. varius* sapsuckers in a hybrid zone in west-central Alberta. *Journal of Avian Biology*, 52(8).

Needham, S. and Evans, J., 1987. Honey and dripping: Neolithic food residues from Runnymede Bridge. *Oxford Journal of Archaeology*, 6(1), 21-28.

Negri, V., Maxted, N. and Veteläinen, M., 2009. European landrace conservation: an Introduction. *European Landraces: On-farm Conservation, Management and Use*, 1-22.

Negro, J.J., Torres, M.J. and Godoy, J.A., 2001. RAPD analysis for detection and eradication of hybrid partridges (*Alectoris rufa*× *A. graeca*) in Spain. *Biological Conservation*, 98(1), 19-24.

Nicholls, J.A., Double, M.C., Rowell, D.M. and Magrath, R.D., 2000. The evolution of cooperative and pair breeding in thornbills *Acanthiza (Pardalotidae*). *Journal of Avian Biology*, 31(2), 165-176.

Nolte, A.W., Freyhof, J., Stemshorn, K.C. and Tautz, D., 2005. An invasive lineage of sculpins, *Cottus* sp.(*Pisces, Teleostei*) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proceedings of the Royal Society B: Biological Sciences*, 272(1579), 2379-2387.

Nordborg, M. and Tavaré, S., 2002. Linkage disequilibrium: what history has to tell us. *Trends in Genetics*, 18(2), 83-90.

Norrström, N., Niklasson, M. and Leidenberger, S., 2021. Winter weight loss of different subspecies of honey bee *Apis mellifera colonies* (Linnaeus, 1758) in southwestern Sweden. *Plos one*, 16(10).

Notter, D.R., 1999. The importance of genetic diversity in livestock populations of the future. *Journal of animal science*, *77*(1), 61-69.

Numa-Vergel, S.J., Sandoval-Cáceres, Y.P. and Vergara-Navarro, E.V., 2021. First Record of *Aethina tumida* Murray (Coleoptera: *Nitidulidae*) in Colombia. *Journal of Agricultural and Urban Entomology*, 37(1), 33-37.

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M. and Portnoy, D.S., 2018. These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27, 1-14

Oleksa, A. and Tofilski, A., 2015. Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies. *Apidologie*, 46(1), 49-60.

Oleksa, A., Chybicki, I., Tofilski, A. and Burczyk, J., 2011. Nuclear and mitochondrial patterns of introgression into native dark bees (*Apis mellifera mellifera*) in Poland. *Journal of Apicultural Research*, 50(2), 116-129.

Ollerton, Jeff. 2013 'Are honey bees native to Britain? And does it matter?', Jeff Ollerton blog, 12 July 2013. Available at: https://jeffollerton.co.uk/2013/07/12/are-honey-bees-native-to-britain-and-does-it-matter/ Last Accessed 21st of July 2022

Olsen, K.M. and Wendel, J.F., 2013. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology*, 64, 47-70.

Orlova-Bienkowskaja, M.J., 2014. Ashes in Europe are in danger: the invasive range of *Agrilus planipennis* in European Russia is expanding. *Biological Invasions*, *16*(7), 1345-1349.

Ortiz-Barrientos, D. and Baack, E.J., 2014. Species integrity in trees. *Molecular Ecology*, 23, 4188-4191

PA media, 2021, The BBC, Brexit: Millions of bees could be 'destroyed' over import rules. Available at: https://www.bbc.co.uk/news/uk-england-kent-55901419 Last Accessed: 24th July 2022

Page Jr, R.E. and Peng, C.Y.S., 2001. Aging and development in social insects with emphasis on the honey bee, Apis mellifera L. *Experimental Gerontology*, 36(4-6), 695-711.

Pager, H., 1973. Rock paintings in southern Africa showing bees and honey hunting. *Bee World*, 54(2), 61-68.

Pager, H., 1976. Cave paintings suggest honey hunting activities in Ice Age times. *Bee World*, 57(1), 9-14.

Palmer, K.A. and Oldroyd, B.P., 2000. Evolution of multiple mating in the genus Apis. *Apidologie*, 31(2), 235-248.

Palmeri, V., Scirtò, G., Malacrinò, A., Laudani, F. and Campolo, O., 2015. A scientific note on a new pest for European honey bees: first report of small hive beetle *Aethina tumida*,(Coleoptera: *Nitidulidae*) in Italy. *Apidologie*, 46(4), 527-529.

Palo, J.U., O'Hara, R.B., Laugen, A.T., Laurila, A., Primmer, C.R. and Merilä, J., 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular Ecology*, 12(7), 1963-1978.

Parejo, M., Henriques, D., Pinto, M.A., Soland-Reckeweg, G. and Neuditschko, M., 2018. Empirical comparison of microsatellite and SNP markers to estimate introgression in *Apis mellifera mellifera*. *Journal of Apicultural Research*, 57(4), 504-506.

Parejo, M., Wragg, D., Gauthier, L., Vignal, A., Neumann, P. and Neuditschko, M., 2016. Using whole-genome sequence information to foster conservation efforts for the European Dark Honey Bee, *Apis mellifera mellifera*. *Frontiers in Ecology and Evolution*, *4*, 140.

Parejo, M., Wragg, D., Henriques, D., Vignal, A. and Neuditschko, M., 2017. Genome-wide scans between two honeybee populations reveal putative signatures of human-mediated selection. *Animal Genetics*, 48(6), 704-707.

Parker, K., Ryall, K., Aukema, B.H. and Silk, P., 2020. Early detection of *Agrilus planipennis*: investigations into the attractive range of the sex pheromone (3Z)-lactone. *Entomologia Experimentalis et Applicata*, *168*(2), 166-173.

Parliamentary Office of Science and Technology (POST)., 2010. Postnote: Insect Pollination. Number 348. Parliamentary Office of Science and Technology, London. Available at: https://post.parliament.uk/research-briefings/post-pn-348/ Last Accessed 21st of July 2022.

Patton, A.H., Margres, M.J., Epstein, B., Eastman, J., Harmon, L.J. and Storfer, A., 2020. Hybridizing salamanders experience accelerated diversification. *Scientific Reports*, 10(1), 1-12.

Pautasso, M., Aas, G., Queloz, V. and Holdenrieder, O., 2013. Dieback of European ash: Let's avoid this kind of opportunities for nature conservation. *Biological conservation*, *167*, 452-453.

Perry, A., Wachowiak, W., Downing, A., Talbot, R. and Cavers, S., 2020. Development of a single nucleotide polymorphism array for population genomic studies in four European pine species. *Molecular Ecology* Resources, 20(6), 1697-1705.

Petersen, G.E.L., Fennessy, P.F., Amer, P.R. and Dearden, P.K., 2020. Designing and implementing a genetic improvement program in commercial beekeeping operations. *Journal of Apicultural Research*, 59(4), 638-647.

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), 37135.

Pfeifer, B. and Kapan, D.D., 2019. Estimates of introgression as a function of pairwise distances. *BMC Bioinformatics*, 20(1), 1-11.

Pfeifer, S.P., Laurent, S., Sousa, V.C., Linnen, C.R., Foll, M., Excoffier, L., Hoekstra, H.E. and Jensen, J.D., 2018. The evolutionary history of Nebraska deer mice: local adaptation in the face of strong gene flow. *Molecular Biology and Evolution*, 35(4), 792-806.

Phillimore, A.B. and Owens, I.P., 2006. Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences*, 273(1590), 1049-1053.

Pinto, M.A., Henriques, D., Chávez-Galarza, J., Kryger, P., Garnery, L., van der Zee, R., Dahle, B., Soland-Reckeweg, G., De la Rúa, P., Dall'Olio, R. and Carreck, N.L., 2014. Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, 53(2), 269-278.

Potter, C., Harwood, T., Knight, J. and Tomlinson, I., 2011. Learning from history, predicting the future: the UK Dutch elm disease outbreak in relation to contemporary tree disease threats. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1573), 1966-1974.

Potts, S.G., Imperatriz-Fonseca, V., Ngo, H.T., Aizen, M.A., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J. and Vanbergen, A.J., 2016. Safeguarding pollinators and their values to human well-being. *Nature,* 540(7632), 220-229.

Potts, S.G., Imperatriz-Fonseca, V., Ngo, H.T., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J. and Vanbergen, A.J., 2016. The assessment report on pollinators, pollination and food production: summary for policymakers. *Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. 36

Poulsen, M., 2015. Towards an integrated understanding of the consequences of fungus domestication on the fungus-growing termite gut microbiota. *Environmental Microbiology*, 17(8), 2562-2572.

Preininger, A.M. and Hamm, H.E., 2004. G protein signalling: insights from new structures. Sciences STKE Reviews, 2004(218).

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

Pritchard, J.K., Wen, X. and Falush, D., 2010. Documentation for structure software: Version 2.3. University of Chicago, Chicago, IL, 1-37.

Protection of the Native Irish Honey Bee 2021. Bill by the Irish Senate. (Bill 133). Available at: https://www.oireachtas.ie/en/bills/bill/2021/133/#:~:text=Bill%20entitled%20an%20Act%20to,importation%20of%20non%2Dnative%20species Last Accessed: 18th July 2022.

Przewieslik-Allen, A.M., Burridge, A.J., Wilkinson, P.A., Winfield, M.O., Shaw, D.S., McAusland, L., King, J., King, I.P., Edwards, K.J. and Barker, G.L., 2019. Developing a high-throughput SNP-based marker system to facilitate the introgression of traits from Aegilops species into bread wheat (*Triticum aestivum*). *Frontiers in Plant Science*, 9, 1993.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* 81(3), 559-575.

Quezada-Euán, J.J.G., 2018. Stingless bees of Mexico: The biology, management and conservation of an ancient heritage. *Springer International Publishing*.

Quigley, K.M., Bay, L.K. and van Oppen, M.J., 2020. Genome-wide SNP analysis reveals an increase in adaptive genetic variation through selective breeding of coral. *Molecular Ecology,* 29(12), 2176-2188.

R Core Team 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/

Ramasamy, R.K., Ramasamy, S., Bindroo, B.B. and Naik, V.G., 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springer Plus*, 3(1), 1-3.

Randi, E. and Lucchini, V., 2002. Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Conservation Genetics*, 3(1), 29-43.

Rato, C., Carranza, S., Perera, A., Carretero, M.A. and Harris, D.J., 2010. Conflicting patterns of nucleotide diversity between mtDNA and nDNA in the Moorish gecko, *Tarentola mauritanica*. *Molecular Phylogenetics and Evolution*, 56(3), 962-971.

Ravinet, M., Kume, M., Ishikawa, A. and Kitano, J., 2021. Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *Journal of Evolutionary Biology*, 34(1), 114-127.

Regan, T., Barnett, M.W., Laetsch, D.R., Bush, S.J., Wragg, D., Budge, G.E., Highet, F., Dainat, B., de Miranda, J.R., Blaxter, M. and Freeman, T.C., 2018. Characterisation of the UK honey bee (Apis mellifera) metagenome. Nature Communications, 9, 4995.

Reid, D.G., 1996. Systematics and evolution of *Littorina, Ray Scoiety,* 164

Reid, N., Hird, S., Schulte-Hostedde, A. and Sullivan, J., 2010. Examination of nuclear loci across a zone of mitochondrial introgression between *Tamias ruficaudus* and *T. amoenus*. *Journal of Mammalogy*, 91(6), 1389-1400.

Requier, F., Garnery, L., Kohl, P.L., Njovu, H.K., Pirk, C.W., Crewe, R.M. and Steffan-Dewenter, I., 2019. The conservation of native honey bees is crucial. *Trends in Ecology & Evolution*, 34(9), 789-798.

Requier, F., Paillet, Y., Laroche, F., Rutschmann, B., Zhang, J., Lombardi, F., Svoboda, M. and Steffan-Dewenter, I., 2020. Contribution of European forests to safeguard wild honeybee populations. *Conservation Letters*, 13(2), 12693.

Rhodes, C.J., 2018. Pollinator decline–an ecological calamity in the making?. *Science Progress*, 101(2), 121-160.

Rhymer, J.M. and Simberloff, D., 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, 83-109.

Rieseberg, L.H., 2006. Hybrid speciation in wild sunflowers. *Annals of the Missouri Botanical Garden*, 93(1), 34-48.

Rocchi, L., Paolotti, L., Cortina, C. and Boggia, A., 2016. Conservation of landrace: The key role of the value for agrobiodiversity conservation. An application on ancient tomatoes varieties. *Agriculture and Agricultural Science Procedia*, 8, 307-316.

Rode, N.O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J. and Gay, L., 2018. How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Molecular Ecology Resources*, 18(2), 194-203.

Ronges, D., Walsh, J.P., Sinclair, B.J. and Stillman, J.H., 2012. Changes in extreme cold tolerance, membrane composition and cardiac transcriptome during the first day

of thermal acclimation in the porcelain crab *Petrolisthes cinctipes*. *Journal of Experimental Biology*, 215(11), 1824-1836.

Rooke, J., Pan, D., Xu, T. and Rubin, G.M., 1996. KUZ, a conserved metalloprotease-disintegrin protein with two roles in Drosophila neurogenesis. *Science*, 273(5279), 1227-1231.

Ropars, L., Dajoz, I., Fontaine, C., Muratet, A. and Geslin, B., 2019. Wild pollinator activity negatively related to honey bee colony densities in urban context. *PloS One*, 14(9), 0222316.

Rortais, A., Arnold, G., Alburaki, M., Legout, H. and Garnery, L., 2011. Review of the DraI COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). *Conservation Genetics Resources*, 3(2), 383-391.

Rosenblum, E.B., Römpler, H., Schöneberg, T. and Hoekstra, H.E., 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences*, 107(5), 2113-2117.

Rosenkranz, P., Aumeier, P. and Ziegelmann, B., 2010. Biology and control of *Varroa destructor*. *Journal of Invertebrate Pathology*, 103, 96-119.

Ruegg, K., Anderson, E.C., Boone, J., Pouls, J. and Smith, T.B., 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, 23(19), 4757-4769.

Ruottinen, L., Berg, P., Kantanen, J., Kristensen, T.N., Praebel, A. and Groeneveld, L.F., 2014. Status and conservation of the Nordic brown bee. Available at: https://vbn.aau.dk/ws/files/207960984/Ruottinen_et_al_2014.pdf Last accessed: 21th July 2022.

Ruttner, F. 1988. Biogeography and Taxonomy of Honey bees. *Springer-Verlag*.

Ruttner, F., Tassencourt, L. and Louveaux, J., 1978. Biometrical-statistical analysis of the geographic variability of *Apis mellifera*. *Apidologie*, 9(4), 363-381.

Sabella, G., Mulè, R., Robba, L., Agrò, A. and Manachini, B., 2022. Could Europe Apply a Suitable Control Method for the Small Hive Beetle (Coleoptera: *Nitidulidae*)? *Journal of Economic Entomology*, 115(2), 401-411.

Sainsbury, K.A., Shore, R.F., Schofield, H., Croose, E., Campbell, R.D. and Mcdonald, R.A., 2019. Recent history, current status, conservation and management of native mammalian carnivore species in Great Britain. *Mammal Review*, 49(2), 171-188.

Saitoh, K., Chen, W.J. and Mayden, R.L., 2010. Extensive hybridization and tetrapolyploidy in spined loach fish. *Molecular Phylogenetics and Evolution*, 56(3), 1001-1010.

Salisbury, A., Armitage, J., Bostock, H., Perry, J., Tatchell, M. and Thompson, K., 2015. Editors Choice: Enhancing gardens as habitats for flower-visiting aerial insects (pollinators): should we plant native or exotic species?. *Journal of Applied Ecology,* 52(5), 1156-1164.

Salisbury, S.J., Delgado, M.L. and Dalziel, A.C., 2021. Alternative splicing: An overlooked mechanism contributing to local adaptation?

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, M.E., Smith, T.J. and Rader, R., 2018. Bee conservation: Key role of managed bees. *Science*, 360(6387), 389-389.

Schmidt, L.S., Schmidt, J.O., Rao, H., Wang, W. and Xu, L., 1995. Feeding preference and survival of young worker honey bees (Hymenoptera: Apidae) fed rape, sesame, and sunflower pollen. *Journal of Economic Entomology*, 88(6), 1591-1595.

Schneider, J., Atallah, J. and Levine, J.D., 2017. Social structure and indirect genetic effects: genetics of social behaviour. *Biological Reviews*, 92(2), 1027-1038.

Sheppard, W.S., Arias, M.C., Grech, A. and Meixner, M.D., 1997. *Apis mellifera ruttneri*, a new honey bee subspecies from Malta. *Apidologie*, *28*(5), 287-293.

Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T. and Chang, W., 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, 194.

Slatkin, M., 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6), 477-485.

Smeds, L., Aspi, J., Berglund, J., Kojola, I., Tirronen, K. and Ellegren, H., 2021. Whole-genome analyses provide no evidence for dog introgression in Fennoscandian wolf populations. *Evolutionary Applications*, 14(3), 721-734.

Smith, R.M., Baker, R.H., Malumphy, C.P., Hockland, S., Hammon, R.P., Ostojá-Starzewski, J.C. and Collins, D.W., 2007. Recent non-native invertebrate plant pest establishments in Great Britain: origins, pathways, and trends. *Agricultural and Forest Entomology*, 9(4), 307-326.

Smout, T.C., 2014. What's natural: a species history of Scotland in the last 10,000 years. *Glasgow Naturalist,* 26(1), 11-16.

Sokolowski, M.B., 2020. Honey bee colony aggression and indirect genetic effects. *Proceedings of the National Academy of Sciences*, 117(31), 18148-18150.

Soland-Reckeweg, G., Heckel, G., Neumann, P., Fluri, P. and Excoffier, L., 2009. Gene flow in admixed populations and implications for the conservation of the Western honeybee, *Apis mellifera*. *Journal of Insect Conservation*, 13(3), 317-328.

Solignac, M., Vautrin, D., Loiseau, A., Mougel, F., Baudry, E., Estoup, A., Garnery, L., Haberl, M. and Cornuet, J.M., 2003. Five hundred and fifty microsatellite markers for the study of the honeybee (*Apis mellifera* L.) genome. *Molecular Ecology Notes*, 3(2), 307-311.

Sork, V.L., Aitken, S.N., Dyer, R.J., Eckert, A.J., Legendre, P. and Neale, D.B., 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, 9(4), 901-911.

Soyano, K. and Mushirobira, Y., 2018. The Mechanism of Low-Temperature Tolerance in Fish. Survival Strategies in Extreme Cold and Desiccation: *Adaptation Mechanisms and Their Applications*, 1081, 149.

Spötter, A., Gupta, P., Mayer, M., Reinsch, N. and Bienefeld, K., 2016. Genome-wide association study of a *Varroa*-specific defense behavior in honey bees (*Apis mellifera*). Journal of Heredity, 107(3), 220-227.

Stapley, J., Reger, J., Feulner, P.G., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P. and Slate, J., 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, 25(12), 705-712.

Steffan-Dewenter, I. and Tscharntke, T., 2000. Resource overlap and possible competition between honey bees and wild bees in central Europe. *Oecologia*, 122(2), 288-296.

Straw, N.A., Fielding, N.J., Tilbury, C., Williams, D.T. and Inward, D., 2015. Host plant selection and resource utilisation by Asian longhorn beetle *Anoplophora glabripennis* (Coleoptera: *Cerambycidae*) in southern England. *Forestry: An International Journal of Forest Research*, 88(1), 84-95.

Strickland, S.S., 1982. Honey hunting by the Gurungs of Nepal. *Bee World*, 63(4), 153-161.

Stukenbrock, E.H. and Dutheil, J.Y., 2018. Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics*, 208(3), 1209-1229.

Suarez-Gonzalez, A., Lexer, C. and Cronk, Q.C., 2018. Adaptive introgression: a plant perspective. *Biology Letters*, 14(3), 20170688.

Tagliotti, M.E., Deperi, S.I., Bedogni, M.C. and Huarte, M.A., 2021. Genome-wide association analysis of agronomical and physiological traits linked to drought tolerance in a diverse potatoes (*Solanum tuberosum*) panel. *Plant Breeding*, 140(4), 654-664.

Takai, Y., Sasaki, T. and Matozaki, T., 2001. Small GTP-binding proteins. *Physiological Reviews*, 81(1), 153-208.

Tantillo, G., Bottaro, M., Di Pinto, A., Martella, V., Di Pinto, P. and Terio, V., 2015. Virus infections of honey bees *Apis mellifera*. Italian journal of food safety, 4(3).

Tarpy, D.R. and Seeley, T.D., 2006. Lower disease infections in honeybee (*Apis mellifera*) colonies headed by polyandrous vs monandrous queens. *Naturwissenschaften*, 93(4), 195-199.

Taylor, H.R., 2015. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, 5(15), 3140-3150.

Templeton, A.R., Hemmer, H., Mace, G., Seal, U.S., Shields, W.M. and Woodruff, D.S., 1986. Local adaptation, coadaptation, and population boundaries. *Zoo Biology*, 5(2), 115-125.

The Bee Keeping Order, 2013. The Bee Keeping (Colonsay and Oronsay) Order 2013, Scottish Statutory Instruments. Available at: http://www.legislation.gov.uk/ssi/2013/279/contents/made Last accessed 21st July 2022.

The National Bee Unit Managing *Varroa*, 2020. The Animal and Plant Health Agency, Avalible at: https://nationalbeeunit.com/downloadDocument.cfm?id=1389 Last Acessed 25th of July 2022

Thibaut Jombart, Caitlin Collins., 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. MRC Centre for Outbreak Analysis and Modelling. PDF. Available at https://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf

Thompson, C.E., Biesmeijer, J.C., Allnutt, T.R., Pietravalle, S. and Budge, G.E., 2014. Parasite pressures on feral honey bees (*Apis mellifera* sp.). *PloS one*, 9(8), 105164.

Thompson, H.M., 2003. Behavioural effects of pesticides in bees–their potential for use in risk assessment. *Ecotoxicology*, 12(1), 317-330.

Tigano, A. and Friesen, V.L., 2016. Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10), 2144-2164.

Tihelka, E., Cai, C., Pisani, D. and Donoghue, P.C., 2020. Mitochondrial genomes illuminate the evolutionary history of the Western honey bee (*Apis mellifera*). Scientific reports, 10(1), 1-10.

Tizzani, P., Boano, G., Mosso, M., Pelazza, M., Carolfi, S., Ferra, M., Marletta, N., Pio, G.P.I., Meneguz, P.G., Silvano, F. and Negri, E., 2013. Recent distribution of Red-legged Partridge *Alectoris rufa* in Piedmont (North Western Italy): signs of recent spreading.

Tofilski, A., 2004. DrawWing, a program for numerical description of insect wings. *Journal of Insect Science*, 4(1). 17

Torné-Noguera, A., Rodrigo, A., Osorio, S. and Bosch, J., 2016. Collateral effects of beekeeping: Impacts on pollen-nectar resources and wild bee communities. *Basic and Applied Ecology*, 17(3), 199-209.

Toro, M.A., Fernández, J. and Caballero, A., 2009. Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120(3), 174-195.

Townsend, M., 2008. Report on Survey for Oak Processionary Moth *Thaumetopoea processionea* (Linnaeus) (Lepidoptera: *Thaumetopoeidae*)(OPM) in London in 2007. Report to the Forestry Commission, UK.

Turner, A., 2021. 'Honey bees are voracious; is it time to put the brakes on the boom in bee keeping', The Guardian, London, 24th July. Available at: https://www.theguardian.com/environment/2021/jul/24/this-only-saves-honeybees-the-trouble-with-britains-beekeeping-boom-aoe last Accessed 19th July 2022.

Turner, J.R., 1981. Adaptation and evolution in *Heliconius*: a defense of NeoDarwinism. *Annual Review of Ecology and Systematics*, 12(1), 99-121.

Upadhyay, M., Kunz, E., Sandoval-Castellanos, E., Hauser, A., Krebs, S., Graf, A., Blum, H., Dotsev, A., Okhlopkov, I., Shakhin, A. and Bagirov, V., 2021. Whole genome

sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Molecular Ecology*, 30(24), 6701-6717.

Utaipanon, P., Holmes, M.J., Chapman, N.C. and Oldroyd, B.P., 2019. Estimating the density of honey bee (*Apis mellifera*) colonies using trapped drones: area sampled and drone mating flight distance. *Apidologie,* 50(4), 578-592.

Uzunov, A., Brascamp, E.W. and Büchler, R., 2017. The basic concept of honey bee breeding programs. *Bee World*, 94(3), 84-87.

Vähä, J.P. and Primmer, C.R., 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, 15(1), 63-72.

Van der Auwera, G.A. and O'Connor, B.D., 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J. and Banks, E., 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current protocols in bioinformatics, 43(1), 11-10.

Van der Have, T.M. and De Jong, G., 1996. Adult size in ectotherms: temperature effects on growth and differentiation. *Journal of Theoretical Biology*, 183(3), 329-340.

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

Vanbergen, A. J., Heard, M. S., Breeze, T. D., Potts, S. G. & Hanley, N. Status and Value of Pollinators and Pollination Services (Defra, UK, 2014). Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/210926/pb13981-bees-pollinators-review.pdf Last Accessed: 21st July 2022.

VanEngelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y. and Underwood, R., 2009. Colony collapse disorder: a descriptive study. *PloS one*, 4(8), 6481.

Verslues, P.E., Lasky, J.R., Juenger, T.E., Liu, T.W. and Kumar, M.N., 2014. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis. Plant Physiology*, 164(1), 144-159.

Vince Gaffney, Simon Fitch & David Smith, 2009. Europe's Lost World: the Rediscovery of Doggerland. York: Council for British Archaeology. (CBA Research Report 160.); ISBN 978-1-902771-77-9, Cambridge Archaeological Journal, 20(1), 145-146.

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.

Wallace, D.C., 2007. Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annual Review of Biochemistry*, 76(1), 781-821.

Wallberg, A., Schoening, C., Webster, M.T. and Hasselmann, M., 2017. Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genetics*, 13(5), 1006792.

Walsh, J., Kovach, A.I., Olsen, B.J., Shriver, W.G. and Lovette, I.J., 2018. Bidirectional adaptive introgression between two ecologically divergent sparrow species. *Evolution*, 72(10), 2076-2089.

Wang, H., Jin, S., Chen, X., Gen, X. and He, Y., 2012. Target deletion of the AAA ATPase PpCDC48II in *Physcomitrella* patens results in freezing sensitivity after cold acclimation. *Science China Life Sciences*, 55(2), 150-157.

Wang, J., 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, 11(1), 141-145.

Waples R.S, Hendry A.P., 2008 Evolutionary perspectives on salmonid conservation and management. *Evolutionary Applications*, 1, 183–188.

Webb, D.A., 1985. What are the criteria for presuming native status?. *Watsonia*, 15(3), 231-236.

Whitfield, C.W., Behura, S.K., Berlocher, S.H., Clark, A.G., Johnston, J.S., Sheppard, W.S., Smith, D.R., Suarez, A.V., Weaver, D. and Tsutsui, N.D., 2006. Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, *314*(5799), 642-645.

Whitney, K.D., Randell, R.A. and Rieseberg, L.H., 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist*, 167(6), 794-807.

Whitney, K.D., Randell, R.A. and Rieseberg, L.H., 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, 187(1), 230-239.

Wickham, H., 2016. Package 'ggplot2': elegant graphics for data analysis. *Springer-Verlag, 10*, 978.

Widdicombe, J. 2015 The Principles of Bee Improvement. Published in Mytholmroyd, West Yorkshire by Northern Bee Books.

Wilfert, L., Gadau, J. and Schmid-Hempel, P., 2007. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity,* 98(4), 189-197.

Wilkinson, P., 1987. Red-legged impostors. *British Trust 0rnithology News* 152, 1–2.

Wilkinson, P., 1991. Mystery photographs. *British Birds* 84, 264–265

Williams, G.C., 2018. Adaptation and natural selection: A critique of some current evolutionary thought, *Princeton university press*, 75.

Williams, G.R., Tarpy, D.R., Chauzat, M.P., Cox-Foster, D.L., Delaplane, K.S., Neumann, P., Pettis, J.S., Rogers, R.E. and Shutler, D., 2010. Colony collapse disorder in context. *Bioessays*, *32*(10), 845.

Winger, B.M., 2017. Consequences of divergence and introgression for speciation in Andean cloud forest birds. *Evolution,* 71(7), 1815-1831.

Wolf, J.B., Brodie III, E.D., Cheverud, J.M., Moore, A.J. and Wade, M.J., 1998. Evolutionary consequences of indirect genetic effects. Trends in Ecology & Evolution, 13(2), 64-69.

Woodward, S. and Boa, E., 2013. Ash dieback in the UK: a wake-up call. *Molecular Plant Pathology*, *14*(9), 856.

Wragg, D., Marti-Marimon, M., Basso, B., Bidanel, J.P., Labarthe, E., Bouchez, O., Le Conte, Y. and Vignal, A., 2016. Whole-genome resequencing of honeybee drones to

detect genomic selection in a population managed for royal jelly. *Scientific reports*, 6(1), 1-13.

Wragg, D., Marti-Marimon, M., Basso, B., Bidanel, J.P., Labarthe, E., Bouchez, O., Le Conte, Y. and Vignal, A., 2016. Whole-genome resequencing of honey bee drones to detect genomic selection in a population managed for royal jelly. *Scientific Reports*, 6(1), 1-13.

Wright, C.M., Keiser, C.N. and Pruitt, J.N., 2016. Colony personality composition alters colony-level plasticity and magnitude of defensive behaviour in a social spider. *Animal Behaviour*, 115, 175-183.

Wu, C.I., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6), 851-865.

Wu, J.Y., Anelli, C.M. and Sheppard, W.S., 2011. Sub-lethal effects of pesticide residues in brood comb on worker honey bee (*Apis mellifera*) development and longevity. *PloS One*, 6(2), 14720.

Xu, Y., Lin, Z., Tang, C., Tang, Y., Cai, Y., Zhong, H., Wang, X., Zhang, W., Xu, C., Wang, J. and Wang, J., 2019. A new massively parallel nanoball sequencing platform for whole exome research. *BMC Bioinformatics*, 20(1), 1-9.

Yang, F., Zhao, R., Fang, X., Huang, H., Xuan, Y., Ma, Y., Chen, H., Cai, T., Qi, Y. and Xi, R., 2015. The RNA surveillance complex Pelo-H bs1 is required for transposon silencing in the *Drosophila* germline. EMBO Reports, 16(8), 965-974.

Zakour, M.K. and Bienefeld, K., 2014. Basic considerations in the development of breeding plans for honey bees, illustrated by data on the native Syrian honey bee (*Apis mellifera syriaca*). *Journal of Apicultural Research*, 53(2), 314-326.

Zayed, A. and Whitfield, C.W., 2008. A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences*, *105*(9), 3421-3426.

Zeuner, F.E. and FJ, Manning., 1976. A monograph on fossil bees (*Hymenoptera: Apoidea*). *Bulletins of the British Museum of Natural History*. 27(3), 151-268

Zeven, A.C., 1998. Landraces: a review of definitions and classifications. *Euphytica*, 104(2), 127-139.

Zhang, H.Y., Zhao, Z.X., Xu, J., Xu, P., Bai, Q.L., Yang, S.Y., Jiang, L.K. and Chen, B.H., 2018. Population genetic analysis of aquaculture salmonid populations in China using a 57K rainbow trout SNP array. *PloS One*, 13(8), 0202582.

Zhou, T., Ko, E.A., Gu, W., Lim, I., Bang, H. and Ko, J.H., 2012. Non-silent story on synonymous sites in voltage-gated ion channel genes. *PloS One*, 7(10), 48541.

Zieliński, P., Nadachowska-Brzyska, K., Wielstra, B., Szkotak, R., Covaciu-Marcov, S.D., Cogălniceanu, D. and Babik, W., 2013. No evidence for nuclear introgression despite complete mt DNA replacement in the C arpathian newt (L issotriton montandoni). *Molecular Ecology,* 22(7), 1884-1903.

Zimmerman, S.J., Aldridge, C.L. and Oyler-McCance, S.J., 2020. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics*, 21(1), 1-16.

8. Appendix

8.1 Appendix 1: Supplementary tables for chapter 3

Table 8.1. ADMIXTURE results for the Pooled colony ddRADseq at K=2. All samples were generated in this study.

Subspecies	Sample origin	Data generated by	Sample ID	Q value	
				C-lineage cluster	M-lineage cluster
Unknown	South West	This study	c1	0.28	0.72
Unknown	South West	This study	c2	0.40	0.60
Unknown	South West	This study	c3	0.26	0.74
Unknown	South West	This study	c4	0.27	0.73
Unknown	South West	This study	c5	0.35	0.65
Unknown	South West	This study	c6	0.26	0.74
Unknown	South West	This study	c7	0.20	0.80
Unknown	South West	This study	c8	0.27	0.73
Unknown	South West	This study	c9	0.24	0.76
Unknown	South West	This study	c10	0.58	0.42
Unknown	South West	This study	c11	0.38	0.62
Unknown	South West	This study	c12	0.30	0.70
Unknown	South West	This study	c13	0.47	0.53
Unknown	South West	This study	c14	0.28	0.72
Unknown	South West	This study	c15	0.47	0.53
Unknown	South West	This study	c16	0.49	0.51
Unknown	South West	This study	c17	0.27	0.73
Unknown	South West	This study	c18	0.24	0.76
Unknown	South West	This study	c19	0.57	0.43
Unknown	South West	This study	c20	0.52	0.48
Unknown	South West	This study	c21	0.64	0.36
Unknown	South West	This study	c22	0.82	0.18
Unknown	South West	This study	c23	0.20	0.80
Unknown	South West	This study	c24	0.31	0.69
Unknown	South West	This study	c25	0.20	0.80
Unknown	South West	This study	c27	0.35	0.65
Unknown	South West	This study	c29	0.26	0.74
Unknown	South West	This study	c30	0.30	0.70
A. m. carnica	Germany	This study	6	0.96	0.04
A. m. carnica	Sweden	This study	60	0.96	0.04

A. m. carnica	Sweden	This study	62	1.00	0.00
A. m. carnica	Sweden	This study	63	0.83	0.17
A. m. carnica	Germany	This study	7	0.97	0.03
A. m. carnica	Sweden	This study	48	1.00	0.00
A. m. carnica	Germany	This study	5	1.00	0.00
A. m. carnica	Germany	This study	AMC2	0.95	0.05
A. m. carnica	Sweden	This study	46	1.00	0.00
A. m. carnica	Sweden	This study	39	0.94	0.06
A. m. carnica	Germany	This study	4	0.62	0.38
A. m. carnica	Sweden	This study	35	0.93	0.07
A. m. carnica	Sweden	This study	29	0.86	0.14
A. m. carnica	Germany	This study	3	1.00	0.00
A. m. carnica	Sweden	This study	23	0.90	0.10
A. m. carnica	Germany	This study	AMC1	0.98	0.02
A. m. carnica	Germany	This study	2	1.00	0.00
A. m. carnica	Norway	This study	163	0.83	0.17
A. m. carnica	Norway	This study	156	0.99	0.01
A. m. carnica	Norway	This study	157	1.00	0.00
A. m. carnica	Norway	This study	158	0.99	0.01
A. m. carnica	Norway	This study	151	0.98	0.02
A. m. carnica	Germany	This study	1	0.99	0.01
A. m. carnica	Norway	This study	142	0.96	0.04
A. m. carnica	Norway	This study	143	0.98	0.02
A. m. carnica	Norway	This study	137	0.97	0.03
A. m. carnica	Norway	This study	138	0.72	0.28
A. m. carnica	Norway	This study	139	0.59	0.41
A. m. ligustica	Italy	This study	10	1.00	0.00
A. m. ligustica	Italy	This study	12	1.00	0.00
A. m. ligustica	Italy	This study	13	1.00	0.00
A. m. ligustica	Italy	This study	8	0.85	0.15
A. m. ligustica	Italy	This study	9	1.00	0.00
A. m. ligustica	Italy	This study	AML_2	1.00	0.00
A. m. ligustica	Italy	This study	AML2	1.00	0.00
A. m. ligustica	Sweden	This study	26	0.81	0.19
A. m. ligustica	Sweden	This study	58	0.70	0.30
A. m. ligustica	Sweden	This study	54	0.80	0.20
A. m. ligustica	Sweden	This study	27	0.78	0.22
A. m. ligustica	Sweden	This study	50	0.80	0.20
A. m. ligustica	Sweden	This study	43	0.73	0.27
A. m. ligustica	Sweden	This study	45	0.89	0.11
A. m. ligustica	Sweden	This study	41	0.79	0.21
A. m. ligustica	Sweden	This study	34	1.00	0.00
A. m. ligustica	Sweden	This study	21	1.00	0.00
A. m. ligustica	Sweden	This study	22	0.82	0.18
A. m. mellifera	Norway	This study	136	0.00	1.00

A. m. mellifera	Switzerland	This study	AMM2	0.00	1.00
A. m. mellifera	Switzerland	This study	14	0.21	0.79
A. m. mellifera	Norway	This study	140	0.11	0.89
A. m. mellifera	Norway	This study	144	0.26	0.74
A. m. mellifera	Switzerland	This study	15	0.20	0.80
A. m. mellifera	Norway	This study	150	0.00	1.00
A. m. mellifera	Norway	This study	152	0.27	0.73
A. m. mellifera	Norway	This study	153	0.00	1.00
A. m. mellifera	Switzerland	This study	16	0.18	0.82
A. m. mellifera	Norway	This study	162	0.00	1.00
A. m. mellifera	Norway	This study	164	0.00	1.00
A. m. mellifera	Switzerland	This study	18	0.18	0.82
A. m. mellifera	Switzerland	This study	20	0.17	0.83
A. m. mellifera	Sweden	This study	24	0.34	0.66
A. m. mellifera	Sweden	This study	28	0.00	1.00
A. m. mellifera	Sweden	This study	30	0.00	1.00
A. m. mellifera	Sweden	This study	38	0.03	0.97
A. m. mellifera	Sweden	This study	40	0.00	1.00
A. m. mellifera	Switzerland	This study	AMM1	0.00	1.00
A. m. mellifera	Sweden	This study	42	0.00	1.00
A. m. mellifera	Sweden	This study	47	0.00	1.00
A. m. mellifera	Sweden	This study	52	0.00	1.00
A. m. mellifera	Sweden	This study	55	0.00	1.00

Table 8.2. ADMIXTURE results for Individual ddRADseq at K=2. The majority of the subspecies standards for this data set were originally sequenced by Wallberg et al (2014) and were downloaded from the Sequence Read Archive under the project number for these data is PRJNA236426

Subspecies	Sample origin	Data generated by	Sample ID	Q Value	
				C-lineage cluster	M-lineage cluster
Unknown	South West	This study	c2	0.22	0.78
Unknown	South West	This study	c5	0.14	0.86
Unknown	South West	This study	c6	0.17	0.83
Unknown	South West	This study	c7	0.11	0.89
Unknown	South West	This study	c8	0.21	0.79
Unknown	South West	This study	c10	0.71	0.29
Unknown	South West	This study	c11	0.23	0.77
Unknown	South West	This study	c12	0.37	0.63
Unknown	South West	This study	c13	0.27	0.73

Unknown	South West	This study	c14	0.26	0.74
Unknown	South West	This study	c16	0.57	0.43
Unknown	South West	This study	c17	0.15	0.85
Unknown	South West	This study	c18	0.14	0.86
Unknown	South West	This study	c21	0.63	0.37
Unknown	South West	This study	c22	0.82	0.18
Unknown	South West	This study	c25	0.12	0.88
Unknown	South West	This study	c26	0.19	0.81
Unknown	South West	This study	c27	0.17	0.83
Unknown	South West	This study	c28	0.16	0.84
Unknown	South West	This study	c29	0.17	0.83
Unknown	South West	This study	c30	0.13	0.87
A. m. carnica	Austria	Wallberg *et al*	SAMN02596344	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596341	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596345	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596346	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596342	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596340	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596347	1.00	0.00
A. m. carnica	Austria	Wallberg *et al*	SAMN02596338	1.00	0.00
A. m. carnica	Germany	This study	amc33	0.95	0.05
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596289	0.98	0.02
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596288	0.95	0.05
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596297	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596295	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596290	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596293	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596294	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596292	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596296	1.00	0.00
A. m. ligustica	Italy	Wallberg *et al*	SAMN02596291	1.00	0.00
A. m. ligustica	Italy	This study	aml35	1.00	0.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596337	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596422	0.21	0.79
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596415	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596417	0.06	0.94
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596330	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596328	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596423	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596331	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596332	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596329	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596414	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596420	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596416	0.00	1.00

A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596418	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596421	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596334	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596335	0.00	1.00
A. m. mellifera	Sweden	Wallberg *et al*	SAMN02596419	0.00	1.00
A. m. mellifera	Norway	Wallberg *et al*	SAMN02596333	0.00	1.00
A. m. mellifera	Switzerland	This study	amm34	0.00	1.00

Table 8.3 ADMIXURE for SNP Array results at K=2. The subspecies standards for the SNP Array were generated by Pinto et al (2014) and the SNP Aray was developed by Munoz et al (2015) and Henriques et al (2018)

Subspecies	Sample Origin	Data generated by	Sample	Q Value	
				M-lineage cluster	C-lineage cluster
Unknown	South West, UK	This study	c1	0.81	0.19
Unknown	South West, UK	This study	c2	0.80	0.20
Unknown	South West, UK	This study	c3	0.86	0.14
Unknown	South West, UK	This study	c4	0.77	0.23
Unknown	South West, UK	This study	c5	0.79	0.21
Unknown	South West, UK	This study	c6	0.82	0.18
Unknown	South West, UK	This study	c7	0.87	0.13
Unknown	South West, UK	This study	c8	0.79	0.21
Unknown	South West, UK	This study	c9	0.78	0.22
Unknown	South West, UK	This study	c10	0.35	0.65
Unknown	South West, UK	This study	c11	0.83	0.17
Unknown	South West, UK	This study	c12	0.67	0.33
Unknown	South West, UK	This study	c13	0.74	0.26
Unknown	South West, UK	This study	c14	0.79	0.21
Unknown	South West, UK	This study	c15	0.81	0.19
Unknown	South West, UK	This study	c16	0.47	0.53
Unknown	South West, UK	This study	c17	0.85	0.15
Unknown	South West, UK	This study	c18	0.85	0.15
Unknown	South West, UK	This study	c19	0.63	0.37
Unknown	South West, UK	This study	c20	0.57	0.43
Unknown	South West, UK	This study	c21	0.37	0.63
Unknown	South West, UK	This study	c22	0.21	0.79
Unknown	South West, UK	This study	c23	0.86	0.14
Unknown	South West, UK	This study	c24	0.74	0.26
Unknown	South West, UK	This study	c25	0.84	0.16
Unknown	South West, UK	This study	1.2	0.74	0.26
Unknown	South West, UK	This study	1.6	0.80	0.20

Unknown	South West, UK	This study	1.11	0.83	0.17
Unknown	South West, UK	This study	2.1	0.58	0.42
Unknown	South West, UK	This study	2.4	0.84	0.16
Unknown	South West, UK	This study	2.35	0.90	0.10
Unknown	South West, UK	This study	3.2	0.81	0.19
Unknown	South West, UK	This study	3.3	0.63	0.37
Unknown	South West, UK	This study	3.5	0.77	0.23
Unknown	South West, UK	This study	3.6	0.82	0.18
Unknown	South West, UK	This study	3.7	0.80	0.20
Unknown	South West, UK	This study	4.2	0.55	0.45
Unknown	South West, UK	This study	5.1	0.86	0.14
Unknown	South West, UK	This study	6.2	0.55	0.45
Unknown	South West, UK	This study	7.1	0.46	0.54
Unknown	South West, UK	This study	8.2	0.80	0.20
Unknown	South West, UK	This study	8.4	0.77	0.23
Unknown	South West, UK	This study	10.1	0.71	0.29
Unknown	South West, UK	This study	10.3	0.83	0.17
Unknown	South West, UK	This study	11.2	0.49	0.51
Unknown	South West, UK	This study	11.3	0.49	0.51
Unknown	South West, UK	This study	13.2	0.76	0.24
Unknown	South West, UK	This study	13.3	0.84	0.16
Unknown	South West, UK	This study	13.4	0.84	0.16
Unknown	South West, UK	This study	13.7	0.76	0.24
Unknown	South West, UK	This study	13.8	0.81	0.19
Unknown	South West, UK	This study	13.9	0.83	0.17
Unknown	South West, UK	This study	14.1	0.94	0.06
Unknown	South West, UK	This study	14.2	0.93	0.07
Unknown	South West, UK	This study	14.4	0.87	0.13
Unknown	South West, UK	This study	15.1	0.35	0.65
Unknown	South West, UK	This study	16.1	0.86	0.14
Unknown	South West, UK	This study	16.2	0.50	0.50
Unknown	South West, UK	This study	17.2	0.73	0.27
Unknown	South West, UK	This study	18.1	0.71	0.29
Unknown	South West, UK	This study	18.2	0.62	0.38
Unknown	South West, UK	This study	18.3	0.23	0.77
Unknown	South West, UK	This study	19.1	0.65	0.35
Unknown	South West, UK	This study	19.2	0.47	0.53
Unknown	South West, UK	This study	19.4	0.77	0.23
Unknown	South West, UK	This study	20.1	0.64	0.36
Unknown	South West, UK	This study	20.2	0.83	0.17
Unknown	South West, UK	This study	21.2	0.84	0.16
Unknown	South West, UK	This study	21.3	0.74	0.26
Unknown	South West, UK	This study	21.4	0.83	0.17
Unknown	South West, UK	This study	21.6	0.85	0.15
Unknown	South West, UK	This study	21.7	0.84	0.16

Unknown	South West, UK	This study	22.2	0.42	0.58
Unknown	South West, UK	This study	22.3	0.61	0.39
Unknown	South West, UK	This study	22.4	0.69	0.31
Unknown	South West, UK	This study	22.5	0.56	0.44
Unknown	South West, UK	This study	24.1	0.82	0.18
Unknown	South West, UK	This study	24.2	0.79	0.21
Unknown	South West, UK	This study	24.5	0.79	0.21
Unknown	South West, UK	This study	24.7	0.73	0.27
Unknown	South West, UK	This study	24.8	0.65	0.35
Unknown	South West, UK	This study	24.16	0.80	0.20
Unknown	South West, UK	This study	25.1	0.89	0.11
Unknown	South West, UK	This study	25.2	0.78	0.22
Unknown	South West, UK	This study	25.3	0.82	0.18
Unknown	South West, UK	This study	25.5	0.67	0.33
Unknown	South West, UK	This study	25.7	0.78	0.22
Unknown	South West, UK	This study	41	0.71	0.29
Unknown	South West, UK	This study	42	0.80	0.20
Unknown	South West, UK	This study	43	0.81	0.19
Unknown	South West, UK	This study	44	0.51	0.49
Unknown	South West, UK	This study	45	0.67	0.33
Unknown	South West, UK	This study	46	0.77	0.23
Unknown	South West, UK	This study	51	0.76	0.24
Unknown	South West, UK	This study	53	0.70	0.30
Unknown	South West, UK	This study	54	0.83	0.17
Unknown	South West, UK	This study	57	0.75	0.25
Unknown	South West, UK	This study	58	0.68	0.32
A. m. ligustica	Italy	Pinto *et al*	2681	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2682	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2683	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2684	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2686	0.04	0.96
A. m. ligustica	Italy	Pinto *et al*	2687	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2688	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2692	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2695	0.00	1.00
A. m. ligustica	Italy	Pinto *et al*	2696	0.00	1.00
A. m. mellifera	France	Pinto *et al*	2698	0.41	0.59
A. m. mellifera	France	Pinto *et al*	2701	0.43	0.57
A. m. mellifera	France	Pinto *et al*	2702	0.55	0.45
A. m. mellifera	France	Pinto *et al*	2703	0.77	0.23
A. m. mellifera	France	Pinto *et al*	2704	1.00	0.00
A. m. carnica	Croatia	Pinto *et al*	2721	0.00	1.00
A. m. carnica	Croatia	Pinto *et al*	2723	0.00	1.00
A. m. carnica	Croatia	Pinto *et al*	2725	0.00	1.00
A. m. carnica	Croatia	Pinto *et al*	2726	0.00	1.00

A. m. carnica	Croatia	Pinto *et al*	2730	0.00	1.00
A. m. carnica	Serbia	Pinto *et al*	2735	0.02	0.98
A. m. carnica	Serbia	Pinto *et al*	2736	0.06	0.94
A. m. carnica	Serbia	Pinto *et al*	2738	0.00	1.00
A. m. carnica	Serbia	Pinto *et al*	2739	0.00	1.00
A. m. carnica	Serbia	Pinto *et al*	2740	0.00	1.00
A. m. mellifera	Denmark	Pinto *et al*	2744	0.77	0.23
A. m. mellifera	Denmark	Pinto *et al*	2745	0.70	0.30
A. m. mellifera	Denmark	Pinto *et al*	2746	0.80	0.20
A. m. mellifera	Denmark	Pinto *et al*	2747	0.89	0.11
A. m. mellifera	Denmark	Pinto *et al*	2748	0.95	0.05
A. m. mellifera	Denmark	Pinto *et al*	2756	0.97	0.03
A. m. mellifera	Denmark	Pinto *et al*	2757	0.91	0.09
A. m. mellifera	Denmark	Pinto *et al*	2758	0.85	0.15
A. m. mellifera	Denmark	Pinto *et al*	2760	0.82	0.18
A. m. mellifera	Netherlands	Pinto *et al*	2768	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2769	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2770	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2771	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2772	1.00	0.00
A. m. mellifera	Switzerland	Pinto *et al*	2773	0.92	0.08
A. m. mellifera	Switzerland	Pinto *et al*	2774	0.86	0.14
A. m. mellifera	Switzerland	Pinto *et al*	2775	0.87	0.13
A. m. mellifera	Switzerland	Pinto *et al*	2776	0.92	0.08
A. m. mellifera	Scotland	Pinto *et al*	2777	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2778	0.96	0.04
A. m. mellifera	Scotland	Pinto *et al*	2779	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2780	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2781	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2782	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2783	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2784	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2785	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2794	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2795	0.92	0.08
A. m. mellifera	Netherlands	Pinto *et al*	2796	1.00	0.00
A. m. mellifera	Switzerland	Pinto *et al*	2797	0.83	0.17
A. m. mellifera	Switzerland	Pinto *et al*	2798	0.95	0.05
A. m. mellifera	Norway	Pinto *et al*	2799	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2800	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2801	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2802	1.00	0.00
A. m. mellifera	Norway	Pinto *et al*	2803	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2811	1.00	0.00
A. m. mellifera	Netherlands	Pinto *et al*	2812	0.99	0.01

A. m. mellifera	Netherlands	Pinto *et al*	2813	0.97	0.03
A. m. mellifera	Netherlands	Pinto *et al*	2814	0.93	0.07
A. m. mellifera	Netherlands	Pinto *et al*	2815	0.99	0.01
A. m. mellifera	Netherlands	Pinto *et al*	2816	0.46	0.54
A. m. mellifera	Netherlands	Pinto *et al*	2817	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2820	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2821	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2822	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2823	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2824	1.00	0.00
A. m. mellifera	Scotland	Pinto *et al*	2825	1.00	0.00
A. m. mellifera	France	Pinto *et al*	2826	0.93	0.07
A. m. mellifera	France	Pinto *et al*	2827	1.00	0.00
A. m. mellifera	France	Pinto *et al*	2828	0.62	0.38
A. m. mellifera	France	Pinto *et al*	2829	0.86	0.14
A. m. mellifera	Belgium	Pinto *et al*	2830	0.95	0.05
A. m. mellifera	France	Pinto *et al*	2831	0.68	0.32
A. m. mellifera	Belgium	Pinto *et al*	2832	0.95	0.05
A. m. mellifera	Belgium	Pinto *et al*	2833	0.87	0.13
A. m. mellifera	France	Pinto *et al*	2834	1.00	0.00
A. m. mellifera	France	Pinto *et al*	2835	0.93	0.07
A. m. mellifera	France	Pinto *et al*	2836	0.94	0.06
A. m. mellifera	France	Pinto *et al*	2837	0.97	0.03
A. m. mellifera	France	Pinto *et al*	2838	0.98	0.02
A. m. mellifera	England	Pinto *et al*	2895	0.88	0.12
A. m. mellifera	England	Pinto *et al*	2896	0.76	0.24
A. m. mellifera	England	Pinto *et al*	2897	0.82	0.18
A. m. mellifera	England	Pinto *et al*	2898	0.86	0.14
A. m. mellifera	England	Pinto *et al*	2899	0.86	0.14
A. m. mellifera	England	Pinto *et al*	2900	0.62	0.38
A. m. mellifera	England	Pinto *et al*	2901	0.83	0.17
A. m. mellifera	England	Pinto *et al*	2902	0.87	0.13

Table A 1.4. Summary of ABBA BABA statistics for pooled colonies. Average D statistics (an indicator of the occurance of introgression) and admixture proportions (*f*) from 72 tests per colony for introgression from the two C lineage groups (36 tests for Mel:SW:Car and 36 tests for Mel:SW:Lig). The astirix represents colonies that were significant for all 36 tests for either *A. m. carnica* (Car) introgresison or *A. m. ligutsic*s (Lig) introgression.

ABBA BABA statistics for pooled colonies								
Colony Sample	Trios							
	Mel; South West ; Lig				Mel ; South West ; Car			
	Average D	SD	Average *f*	SD	Average D	SD	Average *f*	SD
c1	0.195*	0.019	0.149	0.012	0.167*	0.020	0.146	0.014
c2	0.202*	0.019	0.177	0.012	0.173*	0.019	0.172	0.015
c3	0.171*	0.019	0.122	0.012	0.147*	0.021	0.122	0.016
c4	0.119*	0.020	0.089	0.014	0.094	0.020	0.080	0.015
c5	0.091*	0.018	0.080	0.014	0.070	0.013	0.063	0.016
c6	0.091	0.016	0.065	0.013	0.077	0.006	0.074	0.016
c7	0.206*	0.021	0.144	0.012	0.182*	0.022	0.141	0.014
c8	0.131*	0.019	0.100	0.013	0.100*	0.020	0.087	0.015
c9	0.115*	0.020	0.084	0.013	0.084	0.160	0.063	0.016
c10	0.389*	0.015	0.447	0.009	0.352*	0.160	0.467	0.009
c26	0.107*	0.019	0.085	0.014	0.080	0.012	0.065	0.016
c11	0.299*	0.016	0.271	0.010	0.258*	0.018	0.264	0.013
c13	0.263*	0.015	0.247	0.011	0.233*	0.017	0.248	0.012
c14	0.173*	0.018	0.138	0.013	0.142*	0.019	0.132	0.016
c15	0.346*	0.014	0.327	0.009	0.315*	0.018	0.340	0.012
c30	0.188*	0.018	0.144	0.012	0.170*	0.020	0.147	0.014
c16	0.268*	0.014	0.263	0.010	0.236*	0.017	0.267	0.013
c27	0.213*	0.020	0.164	0.012	0.191*	0.020	0.173	0.014
c17	0.182*	0.020	0.133	0.012	0.153*	0.008	0.126	0.015
c18	0.084	0.019	0.061	0.014	0.066	0.016	0.042	0.016
c19	0.399*	0.014	0.452	0.008	0.366*	0.018	0.472	0.011
c20	0.326*	0.015	0.345	0.009	0.289*	0.017	0.351	0.013
c21	0.394*	0.012	0.478	0.008	0.362*	0.017	0.504	0.013
c29	0.084	0.014	0.059	0.013	0.062	0.007	0.043	0.016
c22	0.450*	0.016	0.629	0.009	0.415*	0.016	0.669	0.006
c23	0.114*	0.019	0.086	0.014	0.084	0.016	0.069	0.015
c24	0.140*	0.019	0.110	0.013	0.112*	0.019	0.100	0.015
c25	0.080	0.014	0.051	0.013	0.060	0.005	0.031	0.016

8.2 Appendix 2: Supplementary tables for chapter 4

Table 8.4 ADMIXTURE results for K=2 performed on pooled colony WGS data from across UK and Ireland.

Q values		Sample information		
A. m. mellifera	C-lineage	Sample code	Subspecies assignment	Approximate sample location
0.021	0.979	1	*A. m. carnica*	Germany
0.050	0.950	137	*A. m. carnica*	Norway
0.321	0.679	138	*A. m. carnica*	Norway
0.437	0.563	139	*A. m. carnica*	Norway
0.045	0.955	142	*A. m. carnica*	Norway
0.034	0.966	143	*A. m. carnica*	Norway
0.047	0.953	151	*A. m. carnica*	Norway
0.000	1.000	156	*A. m. carnica*	Norway
0.000	1.000	157	*A. m. carnica*	Norway
0.029	0.971	158	*A. m. carnica*	Norway
0.235	0.765	163	*A. m. carnica*	Norway
0.000	1.000	2	*A. m. carnica*	Germany
0.096	0.904	23	*A. m. carnica*	Sweden
0.082	0.918	29	*A. m. carnica*	Sweden
0.000	1.000	3	*A. m. carnica*	Germany
0.035	0.965	35	*A. m. carnica*	Sweden
0.069	0.931	39	*A. m. carnica*	Sweden
0.379	0.621	4	*A. m. carnica*	Germany
0.000	1.000	46	*A. m. carnica*	Sweden
0.000	1.000	48	*A. m. carnica*	Sweden
0.000	1.000	5	*A. m. carnica*	Germany
0.038	0.962	6	*A. m. carnica*	Sweden
0.026	0.974	60	*A. m. carnica*	Sweden
0.000	1.000	62	*A. m. carnica*	Sweden
0.138	0.862	63	*A. m. carnica*	Sweden
0.016	0.984	7	*A. m. carnica*	Germany
0.000	1.000	10	*A. m. ligustica*	Bologna Italy
0.000	1.000	12	*A. m. ligustica*	Bologna Italy
0.000	1.000	13	*A. m. ligustica*	Bologna Italy
0.124	0.876	8	*A. m. ligustica*	Bologna Italy
0.000	1.000	9	*A. m. ligustica*	Bologna Italy
0.000	1.000	21	*A. m. ligustica*	Sweden
0.176	0.824	22	*A. m. ligustica*	Sweden
0.170	0.830	26	*A. m. ligustica*	Sweden

0.194	0.806	27	*A. m. ligustica*	Sweden
0.000	1.000	34	*A. m. ligustica*	Sweden
0.177	0.823	41	*A. m. ligustica*	Sweden
0.251	0.749	43	*A. m. ligustica*	Sweden
0.103	0.897	45	*A. m. ligustica*	Sweden
0.141	0.859	50	*A. m. ligustica*	Sweden
0.180	0.820	54	*A. m. ligustica*	Sweden
0.273	0.727	58	*A. m. ligustica*	Sweden
1.000	0.000	136	*A. m. mellifera*	Norway
0.829	0.171	14	*A. m. mellifera*	Switzerland
0.863	0.137	140	*A. m. mellifera*	Norway
0.688	0.312	144	*A. m. mellifera*	Norway
0.540	0.460	147	*A. m. mellifera*	Norway
0.793	0.207	15	*A. m. mellifera*	Switzerland
1.000	0.000	150	*A. m. mellifera*	Norway
0.642	0.358	152	*A. m. mellifera*	Norway
1.000	0.000	153	*A. m. mellifera*	Norway
0.545	0.455	155	*A. m. mellifera*	Norway
0.819	0.181	16	*A. m. mellifera*	Switzerland
0.989	0.011	162	*A. m. mellifera*	Norway
1.000	0.000	164	*A. m. mellifera*	Norway
0.619	0.381	17	*A. m. mellifera*	Switzerland
0.805	0.195	18	*A. m. mellifera*	Switzerland
0.666	0.334	19	*A. m. mellifera*	Switzerland
0.836	0.164	20	*A. m. mellifera*	Switzerland
0.675	0.325	24	*A. m. mellifera*	Sweden
0.977	0.023	28	*A. m. mellifera*	Sweden
1.000	0.000	30	*A. m. mellifera*	Sweden
0.746	0.254	37	*A. m. mellifera*	Sweden
0.935	0.065	38	*A. m. mellifera*	Sweden
1.000	0.000	40	*A. m. mellifera*	Sweden
1.000	0.000	42	*A. m. mellifera*	Sweden
1.000	0.000	47	*A. m. mellifera*	Sweden
0.972	0.028	52	*A. m. mellifera*	Sweden
0.727	0.273	53	*A. m. mellifera*	Sweden
1.000	0.000	55	*A. m. mellifera*	Sweden
0.502	0.498	100	N/A	Dorset
0.609	0.391	101	N/A	Shropshire
0.291	0.709	102	N/A	Surrey
0.380	0.620	103	N/A	Linconshire
0.725	0.275	104	N/A	West Cornwall
0.877	0.123	105	N/A	Newcastle upon Tyne
0.751	0.249	106	N/A	Rame Penninsular
0.733	0.267	107	N/A	Tamar Valley
0.448	0.552	108	N/A	Middlesbrough

0.272	0.728	109	N/A	Isle of Wight
0.365	0.635	110	N/A	Sheffield
0.356	0.644	111	N/A	North Wales
0.619	0.381	112	N/A	Galway
0.014	0.986	113	N/A	Keswick
0.596	0.404	114	N/A	Norwich
0.509	0.491	115	N/A	Edinburgh
0.630	0.370	116	N/A	Hereford
0.472	0.528	117	N/A	Derby and Nottingham
0.418	0.582	118	N/A	Yeovil
0.456	0.544	119	N/A	Nottingham
0.471	0.529	120	N/A	Canterbury
0.672	0.328	121	N/A	Newcastle upon Tyne
0.626	0.374	122	N/A	Colchester
0.319	0.681	123	N/A	Reading
0.529	0.471	124	N/A	Scottish Boarders
0.000	1.000	125	N/A	Galway
0.458	0.542	126	N/A	Newlyn
0.627	0.373	127	N/A	Alladale
0.564	0.436	128	N/A	Aberdeen
0.626	0.374	129	N/A	Dundee
0.538	0.462	130	N/A	Scottish boarder
0.586	0.414	131	N/A	Bridgewater
0.558	0.442	132	N/A	Belfast
0.895	0.105	133	N/A	Jersey
0.394	0.606	134	N/A	Birmingham
0.570	0.430	64	N/A	Colchester
0.253	0.747	65	N/A	Oxford
0.528	0.472	66	N/A	Liskeard
0.000	1.000	67	N/A	Luton
0.444	0.556	68	N/A	Plymouth
0.508	0.492	69	N/A	Henley on Thames
0.657	0.343	70	N/A	Sheffield
0.494	0.506	71	N/A	Liskeard
0.395	0.605	72	N/A	Horsham
0.586	0.414	73	N/A	Isle of Man
0.928	0.072	74	N/A	Isle of Man
0.695	0.305	75	N/A	Bradford
0.671	0.329	76	N/A	Lizard penninsular
0.526	0.474	77	N/A	Burnley and Blackburn
1.000	0.000	78	N/A	Dublin
1.000	0.000	79	N/A	Colonsay
0.769	0.231	80	N/A	Northumberland
0.891	0.109	81	N/A	Belfast
0.495	0.505	82	N/A	York

0.510	0.490	83	N/A	Taunton
0.418	0.582	85	N/A	Birmingham
0.958	0.042	86	N/A	Co Monaghan
0.671	0.329	87	N/A	Abagaveny
0.953	0.047	88	N/A	Co Kildare
1.000	0.000	89	N/A	Inverness
0.918	0.082	90	N/A	Jersey
0.973	0.027	91	N/A	Co Wicklow
0.554	0.446	92	N/A	Cotswolds
0.244	0.756	93	N/A	Plymouth
0.707	0.293	94	N/A	Bradford
0.412	0.588	95	N/A	Dorcester
1.000	0.000	96	N/A	Co tipperary
0.611	0.389	97	N/A	Snodownia
0.162	0.838	98	N/A	Stratford-upon-avon
0.013	0.987	99	N/A	Leeds

Table 8.5 ABBA BABA results for the UK and Ireland colonies. Average *f* statistics proportion of admixture reported for all colonies. Two trios are presented. The first Mel: UK and Ire; Lig, where the *f* proportion represented *A. m. ligustica* introgression, and Mel: UK or Ire; Car, where the *f* proportion represents *A. m. carnica*

Sample approximate location	Colony ID	Trios			
		Mel; UK or Ire ; Lig		Mel ; UK or Ire ; Car	
		Average *f*	SD	Average *f*	SD
Colchester	64	0.224	0.012	0.238	0.017
Oxford	65	0.687	0.006	0.805	0.008
Liskeard	66	0.318	0.010	0.364	0.014
Luton	67	0.769	0.004	0.943	0.003
Plymouth	68	0.457	0.008	0.536	0.011
Henley on Thames	69	0.197	0.014	0.199	0.021
Sheffield	70	0.155	0.013	0.178	0.017
Liskeard	71	0.276	0.011	0.298	0.016
Horsham	72	0.279	0.011	0.293	0.023
Isle of Man	73	0.187	0.012	0.186	0.022
Isle of Man	74	-0.003	0.015	-0.010	0.022
Rame penninsular	75	0.226	0.012	0.266	0.016
Bradford	76	0.122	0.013	0.139	0.018
Lizard penninsular	77	0.356	0.010	0.422	0.013
Burnley	78	-0.130	0.018	-0.174	0.026
Dublin	79	0.002	0.015	0.003	0.021
Colonsay	80	0.110	0.013	0.129	0.018
Belfast	81	-0.041	0.016	-0.065	0.024
York	82	0.189	0.013	0.188	0.023
Taunton	83	0.319	0.010	0.370	0.014

Birmingham	85	0.251	0.012	0.272	0.017
Co Monaghan	86	-0.021	0.015	-0.034	0.022
Nr Abagaveny	87	0.084	0.015	0.066	0.021
Co Kildare	88	-0.124	0.018	-0.173	0.027
Inverness	89	-0.160	0.019	-0.221	0.030
Jersey	90	-0.009	0.015	-0.026	0.022
Co Wicklow	91	0.012	0.015	0.016	0.021
Cotswolds	92	0.073	0.015	0.051	0.025
Plymouth	93	0.576	0.007	0.686	0.009
Bradford	94	0.097	0.014	0.100	0.020
Dorcester	95	0.242	0.012	0.257	0.018
Co tipperary	96	-0.124	0.017	-0.168	0.026
Snodownia	97	0.143	0.013	0.147	0.020
Stratford-upon-avon	98	0.755	0.008	0.896	0.010
Leeds	99	0.885	0.005	1.081	0.005
Sherborne	100	0.364	0.010	0.435	0.013
Shropshire hills	101	0.260	0.011	0.289	0.018
Guilford	102	0.500	0.008	0.574	0.013
North of Lincon	103	0.385	0.009	0.430	0.016
Treluswell	104	0.082	0.014	0.081	0.020
Newcastle upon Tyne	105	0.038	0.014	0.038	0.020
Northumberland	106	0.039	0.015	0.028	0.021
Tamar Valley	107	-0.002	0.016	-0.037	0.026
Darlington	108	0.372	0.011	0.441	0.013
Isle of wight	109	0.501	0.008	0.569	0.011
Sheffield	110	0.551	0.007	0.655	0.010
North Wales	111	0.559	0.007	0.676	0.022
Galway	112	0.091	0.014	0.069	0.022
Keswick	113	0.932	0.007	1.142	0.007
Norwich	114	0.153	0.013	0.154	0.019
Edinburgh	115	0.249	0.011	0.282	0.017
Hereford	116	0.055	0.016	0.017	0.027
Derby	117	0.287	0.011	0.295	0.019
Yeovil	118	0.481	0.008	0.573	0.010
Nottingham	119	0.443	0.009	0.525	0.012
Canterbury	120	0.264	0.011	0.299	0.016
Newcastle upon Tyne	121	0.118	0.013	0.125	0.019
Colchester	122	0.063	0.015	0.034	0.024
Reading	123	0.544	0.007	0.642	0.009
Scottish Boarders	124	0.175	0.012	0.202	0.018
Galway	125	0.807	0.004	0.996	0.004
Newlyn	126	0.405	0.009	0.474	0.011
Alladale	127	0.185	0.012	0.216	0.017
Aberdeen	128	0.180	0.012	0.198	0.017
Dundee	129	0.135	0.013	0.140	0.019
Scottish boarder	130	0.223	0.012	0.265	0.016
Bridgewater	131	0.291	0.011	0.338	0.015

Belfast	132	0.250	0.011	0.311	0.015
Jersey	133	0.012	0.015	0.005	0.021
Birmingham	134	0.473	0.008	0.537	0.010