DOCTOR OF PHILOSOPHY

Investigating the foraging preferences of the honeybee, Apis mellifera L., using DNA metabarcoding

Jones, Laura

Award date:
2020

Awarding institution:
Bangor University

Investigating the foraging preferences of the honeybee, *Apis mellifera* L., using DNA metabarcoding

Laura Jones

2019

A thesis submitted for the degree of Doctor of Philosophy

Bangor University

PRIFYSGOL BANGOR UNIVERSITY

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: ……...

Title: ……..……..

Supervisor/Department: .....

Funding body (if any): .....

Qualification/Degree obtained: ……..

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed …….. (candidate)

Date ……..

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed …….. (candidate)

Date ……..

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed …….. (candidate)

Date ……..

NB: Candidates on whose behalf a bar on access has been approved by the Academic Registry should use the following version of **Statement 3:**

Statement 3 (bar):

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loans and for electronic storage (subject to any constraints as defined in statement 4), after expiry of a bar on access.

Signed …….. (candidate)

Date ……..

Statement 4:

Choose **one** of the following options

a)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	
b)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c)	I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.

2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.

3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.

4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.

5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.

6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: …… Date : ……..

Summary

Decreasing floral resources as a result of habitat loss and fragmentation is one of the key factors in the decline of pollinating insects worldwide, with the resulting impact on food supply and biodiversity receiving increasing global concern. A detailed understanding of which plants honeybees choose to use throughout the season is lacking in the literature and is vital to elucidating honeybee foraging behaviour, and to provide effective recommendations for suitable forage to support hives. Here, using DNA metabarcoding, honeybee foraging has been characterised throughout the season for hives set within a diverse landscape, revealing the plants that are most important to honeybees when offered a wide variety of species. To set this detailed foraging study within a wider UK context, honeybee foraging is then assessed for hives across the UK for the first time since 1952 and in doing so we evidence national scale changes in nectar provision over time. The results of this project provide scientific evidence to support beekeepers, as well as informing landscape level decisions in providing and improving habitat for pollinators. To support this work from a strong knowledge base and provide a high-level of confidence in the taxonomic identification of this and future DNA metabarcoding data, a reference library is created for the flowering plants and conifers of the UK, for three DNA barcode loci. Species coverage and discrimination is assessed for both the native plants alone and when naturalised and horticultural species are included. This work provides a high-quality resource for the honey DNA metabarcoding research presented here, as well as the many current and future applications of plant identification using DNA barcoding.

Acknowledgements

Firstly, thank you to my supervisor Natasha de Vere who developed this project and trusted me to take it on. Your encouragement, advice and support over my career has been invaluable. Thank you to my supervisor Simon Creer for the support from Bangor University, our meetings always encouraged me and made me feel on track.

Thank you to everyone who has been a part of the Science department at the National Botanic Garden of Wales. You have made the Botanic Garden the best place to work. Thank you to Tegan Gilmore, Abigail Lowe, and Jake Moscrop for helping with the surveying set-up. Thank you to the team of conservation volunteers and successive generations of placement students who have helped me with recording all of the plants in flower. Thank you to Abigail Lowe for coming back to the Garden as a PhD researcher, helping to complete the data entry and for supporting the running of the department. Thank you to Col Ford for patiently teaching me Python and always being prepared to help me troubleshoot. Thank you so much to Georgina Brennan for the statistical supervision and support.

I am very grateful to beekeeper Lynda Christie and her team of volunteers Ele Lloyd, Sonia Pritchard, Julian Caruana and more, who have expertly managed the ever-expanding set of apiaries at the Botanic Garden. This work would not have been possible without you. Thank you to Marilyn Caruana and all of the Stitching Botanical team for the artistic interpretations of the research. Thank you to all of the beekeepers across the UK and Ireland, both for donating their honey but also for being incredibly supportive and patient in waiting for the results. Thank you to all of the students, including Lucy Bidgood, Alice Hope and Louisa Smith, who helped with keeping track of the honey samples once they started filling up our post.

This PhD was funded by a Knowledge Economy Skills Scholarship with the National Botanic Garden of Wales as a partner. Additional funding for sequencing was kindly bequeathed by Quentin Kay.

CONTENTS

List of Figures

List of Tables

List of Supplementary Figures

CHAPTER 1

Introduction

1.1. Introduction

Honeybees (*Apis mellifera* L.) have been managed by humans for millennia, with the earliest evidence of hive beekeeping dating from Egypt c. 2400 BC (Crane, 1999). Honeybees contribute directly and indirectly to humans through producing honey, wax and propolis products, and as the pollinator of both wild and crop plants (Potts et al., 2010b). The ease with which honeybees can be managed, compared with other pollinators, makes them important to crop pollination (Potts et al., 2010b) (Figure 1).

Of the leading food crops worldwide, three quarters are dependent on animal pollination contributing to one third of global crop production (Klein et al., 2007). While staple crop plants such as wheat, rice and maize are wind pollinated, crops dependent on animal pollination are among the richest in micronutrients that are essential for human health (Chaplin-Kramer et al., 2014). The yield of most crop species is improved by animal pollination, and has also been shown to increase quality and improve shelf life (Klatt et al., 2014). Specifically, honeybees have been reported as increasing yield in 96% of animal-pollinated crops (Klein et al., 2007; Potts et al., 2010a).

However, there is concern over increased rates of honeybee colony loss worldwide, with a 25% loss of honeybee colonies in central Europe between 1985 and 2005 and a 59% loss of colonies in North America between 1947 and 2005 (Potts et al., 2010b; van Engelsdorp et al., 2008). In contrast, the overall population of honeybee colonies globally is estimated to have increased by 45% from 1961. Despite this worldwide increase in honeybees, within the same period there has been a larger increase in the agriculture that requires animal pollination, by over 300% (Aizen and Harder, 2009). In the mid to late 2000s, large-scale losses of honeybee colonies in the US over winter led to the term Colony Collapse Disorder (van Engelsdorp et al., 2008; VanEngelsdorp et al., 2009), and brought public attention to pollinator declines (Knight et al., 2018).

One of the major factors implicated in the decline of pollinators is the reduction in the abundance and diversity of floral resources due to agricultural intensification and the loss and fragmentation of habitat (Green, 1990; Petit et al., 2003; Robinson and Sutherland, 2002; Stevens et al., 2003). The interaction of this loss of suitable foraging habitat with

factors such as the increased application of pesticides, the spread of pests and diseases, apicultural mismanagement and climate change are all contributing to the loss and poor health of honeybee colonies (Goulson et al., 2015; Potts et al., 2010a). Having a detailed understanding of the plants most important to honeybees is vital to support hives and sustainable beekeeping. To do this, effective and efficient methods of monitoring honeybee foraging are required.

Figure 1. Inspecting the hives at the National Botanic Garden of Wales. The hive used here is a British Standard National hive.

1.1.1. Habitat loss and lack of suitable forage

Habitat loss is named as one of the most major and universal threats to biodiversity, including pollinating insects, with subsequent habitat fragmentation also impacting on surviving populations (Brown and Paxton, 2009; Dziba et al., 2016; Goulson et al., 2015, 2008). With agriculture representing the dominant land use across Europe, agri-environment schemes have been used as a way to mitigate the continuing loss of biodiversity that is associated with habitat loss and habitat quality reduction from agricultural intensification (Batáry et al., 2015). Since the early 20th century, farming has become more intensive, with increased use of machinery, inorganic fertilisers, and pesticides which has led to greater yields (Robinson and Sutherland, 2002). Associated impacts include a reduction in the presence of non-crop plant species, as well as an

estimated 50% reduction in hedgerows to allow access for machinery (Robinson and Sutherland, 2002). Areas which are intensively farmed with mass-flowering crops, such as oil seed rape (*Brassica napus*), provide a large, single forage at one time (Garbuzov et al., 2015; Goulson et al., 2015), and for agricultural landscapes, the effect of isolation from florally diverse natural and semi-natural habitats was found to decrease mean flower-visitor richness, visitation rate and fruit set in crop fields (Garibaldi et al., 2011). In honeybees it has been shown that for normal growth and development they require protein with a suitable volume and variety of amino acids (Brodschneider and Crailsheim, 2010; Standifer, 1967). As such, when examining honeybee foraging in intensive farmland habitats, Requier et al., (2015) highlighted the importance of non-crop species for improving flower availability and increasing the diversity of the honeybee diet in these landscapes.

In addition to agri-environment schemes to improve diversity of forage, gardens can provide important habitat and foraging resources in urban environments (Garbuzov and Ratnieks, 2014a; Salisbury et al., 2015). Gardeners often have an interest in planting for wildlife, with 64% of garden centre customers more likely to purchase plants with an endorsement for being pollinator-friendly (Wignall et al., 2019). To this end, schemes and lists recommending garden plants for pollinators have been created (Garbuzov and Ratnieks, 2014b). However, while research based on empirical evidence is used in the creation of some recommendations, many lists do not provide evidence for their recommended plants, and the overlap of plants between lists can be low (Garbuzov and Ratnieks, 2014b). Notably, Comba et al. (1999) found that horticultural modifications could affect the species composition and number of visiting insects, and that some cultivars with variations such as double headed flowers had reduced floral reward. In addition, floral doubling may suppress nectar secretion in certain species, thereby reducing their worth to insects (Corbet et al., 2001).

1.1.2. Pesticides

An associated impact of agricultural intensification is the increased use of pesticides with pesticide use described as the most controversial cause of bee declines (Goulson et al., 2015). A risk assessment of agrochemicals by Sanchez-Bayo and Goka, (2014) found that

the two exposure routes to honeybees are by direct exposure during and after spray application, and indirectly through exposure to residues in pollen, nectar, wax and honey. Furthermore, 161 pesticide residues were found in hives, and the risk to bees was estimated by considering toxicity, prevalence in the hive and residue loads. Sanchez-Bayo and Goka, (2014) considered five insecticides to pose the most risk, three neonicotinoids and two organophosphates.

Neonicotinoids were developed in the last three decades and are the newest of the major insecticide classes. They work by targeting the central nervous system of insects and are systemic throughout the plant (Goulson et al., 2015; Long and Krupke, 2016; Simon-Delso et al., 2015). Sub-lethal effects of neonicotinoid exposure have been observed (Desneux et al., 2007), resulting in increased levels of pathogens in honeybees (Pettis et al., 2012) and affecting foraging behaviour (Henry et al., 2012; Yang et al., 2008; Vandame et al., 1995). Williamson & Wright (2013) found that prolonged exposure to neonicotinoid and organophosphate pesticides affected olfactory learning and memory formation in honeybees; behaviours important to foraging. In addition, negative synergistic effects of sub-lethal pesticide doses coupled with pathogens have been observed, resulting in increased mortality (Doublet et al., 2015; Vidau et al., 2011). When examining foraging choices between sucrose solutions with or without neonicotinoids, it was shown that honeybees did not avoid the laced sucrose solution and also preferred it to the sucrose only solution (Kessler et al., 2015).

Worldwide, a survey of 198 honey samples detected one of five neonicotinoids in 75% of samples (Mitchell et al., 2017). Honey from beekeepers across the UK was found to contain neonicotinoids even after an EU moratorium of the application of three classes of neonicotinoid seed treatments in 2013 (Woodcock et al., 2018). The majority of honey with neonicotinoids present post-moratorium were sampled earlier in the year, and correlated with the area of oilseed rape in the area surrounding the hive (Woodcock et al., 2018).

1.1.3. Pests and parasites

One of the most serious parasites of *A. mellifera* is the *Varroa destructor* mite (Rosenkranz et al., 2010). Originally a parasite of the Eastern honeybee *Apis cerana, V.*

destructor shifted from its original host to *A. mellifera* where, as a new parasite of the Western honeybee, it shows none of the stable host-parasite relationship it displays in *A. cerana.* The *Varroa* mite feeds on the haemolymph of honeybees and additionally is a vector for pathogens such as Deformed Wing Virus, implicated in Colony Collapse Disorder (VanEngelsdorp et al., 2009). Regular monitoring and treatment with acaricides is required in order to prevent the loss of honeybee colonies (Rosenkranz et al., 2010). More recently the Asian hornet (*Vespa velutina*) has been a concern of beekeepers in the UK after its introduction to France in 2004, with its first appearance in the UK being confirmed in September 2016 (Keeling et al., 2017). The Asian hornet will prey on a variety of pollinating insects, including in its native range the Eastern honeybee, catching individuals as they return to the hive (Arca et al., 2014).

Nosema apis and *Nosema ceranae* are obligate intracellular fungal parasites that affect *A. mellifera*, attacking the epithelial lining of the midgut and affecting nutrient adsorption. *N. ceranae* is thought to have jumped to *A. mellifera* from *A. ceranae* in the 2000s, with its rapid dispersal attributed to the wide scale transportation of honeybees for commercial and hobby purposes (Klee et al., 2007). Other potential concerns include *Aethina tumida*, the small hive beetle, which has been introduced from sub-Saharan Africa to North America and Australia causing significant economic impact to apiculture (Cuthbertson et al., 2013; Schäfer et al., 2019). In 2014, *A. tumida* was found in Italy but is yet to be recorded in the UK (Palmeri et al., 2015).

1.1.4. Genetic variation

In honeybees, genetic variability has been linked to improved resistance to disease (Seeley and Tarpy, 2007; Tarpy, 2003), more stable thermoregulation (Graham et al., 2006; Jones, 2004) and overall improved colony fitness from better foraging rates, food storage and population growth (Mattila and Seeley, 2007). However, the introduction of non-local honeybees is a concern for conservation of local variation. Beekeepers can create admixtures of divergent honeybee subspecies by the introduction of commercial foreign queens (Meixner et al., 2010). As such, arguments for protecting native genetic diversity have been made to prevent the introduction of potentially maladapted ecotypes and the loss of genetic diversity that may negatively affect honeybees' ability to respond

to changing environmental conditions (Pinto et al., 2014). The conservation of the genetic diversity and ecotypes that underlay the adaptability of *A. mellifera* is therefore an important component in responding to climate change (Le Conte and Navajas, 2008).

1.1.5. Subspecies of *Apis mellifera*

The anthropogenic movement of the Western honeybee has greatly changed the species' natural distribution, affecting subspecies and resulting in hybridisation (Ruttner et al., 1978). The honeybee has been grouped into four evolutionary groups: the West Mediterranean, North Mediterranean, Oriental and African (Ruttner et al. 1978; Estoup et al. 1995; Meixner et al. 2013) lineages of *A. mellifera*. Of approximately 30 recognised subspecies of *A. mellifera,* ten are native to Europe (Pinto et al., 2014) and these ten subspecies are represented by two lineages, the M-lineage found in Western Europe and the C-lineage in Eastern Europe (Muñoz et al., 2015; Wallberg et al., 2014). The distinction between the geographic subspecies and lineages was initially based on morphometric data associated with behavioural and ecological characteristics, subsequently also incorporating microsatellite genetic data (Estoup et al. 1995) and genome sequencing (Wallberg et al., 2014).

In the UK, the M-lineage honeybee *Apis mellifera mellifera* (dark honeybee) is under threat, mostly due to introgression from C-lineage honeybees including *Apis mellifera carnica* (Carniolan honeybee) and *Apis mellifera ligustica* (Italian honeybee) (Muñoz et al., 2015). Pinto et al. (2014) compared protected and unprotected populations of the subspecies *A. m. mellifera*, using mitochondrial DNA and a genome-wide scan of single nucleotide polymorphisms (SNPs). While introgression was found to be higher in unprotected populations versus protected, foreign haplotypes did still occur in the protected populations. For honeybees in Ireland, mitochondrial and microsatellite analysis revealed that there were pure populations of *A. m. mellifera* present, with bees that showed resemblance in their genetics to other European bees, but also the presence of distinct Irish haplotypes that had not been sequenced before (Hassett et al., 2018).

1.1.6. Nutritional requirements of honeybees

The nutritional requirements of a honeybee's diet differ according to ontogenetic stage and are met by the nectar and pollen of the plants they forage on, which forms the basis

for the colony's growth and development (Winston, 1987). In honeybee colonies, foraging is completed by sterile female workers, rather than the queen or drones. Throughout the honeybee life-cycle, adult workers require more carbohydrates for high energy activities such as flight or wax production, while larvae and young workers need more protein for growth (Haydak, 1970). Poor quality feeding of the larvae may result in reduced numbers or quality in the next generation of adults, possibly affecting the next round of brood rearing (Brodschneider and Crailsheim, 2010). For example, Scofield & Mattila (2015) found that larvae with restricted access to pollen were lighter and shorter-lived than larvae with normal access. Moreover, as adults, the pollen-restricted individuals were less likely to forage. If they did forage, they were more likely to die after only one day. At the scale of the hive, a decline in the spring pollen harvest was associated with negative health impacts later in the season, including reduced brood production, smaller honey stores in winter and increased colony loss over winter (Requier et al., 2017).

In Alaux et al. (2010) the diversity of pollen intake was linked to increased immuno-competence in honeybees, compared to a monofloral increase in protein quantity. However, Di Pasquale et al. (2013) showed that a polyfloral diet was not necessarily better than monofloral pollen of good quality, noting that not all physiological factors will be equally affected by the pollen diet. In contrast, honeybees fed a pollen substitute were found to have higher levels of the *Nosema* parasite compared to honeybees fed on wildflower pollen, indicating the potential importance of a polyfloral pollen diet to wider bee health (Fleming et al., 2015).

1.1.7. Pollen nutrition and foraging preferences

Pollen is the main source of protein for honeybees, also supplying lipids, sterols, minerals and nutrients (Vaudo et al., 2015) with inadequate pollen consumption early in the life of the bee resulting in increased mortality (Scofield and Mattila, 2015). Notably, pollen from different plants can vary greatly in amino acid composition (Haydak, 1970), with protein concentration varying from around 2 to 60% depending on the plant species (Avni et al., 2014; Pernal and Currie, 2001; Vaudo et al., 2015). De Groot (1953) found 10 amino acids that were highly important for honeybee growth. Arginine, histidine, lysine, tryptophan, phenylalanine, methionine, threonine, leucine, isoleucine and valine were identified as

essential amino acids, with leucine, isoleucine and valine required in the greatest amounts. Pollen also provides the lipid source for honeybees, which is mainly metabolised for energy, while sterols are essential to brood rearing (Brodschneider and Crailsheim, 2010). The lipid concentration in pollen is estimated to range from 1 to 20% (Roulston and Cane, 2000), and a deficiency in omega-3 polyunsaturated fatty acids in the diet has been shown to have detrimental effects on the cognitive ability of the honeybee (Arien et al., 2015).

When adult workers are emerging they are still developing their hypopharyngeal glands and fat bodies, which is dependent on the pollen protein available (Haydak, 1970; Winston, 1987). These hypopharyngeal glands are used to produce jelly that is, in main, fed to the brood, and so the glands are most developed when the worker is nursing (Crailsheim et al., 1992). The development of the hypopharyngeal glands and ovaries in emerging workers was found to be correlated with the amount of protein consumed from pollen, but the consumption preferences of the workers was unrelated to the level of protein contained in the pollen (Pernal & Currie 2000). Furthermore, Pernal & Currie (2001) also showed that honeybees respond to pollen deficiencies by increasing the amount of pollen foraged, rather than selecting for pollen with higher protein content. On average, around 1 kg of pollen stores is maintained by the honeybees, with an estimated requirement for around between 13 and 18 kg of pollen per year (Brodschneider and Crailsheim, 2010; Wright et al., 2018).

1.1.8. Nectar nutrition and foraging preferences

Nectar provides the carbohydrate source required by honeybees, although honeydew from hemipterous insects (aphids, leaf hoppers) is also utilised to some degree (Doner, 1977). The main sugars in nectar are sucrose (disaccharide), glucose (monosaccharide) and fructose (monosaccharide), with other mono- and disaccharides possibly present in smaller amounts. The total concentration of sucrose, glucose and fructose can range from 7 to 70% (Nicolson et al., 2007). Percival (1961) found that across 889 species of angiosperms, nectars could be broadly classified into three categories: dominant with sucrose; balanced between sucrose, fructose, and glucose; or dominant fructose and

glucose. Other sugars can be toxic to bees, including mannose, galactose and arabinose (Barker and Lehner, 1974).

Nectar and honeydew can be used or fed directly by the honeybees but is most commonly processed into honey (Winston, 1987). The process of converting nectar and honeydew into honey involves reducing the water content down to 16-20% and the addition of enzymes (diastase, invertase and glucose oxidase) which break down the sugars in nectar or honeydew to monosaccharide forms (Brodschneider and Crailsheim, 2010; Doner, 1977). In Mao et al. (2013), constituents found in honey and derived from pollen were found to specifically induce detoxification genes in honeybees, highlighting the potential risks of substituting sugar feeds for honey.

Honeybee workers do not store large amounts of energy reserves in their body and do not survive long periods without feeding. As such, they are reliant on the colony food stores and storage in the honey stomach during high level activity, such as foraging (Hrassnigg et al., 2005). The larvae also require carbohydrates for development, but larval needs can be met by pollen and brood food (Vaudo et al., 2015).

Nectar can also contain varying low concentrations of amino acids and other nutrients although their role and importance is unknown (Baker and Baker, 1973; Crailsheim, 1990; Goulson et al., 2015; Wright et al., 2013). Secondary plant compounds such as alkaloids in nectar may deter some nectar feeders, and are generally thought to have a negative effect on visitation (Adler, 2000; Nicolson et al., 2007). However, the compounds nicotine, anabasine, caffeine and amygdalin were examined for feeding preference and only anabasine was found to deter honeybees, while low concentrations of nicotine and caffeine caused a feeding preference (Singaravelan et al., 2005). Caffeine, a plant-produced alkaloid present in the nectar of *Coffea*, *Citrus* and *Camellia* species, has been shown to enhance honeybees' memory of a sucrose reward (Wright et al., 2013) and increase colony level recruitment to foraging sources with caffeine (Couvillon et al., 2015).

There are three main characteristics of nectar that have been found to influence bee foraging: sugar composition, nectar concentration, and nectar volume (Vaudo et al., 2015; Waller, 1972; Wykes, 1952). Nectar concentration can limit which pollinators are able to

mechanically obtain the nectar, with long-tongued bees being potentially limited to more dilute nectars. Honeybees have been found to collect sucrose and fructose in preference to glucose, with a preference for sucrose in a 30-50% concentration (Waller, 1972), while in the field they will collect nectar with a range of concentrations depending on availability of profitable resources (Seeley, 1986).

Foraging honeybees have been shown to exhibit floral constancy, where in one foraging trip they tend to visit only one species or type of flower, while at a colony level forage will be collected from multiple sources (Wright et al., 2018). Wells & Wells (1983) found that honeybees also demonstrated individual constancy to flower colours and morphs even if the sucrose reward quality and frequency was lower. However, when Grüter et al. (2011) tested honeybee flower constancy with ecologically realistic rewards, flower constancy was more likely for the most rewarding situation, in terms of sucrose concentration, volume and frequency.

1.1.9. Honeybee foraging and how it is monitored

Honeybee foraging has been reported to be influenced by the surrounding local vegetation (Requier et al., 2015; Webby, 2004), seasonal changes (Coffey and Breen, 1997; Couvillon et al., 2014), interactions with other species (Balfour et al., 2013), and the requirements of the colony (Schmickl and Crailsheim, 2004). Nectar foraging was not found to be influenced by changes in honey storage levels in the colony (Fewell and Winston, 1996); unlike pollen foraging (Fewell and Winston, 1992).

When foraging resources are highly variable in patch size and quality, distances travelled by honeybees may be larger, with median distances of 6.1 km recorded by Beekman & Ratnieks (2000) with some trips of over 10 km. While in a less patchy landscape, Visscher & Seeley (1982) found the median foraging distance in a deciduous forest was 1.7 km with 95% of foraging occurring within 6.0 km. The mean foraging distance and area were also found to vary according to the annual cycle, with increases witnessed from spring to summer, accompanied by decreases in the autumn likely relating to the availability of forage (Couvillon et al., 2014). Similar patterns were seen by Steffan-Dewenter and Kuhn, (2003) where foraging was highest in June, when the available resources were scarce.

Foraging honeybees communicate the locations of resources to the colony by using waggle dances, a complex communication method (Seeley, 1995). Contained in the waggle dance is the distance, direction and quality of the resource from the hive. Foragers returning from high quality forage sources are more likely to perform dances than those from poorer forage (Seeley and Towne, 1992) with resources being nectar, pollen, water, tree resin and new nest sites (Schürch et al., 2013). By observing whether the returning forager is carrying pollen, a distinction can be drawn between whether the worker is communicating a source of pollen or nectar (Couvillon and Ratnieks, 2015) and studies have used the waggle dance to track honeybee foraging (Schürch et al., 2013; Visscher and Seeley, 1982). However, as not all foraging sites are reported back at the hive, monitoring dances will only give information about the most profitable resources (Seeley, 1995).

Riley et al. (2005) used harmonic radar to compare the flight paths of foragers with their waggle dance and found that the waggle dance is highly effective, but still requires odour and visual cues at the end of the flight. Harmonic radar has also been used to track orientating trips of honeybees, which are shorter than foraging trips as foraging honeybees were found to leave the 700 m range of the harmonic radar tracking (Capaldi et al. 2000). Winston (1987) states that when the foraging worker reports distance to a forage source, the waggle dance is communicating the relative expenditure of energy. Putting small weights on foragers, therefore, can cause them to communicate greater distances to a forage in the waggle dance (Schifferer 1952, in Winston, 1987).

Radio-tracking has been used to track the flight distances and space use by bumblebees, however differences in flower-handling and greater rest periods between bumblebees with and without transmitters was observed, suggesting an energetic cost due to the weight of the tracker (Hagen et al. 2011). This technology would likely be too heavy for honeybees. In mark-recapture methods, bees are marked at the hive or nest and captured while out foraging, which allows the foraging trips to plants to be directly observed (Greenleaf et al., 2007). The bees can be marked in different ways including radio frequency identification (RFID) tags as in Pahl et al. (2011), where artificially dispersed honeybees were tracked returning to the hive. Southwick & Buchmann (1995) used a metal tagging system with magnetic traps at the hive entrances which also assessed

homing. Quantum dots, semiconductor nanocrystals than emit bright light when exposed to UV, have been tested in use of tracking the movement of pollen and could be potentially be utilised in tracking insect movement (Minnaar and Anderson, 2019).

To investigate floral visitation, studies have used the morphological identification of pollen both by using pollen loads collected from honeybees returning to the hive (Synge, 1947), and by examining the pollen present in the honey (Aronne et al., 2012). The extraction and identification of pollen from within honey is termed melissopalynology and requires a high-level of skill and experience, with reproducibility dependent on the experience of the person (Bruni et al., 2015; Louveaux et al., 1978). The level of taxonomic classification in palynological analysis can range from species level to family (Aronne et al., 2012), and is used to characterise the botanical and geographical origin of honey (Louveaux et al., 1978). Establishing botanical origin is vital to legislation regulating the labelling and selling of honey, and for the quality control of an economically important food product (Bruni et al., 2015). However, microscopical analysis of the pollen may underestimate or overestimate certain plant species relative to their actual nectar contribution to a honey; in that the percentage of pollen in the sediment may not correspond to the percentage of nectar in the honey (Louveaux et al., 1978).

In addition to use in the commercial aspects of honey, melissopalynology has also been used to ecologically investigate honeybee foraging. Sampling pollen throughout April to September, Requier et al. (2015) identified 228 different plant species from pollen samples in an agricultural landscape in France. The majority of pollen was collected from species growing as weeds in crops and semi-natural habitats, contributing an important portion of the honeybees' diet between two mass flowering agricultural events. Looking at both honey and pollen stored by the honeybees in Italy, Aronne et al. (2012) found that 69% of the pollen types were represented in both honey and pollen, with 14% of types found only in honey and 17% types only in pollen.

1.1.10. DNA identification of pollen

As an alternative to melissopalynology, the use of DNA identification methods to characterise both collected pollen and the pollen in honey has been used, with the possibility of increasing species differentiation and reducing the level of specific expertise

needed (Bruni et al., 2015; Danner et al., 2017; de Vere et al., 2017; Galimberti et al., 2014; Hawkins et al., 2015; Jain et al., 2013). Both the DNA from the pollen loads collected by honeybees and from the pollen in the honey has been successfully extracted using different methods (de Vere et al., 2017; Guertler et al., 2014; Hawkins et al., 2015; Soares et al., 2015; Waiblinger et al., 2012).

Pollen foraging was investigated using PCR and Sanger sequencing by Wilson et al. (2010) which is suitable for pollen loads consisting of one plant species but ineffective for multifloral mixed samples as with honey. Real-time PCR assays for fifteen plant species were used to develop a method for identifying geographical origin of honey, but such an approach is based on already knowing the species most likely to be found in a specific area, and was tested on regions that are geographically remote to each other (Laube et al., 2010).

1.1.11. DNA metabarcoding

DNA metabarcoding, using DNA barcode regions in combination with high throughput sequencing, allows the analysis of multiple mixed species samples, such as from honey, fecal matter, or gut contents (Deiner et al., 2017; Pompanon et al., 2012). Studies using this approach to investigate pollen DNA metabarcoding have utilised different sequencing platforms: Ion Torrent to examine quantifying airborne pollen (Kraaijeveld et al., 2015); Illumina MiSeq to investigate the spring foraging of honeybees in a diverse system (de Vere et al., 2017); 454 pyrosequencing to look at its ability to characterise commercial honey (Valentini et al., 2010) and to compare DNA metabarcoding of honey with melissopalynology and how it relates to honeybee foraging (Hawkins et al. 2015).

DNA metabarcoding relies upon several components to be effective. A DNA region with universal primers needs to be used in order to amplify a true representation of the variety of species in the sample (Hollingsworth et al., 2011) and the level of taxonomic discrimination that is achievable with the DNA region should be suitable for the requirements of the study. Vitally, the reference library that the unknown DNA regions are compared to needs to be rigorous and comprehensive to ensure accurate identification (Coissac et al., 2012). Without an understanding of the coverage and quality

of the available reference library sequences, the quality of any achieved identification of unknown sequences is more difficult to assess.

The DNA regions used by previous studies include plastid markers *rbcL, trnH-psbA* and *trnL* (Bruni et al., 2015; de Vere et al., 2017; Galimberti et al., 2014; Kraaijeveld et al., 2015; Valentini et al., 2010), and nuclear markers, ITS (Wilson et al., 2010) and ITS2 (Keller et al., 2015; Richardson et al., 2015). Plant barcode markers are often used in combination to gain taxonomic range and species discrimination that is not necessarily achievable with one marker. The *rbcL* marker has high universality and is one of the internationally agreed DNA barcode markers for plants (CBOL Plant Working Group, 2009) while *trnH-psbA* and ITS2 have been recognised as additional markers that can help increase species discrimination (Hollingsworth et al., 2011).

1.1.12. Knowledge gaps and project overview

DNA metabarcoding of pollen and honey allows a new, potentially more efficient, avenue for the identification of honeybee forage plants. To make high-quality, validated identifications from DNA metabarcoding analyses, a vital component is a comprehensive DNA barcode reference library. The Welsh native flora has been DNA barcoded with *rbcL* and *matK* (de Vere et al., 2012). DNA barcodes are required for both the UK native species not present in the Welsh database, in addition to non-native species. To provide wider support for applications, the UK DNA barcode database will also include the ITS2 plant DNA barcode. This resource is vital for identifying the floral source of pollen in honey, and will provide a strong foundation for ongoing and future plant DNA metabarcoding research in the UK.

Detailed analyses of which plants honeybees choose to use throughout the season is lacking within the scientific literature. Honeybees are often considered to be super generalists within the context of pollinator-plant networks, however it has previously been shown that during spring their plant choice can be limited even within a horticulturally diverse landscape (de Vere et al., 2017). Our current understanding of honeybee foraging throughout the season within the UK is based on historical surveys using melissopalynology of honey and analysis of pollen loads (Deans, 1957; Percival, 1947; Synge, 1947), however since the 1940s there have been major changes in

agricultural practices with resulting shifts in the potential availability of forage sources (Baude et al., 2016; Robinson and Sutherland, 2002). To support honeybee colonies from the effects of the interacting stressors of inadequate forage, exposure to pesticides, pests and diseases, and climate change, a fuller understanding of the plants which are most important to honeybees currently is required.

To achieve this, a UK wide survey of honey plants is needed, collecting honey and metadata from beekeepers across the UK. This will provide the first UK wide survey of honey since 1952, giving the picture of current honeybee foraging patterns and the plants most important to honey production. For the full temporal range of honeybee foraging within the UK, this survey is combined with a detailed study of honey collected through the honeybees' active foraging season and compared to a well botanised landscape.

By combining plant forage DNA metabarcoding data with floral plant surveys, landscape habitat information and historical foraging data, the foraging behaviour and preferences of honeybees can be examined, providing an unparalleled account of the current foraging decisions of honeybees within a diverse landscape as well as across the UK.

1.2. Aims and outline of thesis

1.2.1. Aims and objectives

- Create a DNA barcode reference library for the UK native flowering plant and conifer species, using three DNA barcode loci, *rbcL*, *matK*, and ITS2.
    - How do *rbcL*, *matK*, and ITS2 vary in their ability to successfully recover a sequence, within the UK flora?
    - How do *rbcL*, *matK*, and ITS2 vary in their species level discrimination?

- Create a reference library including UK non-native plant species from all available DNA sequences on GenBank to support the applications of DNA barcoding.
    - How representative are the currently available sequences for the UK non-native and native plant species?
    - What is the species level discrimination ability within the reference library for the selected native and non-native UK plant species?

- Investigate the foraging preferences of honeybees using DNA metabarcoding within a diverse floral landscape.
    - Which plants are honeybees choosing throughout the season (April to September) when offered a diverse floral resource?
    - Are there differences in foraging between hives placed near horticultural plants, compared with hives situated further away but within foraging distance?
    - How does the abundance of a plant within the landscape relate to the abundance found within the honey?
    - Is there any pattern to the characteristics of the forage plants found throughout the season, in terms of native status, growth form and associated habitat?

- Characterise the plants honeybees are using across the UK using DNA metabarcoding
    - Are there spatial and temporal patterns to honeybee foraging within the UK?
    - How does the presence of crop species within foraging distance of the hive relate to the floral composition of the honey?
    - Have landscape scale changes in forage availability since the early 20th century affected the use of key honeybee forage plants in the UK?

1.2.2. Thesis outline

1.2.2.1. Chapter 1: Introduction

An overview of the main contributors to pollinator decline are presented, along with the nutritional requirements of honeybees. Traditional ways of tracking honeybee foraging are discussed and DNA barcoding and metabarcoding are introduced as methods to improve monitoring.

1.2.2.2. Chapter 2: Creating a DNA barcoding reference library for the flowering plants and conifers of the UK

To effectively analyse and identify unknown sequences from plant DNA metabarcoding data within the UK, a DNA barcode reference library for UK plant species is required. The 1,482 native flowering plants and conifers present in the UK are DNA barcoded using three barcode loci, *rbcL*, *matK* and ITS2. To facilitate studies with a wider scope than the native plants, including pollinator foraging, the reference library is supplemented using sequence data from GenBank for 5,586 plant species covering native, naturalised and horticultural species, found in the UK. Both databases are analysed to assess their ability in sequence recovery and database representation, as well as the relative ability of the markers to discriminate to different taxonomic levels.

1.2.2.3. Chapter 3: Temporal and spatial patterns of honeybee foraging in a diverse floral landscape

Honeybee foraging at a landscape scale is characterised within a diverse landscape, which contains a Botanic Garden and a National Nature Reserve, managed as an organic farm. Each month, from May to September 2016 and April to September 2017, the available flowering plants were recorded. At the same time, honey samples were taken from hives set in two locations within the study site, one within the Botanic Garden and one within the Nature Reserve. Using DNA metabarcoding, and the DNA barcode reference library created in Chapter 2, the floral source of the honey from six hives was characterised each month over two years. The plants found in the honey are investigated, in terms of their spatial relationship with the surrounding landscape, the phenology of plants through the

season and the characteristics of the plants in terms of native status, growth form and associated habitat.

1.2.2.4. Chapter 4: Agricultural intensification, shifts in crop use and invasive species induce nationwide change in honeybee foraging

Honeybee foraging on a national scale is examined, using 441 honey samples provided by beekeepers from across the UK. Each sample was characterised using DNA metabarcoding with the *rbcL* and ITS2 plant barcode regions. This survey represents the first widespread analysis of UK honey since 1952. Using land cover habitat data and crop data we examine spatial relationships with insect-pollinated crops found in the honey. To examine changes in the major sources of honey between 1952 and 2017, we compare the results of the DNA metabarcoding with the results of a survey conducted in 1952 by A.S.C Deans (Deans, 1958, 1957).

1.2.2.5. Chapter 5: Discussion

This chapter presents the overall discussion for the thesis, covering the future directions of DNA barcoding and metabarcoding, and further applications of this research. Recommendations for planting for pollinators to provide for the requirements of honeybees are discussed.

1.2.2.6. Appendix A: BeeCraft Article: Honeybee foraging, new techniques to barcode the natural world

An article published in the UK beekeeping magazine BeeCraft outlining the spring honeybee forage results of the paper presented in Appendix B. The article was also used to recruit beekeepers for the UK wide honey survey in Chapter 4.

1.2.2.7. Appendix B: Using DNA metabarcoding to investigate honeybee foraging reveals limited flower use despite high floral availability

de Vere, N., **Jones, L.**, Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S., Ford, C.R., 2017. Using DNA metabarcoding to investigate honeybee foraging reveals

limited flower use despite high floral availability. *Scientific Reports* **7**. DOI: 10.1038/srep42838

Honeybee foraging early in the season (April and May) is investigated within a botanic garden, showing that honeybees use a limited proportion of the plant species available to them. This work formed the basis of the project proposal for my PhD and I helped to prepare this paper during my PhD.

1.2.2.8. Appendix C: Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding

Lucas, A., Bodger, O., Brosi, B.J., Ford, C.R., Forman, D.W., Greig, C., Hegarty, M., **Jones, L.,** Neyland, P.J., de Vere, N., 2018. Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding. *Scientific Reports* **8.** DOI: 10.1038/s41598-018-23103-0

The pollen from eleven species of hoverfly, 143 individuals, was examined using DNA metabarcoding with *rbcL*, revealing the pollen transport network structures in grassland communities. Hoverflies were found to carry pollen from 59 plant taxa, predominantly Apiaceae, Cardueae, *Calluna vulgaris*, *Rubus fruticosus* agg. and *Succisa pratensis*.

1.2.2.9. Appendix D: Pollen metabarcoding reveals broad and species-specific resource use by urban bees

Potter, C., de Vere, N., **Jones, L.,** Ford, C.R., Hegarty, M.J., Hodder, K.H., Diaz, A., Franklin, E.L., 2019. Pollen metabarcoding reveals broad and species-specific resource use by urban bees. *PeerJ* **7**. DOI: 10.7717/peerj.5999

The foraging patterns of bee species found using sown wildflower strips in an urban environment were examined using DNA metabarcoding with the *rbcL* DNA barcode region. Bees were found to forage on a wide range of plant taxa, from within and outside the wildflower plots, showing that bees utilise wildflower mixes as part of the wider urban environment.

1.2.2.10. Appendix E: Temperate airborne grass pollen defined by spatio-temporal shifts in community composition

Brennan, G.L., Potter, C., de Vere, N., Griffith, G.W., Skjøth, C.A., Osborne, N.J., Wheeler, B.W., McInnes, R.N., Clewlow, Y., Barber, A., Hanlon, H.M., Hegarty, M., **Jones, L.**, Kurganskiy, A., Rowney, F.M., Armitage, C., Adams-Groom, B., Ford, C.R., Petch, G.M., Creer, S., 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution* **3**. DOI: 10.1038/s41559-019-0849-7

Airborne grass pollen through the season in the UK was characterised using DNA metabarcoding with *rbcL* and ITS2. Grass pollen represents an important contributor to allergenic reactions from airborne pollen. Both the time of sampling and location of sampling were found to drive changes in the community composition of the grass pollen.

1.2.2.11. Appendix F: Conference presentations, art exhibitions and media

Outlines my conference presentations throughout the course of my PhD, as well as art exhibitions inspired by the research presented here. A full list of media appearances is provided, these were used to publicise my results and recruit beekeepers for the UK survey in Chapter 4.

The contribution of authors to the chapters and papers within the appendices are outlined in Table 1.

Table 1. Outlining the contribution of authors to the chapters and my contribution to the papers within the appendices.

	Contributions and acknowledgements
Chapter 2	• Field sampling was completed by Natasha de Vere, Tim Rich, and Heather McHaffie. • I completed the herbarium sampling for the UK species with Helena Davies, Natasha de Vere, Adelaide Griffith, Joseph Moughan, and Tim Rich. • Labwork and DNA sequence processing was completed by myself, Helena Davies, Joe Moughan, and Eleanor Brittain. Sequencing was provided by Laura Forrest at Royal Botanic Garden, Edinburgh. • I completed all of the data and statistical analyses.
Chapter 3	• I completed all of the fieldwork surveying plant species throughout the season with the help of a team of conservation volunteers and students. • The honeybee hives are managed by beekeeper Lynda Christie and a team of volunteers. • Col Ford and I wrote and developed the bioinformatic analysis pipeline. • I completed all of the labwork and data analysis. • Georgina Brennan advised on statistical analysis and model selection.
Chapter 4	• Col Ford and I wrote and developed the bioinformatic analysis pipeline. • I completed all of the labwork and data analysis. • Georgina Brennan advised on the statistical analysis and model selection.
Appendix A	• I wrote this article for BeeCraft after being an invited speaker for the research lectures at the National Honey Show in October 2016.
Appendix B	• I contributed to the development of the honey DNA metabarcoding methods and helped complete the labwork and analyse the data. I created the figures for this paper.
Appendix C	• I contributed to the development of the pollen DNA metabarcoding methods and completed the sequence library preparation labwork for this paper.
Appendix D	• I contributed to the labwork and supervised the first author in pollen DNA metabarcoding methods.
Appendix E	• I contributed to the labwork, methods and bioinformatic analyses with this paper.
Appendix F	• I provided scientific support and expertise to the art installations and exhibitions listed here, which interpreted my research.

1.3. References

Adler, L.S., 2000. The ecological significance of toxic nectar. *Oikos* **91**, 409–420.

Aizen, M.A., Harder, L.D., 2009. The Global Stock of Domesticated Honey Bees Is Growing Slower Than Agricultural Demand for Pollination. *Current Biology* **19**, 915–918.

Alaux, C., Ducloz, F., Crauser, D., Le Conte, Y., 2010. Diet effects on honeybee immunocompetence. *Biology Letters* **6**, 562–565.

Arca, M., Papachristoforou, A., Mougel, F., Rortais, A., Monceau, K., Bonnard, O., Tardy, P., Thiéry, D., Silvain, J.-F., Arnold, G., 2014. Defensive behaviour of *Apis mellifera* against *Vespa velutina* in France: Testing whether European honeybees can develop an effective collective defence against a new predator. *Behavioural Processes* **106**, 122–129.

Arien, Y., Dag, A., Zarchin, S., Masci, T., Shafir, S., 2015. Omega-3 deficiency impairs honey bee learning. *Proceedings of the National Academy of Sciences* **112**, 201517375.

Aronne, G., Giovanetti, M., Guarracino, M.R., de Micco, V., 2012. Foraging rules of flower selection applied by colonies of Apis mellifera: ranking and associations of floral sources. *Functional Ecology* **26**, 1186–1196.

Avni, D., Hendriksma, H.P., Dag, A., Uni, Z., Shafir, S., 2014. Nutritional aspects of honey bee-collected pollen and constraints on colony development in the eastern Mediterranean. *Journal of Insect Physiology* **69**, 65–73.

Baker, H.G., Baker, I., 1973. Amino-acids in Nectar and their Evolutionary Significance. *Nature* **241**, 543–545.

Balfour, N.J., Garbuzov, M., Ratnieks, F.L.W., 2013. Longer tongues and swifter handling: Why do more bumble bees (*Bombus* spp.) than honey bees (*Apis mellifera*) forage on lavender (*Lavandula* spp.)? *Ecological Entomology* **38**, 323–329.

Barker, R.J., Lehner, Y., 1974. Acceptance and sustenance value of naturally occurring sugars fed to newly emerged adult workers of honey bees (*Apis mellifera* L.). *Journal of Experimental Zoology* **187**, 277–285.

Batáry, P., Dicks, L. V., Kleijn, D., Sutherland, W.J., 2015. The role of agri-environment schemes in conservation and environmental management. *Conservation Biology* **29**, 1006–1016.

Baude, M., Kunin, W.E., Boatman, N.D., Conyers, S., Davies, N., Gillespie, M.A.K., Morton, R.D., Smart, S.M., Memmott, J., 2016. Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature* **530**, 85–88.

Beekman, M., Ratnieks, F.L.W., 2000. Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology* **14**, 490–496.

Brodschneider, R., Crailsheim, K., 2010. Nutrition and health in honey bees. *Apidologie* **41**, 278–294.

Brown, M.J.F.F., Paxton, R.J., 2009. The conservation of bees: a global perspective. *Apidologie* **40**, 410–416.

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., Labra, M., 2015. A DNA barcoding approach to identify plant species in multiflower honey. *Food Chemistry* **170**, 308–315.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12794–7.

Chaplin-Kramer, R., Dombeck, E., Gerber, J., Knuth, K. a, Mueller, N.D., Mueller, M., Ziv, G., Klein, A.-M., 2014. Global malnutrition overlaps with pollinator-dependent micronutrient production. *Proceedings of the Royal Society B: Biological Sciences* **281**.

Coffey, M.F., Breen, J., 1997. Seasonal variation in pollen and nectar sources of honey bees in Ireland. *Journal of Apicultural Research* **36**, 63–76.

Coissac, E., Riaz, T., Puillandre, N., 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology* **21**, 1834–1847.

Comba, L., Corbet, S.A., Barron, A., Bird, A., Collinge, S., Miyazaki, N., Powell, M., 1999. Garden Flowers: Insect Visits and the Floral Reward of Horticulturally-modifed Variants. *Annals of Botany* **83**, 73–86.

Corbet, S.A., Bee, J., Dasmahapatra, K., Gale, S., Gorringe, E., La Ferla, B., Moorhouse, T., Trevail, A., Van Bergen, Y., Vorontsova, M., 2001. Native or exotic? Double or single? Evaluating plants for pollinator-friendly gardens. *Annals of Botany* **87**, 219–232.

Couvillon, M.J., Al Toufailia, H., Butterfield, T.M., Schrell, F., Ratnieks, F.L.W., Schürch, R., 2015. Caffeinated Forage Tricks Honeybees into Increasing Foraging and Recruitment Behaviors. *Current Biology* 2815–2818.

Couvillon, M.J., Ratnieks, F.L.W., 2015. Environmental consultancy: dancing bee bioindicators to evaluate landscape "health." *Frontiers in Ecology and Evolution* **3**.

Couvillon, M.J., Schurch, R., Ratnieks, F.L.W., 2014. Waggle dance distances as integrative indicators of seasonal foraging challenges. *PloS ONE* **9**.

Crailsheim, K., 1990. The protein balance of the honey bee worker. *Apidologie* **21**, 417–429.

Crailsheim, K., Schneider, L.H.W., Hrassnigg, N., Bühlmann, G., Brosch, U., Gmeinbauer, R., Schöffmann, B., 1992. Pollen consumption and utilization in worker honeybees (Apis mellifera carnica): Dependence on individual age and function. *Journal of Insect Physiology* **38**, 409–419.

Crane, E., 1999. Recent research on the world history of beekeeping. *Bee World* **80**, 174–186.

Cuthbertson, A.G.S., Wakefield, M.E., Powell, M.E., Marris, G., Anderson, H., Budge, G.E., Mathers, J.J., Blackburn, L.F., Brown, M.A., 2013. The small hive beetle Aethina tumida: A review of its biology and control measures. *Current Zoology* **59**, 644–653.

Danner, N., Keller, A., Härtel, S., Steffan-Dewenter, I., 2017. Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected pollen. *PLoS ONE* **12**, 9–12.

De Groot, A.P., 1953. Protein and Amino Acid Requirements of the Honeybee (Apis mellifica L.). Utrecht University.

de Vere, N., Jones, L.E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S., Ford, C.R., 2017. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports* **7**, 42838.

de Vere, N., Rich, T.C.G., Ford, C.R., Trinder, S.A., Long, C., Moore, C.W., Satterthwaite, D., Davies, H., Allainguillaume, J., Ronca, S., Tatarinova, T., Garbett, H., Walker, K., Wilkinson, M.J., 2012. DNA Barcoding the Native Flowering Plants and Conifers of Wales. *PLoS ONE* **7**, e37945.

Deans, A.S.C., 1957. *Survey of British Honey Sources*. Bee Research Association.

Deans, A.S.C., 1958. The Pollen Analysis of some British Honeys. Thesis for the National Bee Diploma.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* **26**, 5872–5895.

Desneux, N., Decourtye, A., Delpuech, J.-M., 2007. The sublethal effects of pesticides on beneficial arthropods. *Annual Review of Entomology* **52**, 81–106.

Di Pasquale, G., Salignon, M., Le Conte, Y., Belzunces, L.P., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L., Alaux, C., 2013. Influence of Pollen Nutrition on Honey Bee Health: Do Pollen Quality and Diversity Matter? *PloS ONE* **8**.

Doner, L.W., 1977. The Sugars of Honey—A Review. *Journal of the Science of Food and Agriculture* **28**, 443–456.

Doublet, V., Labarussias, M., de Miranda, J.R., Moritz, R.F. a., Paxton, R.J., 2015. Bees under stress: sublethal doses of a neonicotinoid pesticide and pathogens interact to elevate honey bee mortality across the life cycle. *Environmental Microbiology* **17**, 969–983.

Dziba, L., Erpul, G., Fazel, A., Fischer, M., Hernández, A.M., 2016. Summary for

policymakers of the thematic assessment on pollinators, pollination and food production. *Biota Neotropica* **16**.

Estoup, A., Garnery, L., Solignac, M., Cornuet, J.M., 1995. Microsatellite variation in honey bee (Apis mellifera L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* **140**, 679–95.

Fewell, J.H., Winston, M.L., 1992. Colony state and regulation of pollen foraging in the honey bee, *Apis mellifera* L. *Behavioral Ecology and Sociobiology* **30**, 387–393.

Fewell, J.H., Winston, M.L., 1996. Regulation of nectar collection in relation to honey storage levels by honey bees, *Apis mellifera*. *Behavioral Ecology* **7**, 286–291.

Fleming, J.C., Schmehl, D.R., Ellis, J.D., 2015. Characterizing the impact of commercial pollen substitute diets on the level of Nosema spp. in Honey Bees (Apis mellifera L.). *PloS ONE* **10**.

Galimberti, A., De Mattia, F., Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., Casiraghi, M., Labra, M., 2014. A DNA Barcoding Approach to Characterize Pollen Collected by Honeybees. *PloS ONE* **9**.

Garbuzov, M., Ratnieks, F.L.W., 2014a. Quantifying variation among garden plants in attractiveness to bees and other flower-visiting insects. *Functional Ecology* **28**, 364–374.

Garbuzov, M., Ratnieks, F.L.W., 2014b. Listmania: The Strengths and Weaknesses of Lists of Garden Plants to Help Pollinators. *BioScience* **64**, 1019–1026.

Garbuzov, M., Schürch, R., Ratnieks, F.L.W., 2015. Eating locally: dance decoding demonstrates that urban honey bees in Brighton, UK, forage mainly in the surrounding urban area. *Urban Ecosystems* **18**, 411–418.

Garibaldi, L.A., Steffan-Dewenter, I., Kremen, C., Morales, J.M., Bommarco, R., Cunningham, S. a., Carvalheiro, L.G., Chacoff, N.P., Dudenhöffer, J.H., Greenleaf, S.S., Holzschuh, A., Isaacs, R., Krewenka, K., Mandelik, Y., Mayfield, M.M., Morandin, L. a., Potts, S.G., Ricketts, T.H., Szentgyörgyi, H., Viana, B.F., Westphal, C., Winfree, R., Klein, A.M., 2011. Stability of pollination services decreases with isolation from natural areas despite honey bee visits. *Ecology Letters* **14**, 1062–1072.

Goulson, D., Lye, G.C., Darvill, B., 2008. Decline and conservation of bumble bees. *Annual Review of Entomology* **53**, 191–208.

Goulson, D., Nicholls, E., Botias, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **347**.

Graham, S., Myerscough, M.R., Jones, J.C., Oldroyd, B.P., 2006. Modelling the role of intracolonial genetic diversity on regulation of brood temperature in honey bee (Apis mellifera L.) colonies. *Insectes Sociaux* **53**, 226–232.

Green, B.H., 1990. Agricultural intensification and the loss of habitat, species and amenity in British grasslands: a review of historical change and assessment of future prospects. *Grass and Forage Science* **45**, 365–372.

Greenleaf, S.S., Williams, N.M., Winfree, R., Kremen, C., 2007. Bee foraging ranges and their relationship to body size. *Oecologia* **153**, 589–596.

Grüter, C., Moore, H., Firmin, N., Helanterä, H., Ratnieks, F.L.W., 2011. Flower constancy in honey bee workers (Apis mellifera) depends on ecologically realistic rewards. *The Journal of Experimental Biology* **214**, 1397–1402.

Guertler, P., Eicheldinger, A., Muschler, P., Goerlich, O., Busch, U., 2014. Automated DNA extraction from pollen in honey. *Food Chemistry* **149**, 302–306.

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

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., Baillie, L., Adams-Groom, B., 2015. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. *PLOS ONE* **10**.

Haydak, M.H., 1970. Honey Bee Nutrition. *Annual Review of Entomology* **15**, 143–156.

Henry, M., Béguin, M., Requier, F., Rollin, O., Odoux, J.-F.J.F., Aupinel, P., Aptel, J., Tchamitchian, S., Decourtye, A., Beguin, M., Requier, F., Rollin, O., Odoux, J.-F.J.F., Aupinel, P., Aptel, J., Tchamitchian, S., Decourtye, A., 2012. A Common Pesticide Decreases Foraging Success and Survival in Honey Bees. *Science* **337**, 1453–1453.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. *PloS ONE* **6**.

Hrassnigg, N., Brodschneider, R., Fleischmann, P.H., Crailsheim, K., 2005. Unlike nectar foragers, honeybee drones (Apis mellifera) are not able to utilize starch as fuel for flight. *Apidologie* **36**, 547–557.

Jain, S.A., Jesus, F.T. De, Marchioro, G.M., Araújo, E.D. De, 2013. Extraction of DNA from honey and its amplification by PCR for botanical identification. *Food Science and Technology (Campinas)* **33**, 753–756.

Jones, J.C., 2004. Honey Bee Nest Thermoregulation: Diversity Promotes Stability. *Science* **305**, 402–404.

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

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., Ohe, K. Von Der, Ohe, W. Von Der,

Rost, S., Hartel, S., Steffan-Dewenter, I., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology* **17**, 558–566.

Kessler, S.C., Tiedeken, E.J., Simcock, K.L., Derveau, S., Mitchell, J., Softley, S., Stout, J.C., Wright, G. a., 2015. Bees prefer foods containing neonicotinoid pesticides. *Nature* **521**, 74–76.

Klatt, B.K., Holzschuh, A., Westphal, C., Clough, Y., Smit, I., Pawelzik, E., Tscharntke, T., 2014. Bee pollination improves crop quality, shelf life and commercial value. *Proceedings of the Royal Society B: Biological Sciences* **281**.

Klee, J., Besana, A.M., Genersch, E., Gisder, S., Nanetti, A., Tam, D.Q., Chinh, T.X., Puerta, F., Ruz, J.M., Kryger, P., Message, D., Hatjina, F., Korpela, S., Fries, I., Paxton, R.J., 2007. Widespread dispersal of the microsporidian Nosema ceranae, an emergent pathogen of the western honey bee, Apis mellifera. *Journal of Invertebrate Pathology* **96**, 1–10.

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences* **274**, 303–313.

Knight, T.M., Ashman, T.L., Bennett, J.M., Burns, J.H., Passonneau, S., Steets, J.A., 2018. Reflections on, and visions for, the changing field of pollination ecology. *Ecology Letters* **21**, 1282–1295.

Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., den Dunnen, J.T., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources* **15**, 8–16.

Laube, I., Hird, H., Brodmann, P., Ullmann, S., Schöne-Michling, M., Chisholm, J., Broll, H., 2010. Development of primer and probe sets for the detection of plant species in honey. *Food Chemistry* **118**, 979–986.

Le Conte, Y., Navajas, M., 2008. Climate change: impact on honey bee populations and diseases. *Revue scientifique et technique (International Office of Epizootics)* **27**.

Long, E.Y., Krupke, C.H., 2016. Non-cultivated plants present a season-long route of pesticide exposure for honey bees. *Nature Communications* **7**, 11629.

Louveaux, J., Maurizio, A., Vorwohl, G., 1978. Methods of melissopalynology. *Bee World* **5**, 139–153.

Mao, W., Schuler, M.A., Berenbaum, M.R., 2013. Honey constituents up-regulate detoxification and immunity genes in the western honey bee Apis mellifera. *Proceedings of the National Academy of Sciences* **110**, 8842–8846.

Mattila, H.R., Seeley, T.D., 2007. Genetic diversity in honey bee colonies enhances

productivity and fitness. *Science* **317**, 362–364.

Meixner, M.D., Costa, C., Kryger, P., Hatjina, F., Bouga, M., Ivanova, E., Büchler, R., 2010. Conserving diversity and vitality for honey bee breeding. *Journal of Apicultural Research* **49**, 85–92.

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

Minnaar, C., Anderson, B., 2019. Using quantum dots as pollen labels to track the fates of individual pollen grains. *Methods in Ecology and Evolution* **10**, 604–614.

Mitchell, E.A.D., Mulhauser, B., Mulot, M., Mutabazi, A., Glauser, G., Aebi, A., 2017. A worldwide survey of neonicotinoids in honey. *Science* **358**, 109–111.

Muñoz, I., Henriques, D., Johnston, J.S., Chávez-Galarza, J., Kryger, P., Pinto, M.A., 2015. Reduced SNP Panels for Genetic Identification and Introgression Analysis in the Dark Honey Bee (Apis mellifera mellifera). *PloS ONE* **10**.

Nicolson, S.W., Nepi, M., Pacini, E., 2007. *Nectaries and Nectar*. Springer.

Pahl, M., Zhu, H., Tautz, J., Zhang, S., 2011. Large scale homing in honeybees. *PloS ONE* **6**.

Palmeri, V., Scirtò, G., Malacrinò, A., Laudani, F., Campolo, O., 2015. A scientific note on a new pest for European honeybees: first report of small hive beetle Aethina tumida, (Coleoptera: Nitidulidae) in Italy. *Apidologie* **46**, 527–529.

Percival, M., 1947. Pollen collection by *Apis mellifera*. *New Phytologist* **46**, 142–165.

Percival, M.S., 1961. Types of nectar in angiosperms. *New Phytologist* **60**, 235–281.

Pernal, S.F., Currie, R.W., 2000. Pollen quality of fresh and 1-year-old single pollen diets for worker honey bees (Apis mellifera L.). *Apidologie* **31**, 387–409.

Pernal, S.F., Currie, R.W., 2001. The influence of pollen quality on foraging behavior in honeybees (*Apis mellifera* L.). *Behavioral Ecology and Sociobiology* **51**, 53–68.

Petit, S., Stuart, R.C., Gillespie, M.K., Barr, C.J., 2003. Field boundaries in Great Britain: Stock and change between 1984, 1990 and 1998. *Journal of Environmental Management* **67**, 229–238.

Pettis, J.S., VanEngelsdorp, D., Johnson, J., Dively, G., 2012. Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. *Naturwissenschaften* **99**, 153–158.

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., Carreck, N.L., Johnston, J.S., 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**, 269–278.

Pompanon, F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N., Taberlet, P., 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* **21**, 1931–50.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010a. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25**, 345–353.

Potts, S.G., Roberts, S., Dean, R., Marris, G., Brown, M., Jones, R., Neumann, P., Settele, J., 2010b. Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research* **49**, 15.

Requier, F., Odoux, J.-F., Tamic, T., Moreau, N., Henry, M., Decourtye, A., Bretagnolle, V., 2015. Honey bee diet in intensive farmland habitats reveals an unexpectedly high flower richness and a major role of weeds. *Ecological Applications* **25**, 881–890.

Requier, F., Odoux, J.F., Henry, M., Bretagnolle, V., 2017. The carry-over effects of pollen shortage decrease the survival of honeybee colonies in farmlands. *Journal of Applied Ecology* **54**, 1161–1170.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015. Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. *Applications in Plant Sciences* **3**.

Riley, J.R., Greggers, U., Smith, A.D., Reynolds, D.R., Menzel, R., 2005. The flight paths of honeybees recruited by the waggle dance. *Nature* **225**, 205–207.

Robinson, R.A., Sutherland, W.J., 2002. Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology* **39**, 157–176.

Rosenkranz, P., Aumeier, P., Ziegelmann, B., 2010. Biology and control of Varroa destructor. *Journal of Invertebrate Pathology* **103**, S96–S119.

Roulston, T.H., Cane, J.H., 2000. Pollen nutritional content and digestibility for animals. *Plant Systematics and Evolution* **222**, 187–209.

Ruttner, F., Tassencourt, L., Louveaux, J., 1978. Biometrical-Statistical Analysis of the Geographic Variability of Apis Mellifera L. *Apidolodie* **9**, 363–381.

Salisbury, A., Armitage, J., Bostock, H., Perry, J., Tatchell, M., Thompson, K., 2015. Enhancing gardens as habitats for flower-visiting aerial insects (pollinators): should we plant native or exotic species? *Journal of Applied Ecology* **52**, 1156–1164.

Sanchez-Bayo, F., Goka, K., 2014. Pesticide Residues and Bees – A Risk Assessment. *PloS ONE* **9**.

Schäfer, M.O., Cardaio, I., Cilia, G., Cornelissen, B., Crailsheim, K., Formato, G., Lawrence, A.K., Le Conte, Y., Mutinelli, F., Nanetti, A., Rivera-Gomis, J., Teepe, A., Neumann, P.,

2019. How to slow the global spread of small hive beetles, Aethina tumida. *Biological Invasions* **21**, 1451–1459.

Schifferer, G., 1952. Uber die Entfernungsangabe bei den Tanzen der Bienen. Munich.

Schmickl, T., Crailsheim, K., 2004. Inner nest homeostasis in a changing environment with special emphasis on honey bee brood nursing and pollen supply. *Apidologie* **35**, 249–263.

Schürch, R., Couvillon, M.J., Burns, D.D.R., Tasman, K., Waxman, D., Ratnieks, F.L.W., 2013. Incorporating variability in honey bee waggle dance decoding improves the mapping of communicated resource locations. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* **199**, 1143–1152.

Scofield, H.N., Mattila, H.R., 2015. Honey bee workers that are pollen stressed as larvae become poor foragers and waggle dancers as adults. *PloS ONE* **10**.

Seeley, T.D., 1986. Social foraging by honeybees: how colonies allocate foragers among patches of flowers. *Behavioral Ecology and Sociobiology* **19**, 343–354.

Seeley, T.D., 1995. *The Wisdom of the Hive*. Harvard University Press.

Seeley, T.D., Tarpy, D.R., 2007. Queen promiscuity lowers disease within honeybee colonies. *Proceedings. Biological sciences / The Royal Society* **274**, 67–72.

Seeley, T.D., Towne, W.F., 1992. Tactics of dance choice in honey bees: do foragers compare dances? *Behavioral Ecology and Sociobiology* **30**, 59–69.

Simon-Delso, N., Amaral-Rogers, V., Belzunces, L.P., Bonmatin, J.M., Chagnon, M., Downs, C., Furlan, L., Gibbons, D.W., Giorio, C., Girolami, V., Goulson, D., Kreutzweiser, D.P., Krupke, C.H., Liess, M., Long, E., McField, M., Mineau, P., Mitchell, E.A.D., Morrissey, C.A., Noome, D.A., Pisa, L., Settele, J., Stark, J.D., Tapparo, A., Van Dyck, H., Van Praagh, J., Van der Sluijs, J.P., Whitehorn, P.R., Wiemers, M., 2015. Systemic insecticides (neonicotinoids and fipronil): trends, uses, mode of action and metabolites. *Environmental Science and Pollution Research* **22**, 5–34.

Singaravelan, N., Nee'man, G., Inbar, M., Izhaki, I., 2005. Feeding responses of free-flying honeybees to secondary compounds mimicking floral nectars. *Journal of Chemical Ecology* **31**, 2791–2804.

Soares, S., Amaral, J.S., Oliveira, M.B.P.P., Mafra, I., 2015. Improving DNA isolation from honey for the botanical origin identification. *Food Control* **48**, 130–136.

Southwick, E.E., Buchmann, S.L., 1995. Effects of Horizon Landmarks on Homing Success in Honey Bees. *American Society of Naturalists* **146**, 748–764.

Standifer, L.N., 1967. A comparison of the protein quality of pollens for growth-stimulation of the hypopharyngeal glands and longevity of honey bees, *Apis mellifera* L. (Hymenoptera: Apidae). *Insectes Sociaux* **14**, 415–425.

Steffan-Dewenter, I., Kuhn, A., 2003. Honeybee foraging in differentially structured landscapes. *Proceedings of the Royal Society B: Biological Sciences* **270**, 569–575.

Stevens, Hopkins, Mackintosh, Rimes, Robertson, Blackstock, Jefferson, 2003. The extent of semi-natural grassland communities in lowland England and Wales: a review of conservation surveys 1978-96. *Grass and Forage Science* **54**, 1–18.

Synge, A.D., 1947. Pollen collection by honeybees (*Apis mellifera*). *Journal of Animal Ecology* **16**, 122–138.

Tarpy, D.R., 2003. Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 99–103.

Valentini, A., Miquel, C., Taberlet, P., 2010. DNA Barcoding for Honey Biodiversity. *Diversity* **2**, 610–617.

van Engelsdorp, D., Hayes, J., Underwood, R.M., Pettis, J., 2008. A survey of honey bee colony losses in the U.S., Fall 2007 to Spring 2008. *PloS ONE* **3**.

Vandame, R., Meled, M., Colin, M.-E., Belzunces, L.P., 1995. Alteration of the homing-flight in the honey bee Apis mellifera L. exposed to sublethal dose of deltamethrin. *Environmental Toxicology and Chemistry* **14**, 855–860.

VanEngelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpy, D.R., Pettis, J.S., 2009. Colony collapse disorder: a descriptive study. *PloS ONE* **4**.

Vaudo, A.D., Tooker, J.F., Grozinger, C.M., Patch, H.M., 2015. Bee nutrition and floral resource restoration. *Current Opinion in Insect Science* **10**, 133–141.

Vidau, C., Diogon, M., Aufauvre, J., Fontbonne, R., Viguès, B., Brunet, J.L., Texier, C., Biron, D.G., Blot, N., Alaoui, H., Belzunces, L.P., Delbac, F., 2011. Exposure to sublethal doses of fipronil and thiacloprid highly increases mortality of honeybees previously infected by nosema ceranae. *PloS ONE* **6**.

Visscher, P.K., Seeley, T.D., 1982. Foraging Strategy of Honeybee Colonies in a Temperate Deciduous Forest. *Ecology* **63**, 1790.

Waiblinger, H.-U., Ohmenhaeuser, M., Meissner, S., Schillinger, M., Pietsch, K., Goerlich, O., Mankertz, J., Lieske, K., Broll, H., 2012. In-house and interlaboratory validation of a method for the extraction of DNA from pollen in honey. *Journal für Verbraucherschutz und Lebensmittelsicherheit* **7**, 243–254.

Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., Simões, Z.L.P., Allsopp, M.H., Kandemir, I., De la Rúa, P., Pirk, C.W., Webster, M.T., 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee Apis mellifera. *Nature Genetics* **46**, 1081–1088.

Waller, G.D., 1972. Evaluating Responses of Honey Bees to Sugar Solutions Using an Artificial-Flower Feeder. *Annals of the Entomological Society of America* **65**, 857–862.

Webby, R., 2004. Floral origin and seasonal variation of bee-collected pollens from individual colonies in New Zealand. *Journal of Apicultural Research* **43**, 83–92.

Wells, H., Wells, P.H., 1983. Honey Bee Foraging Ecology : Optimal Diet , Minimal Uncertainty or Individual Constancy? *Journal of Animal Ecology* **52**, 829–836.

Wignall, V.R., Alton, K., Ratnieks, F.L.W., 2019. Garden centre customer attitudes to pollinators and pollinator-friendly planting. *PeerJ* **7**, e7088.

Williamson, S.M., Wright, G.A., 2013. Exposure to multiple cholinergic pesticides impairs olfactory learning and memory in honeybees. *The Journal of Experimental Biology* **216**, 1799–807.

Wilson, E.E., Sidhu, C.S., Levan, K.E., Holway, D.A., 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. *Molecular Ecology* **19**, 4823–4829.

Winston, M.L., 1987. *The Biology of the Honey Bee*. Harvard University Press.

Woodcock, B.A., Ridding, L., Freeman, S.N., Pereira, M.G., Sleep, D., Redhead, J., Aston, D., Carreck, N.L., Shore, R.F., Bullock, J.M., Heard, M.S., Pywell, R.F., 2018. Neonicotinoid residues in UK honey despite European Union moratorium. *PLOS ONE* **13**, e0189681.

Wright, G.A., Baker, D.D., Palmer, M.J., Stabler, D., Mustard, J.A., Power, E.F., Borland, A.M., Stevenson, P.C., 2013. Caffeine in Floral Nectar Enhances a Pollinator's Memory of Reward. *Science* **339**, 1202–1204.

Wright, G.A., Nicolson, S.W., Shafir, S., 2018. Nutritional Physiology and Ecology of Honey Bees. *Annual Review of Entomology* **63**, 327–344.

Wykes, G.R., 1952. The Preferences of Honeybees for Solutions of Various Sugars Which Occur in Nectar. *Journal of Experimental Biology* **29**, 511.

Yang, E.C., Chuang, Y.C., Chen, Y.L., Chang, L.H., 2008. Abnormal foraging behavior induced by sublethal dosage of imidacloprid in the honey bee (Hymenoptera: Apidae). *Journal of Economic Entomology* **101**, 1743–1748.

CHAPTER 2

Creating a DNA barcode reference library for the flowering plants and conifers of the UK

2.1. Abstract

DNA barcoding and metabarcoding provide new avenues for investigating biological systems and these techniques are based upon the need for well curated reference libraries. In plant DNA barcoding, herbarium collections represent a resource that provides taxonomically robust source material but also poses challenges in lab processing. Here we present a national DNA barcoding resource, which covers the native flowering plants and conifers of the United Kingdom. This represents 1,482 plant species, with the majority of specimens (81%) sourced from herbaria. Using the plant DNA barcode markers, *rbcL*, *matK*, and ITS2, at least one DNA barcode was retrieved from 97% of targeted species. We sampled from multiple individuals, resulting in a species coverage for *rbcL* of 95% (4,359 sequences), 89% for *matK* (3,116 sequences) and 74% for ITS2 (2,567 sequences). Recovery of sequences is shown to be lower for herbarium material compared to fresh collections, with the age of the specimen having a significant effect on the success of sequence recovery. The level of taxonomic discrimination was assessed, with good discrimination to genus achieved by all three markers individually (98-100%). Species level discrimination was highest with ITS2 for the UK native species with 85% of species discriminating to species level, compared to 77% for *matK* and 64% for *rbcL.* However, this increased species discrimination should be balanced against the ability to successfully retrieve a sequence, which is lowest for ITS2. To augment the UK native flora reference library for wider applications, the current available sequence data for *rbcL* and ITS2 was assessed for 5,586 UK native, naturalised, alien and horticultural plant species. Genus level representation was high for *rbcL* at 96%, while ITS2 was lower at 84%. Overall discrimination to at least genus level was high for the successfully returned sequences for both markers with 90% for *rbcL* and 96% for ITS2. These resources give a high level of coverage for a national native flora and provide the groundwork for the many applications which require plant identification from DNA.

2.2. Introduction

The identification of plant species is vitally important to the monitoring and conservation of biodiversity and is limited by the availability of taxonomic expertise. DNA barcoding, the method of characterising species using an internationally agreed region of DNA (Hebert et al., 2003), has been used to both characterise existing biodiversity and identify new or cryptic species (Hollingsworth et al., 2016). Furthermore, by using DNA, species identification has become possible where morphological identification was previously limited.

The applications of DNA barcoding and DNA metabarcoding techniques (when DNA barcoding is used in combination with high throughput sequencing from multispecies samples) cover a wide range of purposes, with the potential for rapid identification of species composition from many different sources of DNA (Deiner et al., 2017; Hollingsworth et al., 2016). In using DNA barcode resources, questions about ecological systems and community structure can be answered. In environmental monitoring, DNA metabarcoding has been used to detect the presence of a rare species (Harper et al., 2018) while pollinator communities have been revealed by retrieving insect DNA from the flowers they visit (Thomsen and Sigsgaard, 2019). It is also used in diet analysis, as when examining food partitioning in herbivores (Kartzinel et al., 2015) or the trophic specialisation of bats (Arrizabalaga-Escudero et al., 2019). Potential commercial applications involve using DNA barcoding to assess plant based products for sale, such as confirming the claimed identity of herbal medicines (Li et al., 2011), verifying the geographic origin of honey (Bruni et al., 2015), or checking the legality of species present in tea (De Boer et al., 2017; Stoeckle et al., 2011).

DNA barcode markers also provide taxonomic information which can be used to create phylogenetic trees. This has been used to investigate aspects such as plant evolution (Lim et al., 2014) and for phylogenetic community ecology (Kress et al., 2010). Across these applications, DNA can be extracted from a variety of plant material, including leaves and roots, as well as environmental samples such as soil (Yan et al., 2018). In addition, DNA metabarcoding can allow the identification of pollen from different sources, including characterising pollen from the air (Brennan et al., 2019), examining pollen transport

networks using the pollen from the bodies of insects (Lucas et al., 2018) and investigating the foraging preferences of honeybees using pollen from the honey (de Vere et al., 2017).

However, the ability of studies to identify species from unknown, mixed samples of DNA are built upon the need for well curated reference libraries, with associated voucher specimens and sample metadata (Hebert et al., 2003). By using a high-quality reference library, the accuracy achieved within the many applications which require DNA based species identification can be improved (Bell et al., 2016; Landi et al., 2014). The UK has a long history of botanical recording, with a well-studied flora (Preston et al., 2002; Walker and Preston, 2006). By using national herbarium collections, a comprehensive collection of UK plant species can be accessed for DNA barcoding (de Vere et al., 2012), providing accurate taxonomic identification, source material for DNA extraction and voucher specimens which can be attached to sequencing data. With herbarium collections, the cost and time associated with establishing a reference database can be significantly reduced (Kuzmina et al., 2017). In addition, with a well annotated, housed collection of specimens connected to their DNA sequences, the voucher specimens can be returned to for any further research purposes.

The agreed DNA barcode region for animals, the mitochondrial gene COI (Hebert et al., 2003), shows a low rate of nucleotide substitution in plants, making it unsuitable for identifying species differences. As such, the Consortium for the Barcode of Life (CBOL) recommended two core plastid DNA barcode regions for plants, *rbcL* (*c.* 600 bp) and *matK* (c. 800 bp), while noting that these markers may require complementing with alternative regions such as *trnH-psbA* (CBOL Plant Working Group, 2009) and the nuclear marker ITS in its entirety, or the ITS2 subsection (Hollingsworth, 2011). The ideal DNA barcode would provide an ability to amplify a wide range of taxa, while at the same time allowing a high level of species resolution through high interspecific divergence, with a correspondingly low intraspecific divergence. The initial choice of plant DNA barcodes, *rbcL* and *matK*, represented balancing these needs of taxonomic universality, the discrimination ability and the cost associated with using a large number of markers (CBOL Plant Working Group, 2009).

Here we present the creation of a DNA barcode reference library for the native and archaeophyte (naturalised prior to 1500 AD) flowering plants and conifers of the UK. This work builds upon the previous reference library created for the nation of Wales, within the UK (de Vere et al., 2012). For the Welsh flora DNA barcode database, 1,143 plant species were sequenced using *rbcL* and *matK.* In this study, an additional 339 plant species were targeted to gain representation for plant species not found within Wales, but present within England, Scotland and Northern Ireland. This reference library represents the whole UK native flora for *rbcL, matK,* and ITS2, providing a complete three locus library representing 1,482 species, 503 genera, 104 families and 36 orders. We assessed how successful sequence recovery is affected by specimen age and the plant order sampled for each marker, important considerations for plant DNA barcoding efforts focusing on herbarium collections.

To provide a reference library for applications which require non-native plant species within the UK, we assessed the current level of GenBank representation available for a curated list of native and naturalised, alien and horticultural UK plant species and created a reference library of these, suitable for wider applications. For both the UK native reference library and the UK native and non-native library we examined the relative taxonomic discrimination ability of the markers.

2.2.1. Aims and objectives

- Create a DNA barcode reference library for the UK native flowering plant and conifer species, using three DNA barcode loci, *rbcL*, *matK*, and ITS2.
    - How do *rbcL*, *matK*, and ITS2 vary in their ability to successfully recover a sequence, within the UK flora?
    - How do *rbcL*, *matK*, and ITS2 vary in their species level discrimination?
- Create a reference library including UK non-native plant species from all available DNA sequences on GenBank to support the applications of DNA barcoding.
    - How representative are the currently available sequences for the UK non-native and native plant species?
    - What is the species level discrimination ability within the reference library for the selected native and non-native UK plant species?

2.2.2. Author contributions and history of the project

DNA barcoding of the UK native flora began with the Welsh flora as published in de Vere *et al.* 2012, prior to the start of my PhD. The Welsh flora was completed with *rbcL* and *matK* plant DNA barcode regions. For the establishment of the UK DNA barcode reference library, additional species were either collected in the field or sampled from herbarium specimens. DNA barcodes for *rbcL, matK* and ITS2 were created for the additional UK species, while the Welsh flora was additionally barcoded with ITS2. This work formed part of my undergraduate placement at the National Botanic Garden of Wales, prior to my PhD. Field sampling for the UK native species was completed by Natasha de Vere, Tim Rich, and Heather McHaffie. I completed the herbarium sampling for the UK species with Helena Davies, Natasha de Vere, Adelaide Griffith, Joseph Moughan, and Tim Rich. Helena Davies, Joe Moughan Eleanor Brittain and I completed the labwork and DNA sequence processing. Sequencing was provided by Laura Forrest at Royal Botanic Garden, Edinburgh. For my PhD, I completed all of the data and statistical analyses presented here and created the curated reference library with UK non-native species to support DNA metabarcoding applications in subsequent chapters.

2.3. Methods and Analysis

2.3.1. Sample collection

The UK flora targeted here represents 1,482 native and archaeophyte flowering plants and conifers, representing 503 genera, 104 families, and 36 orders (Preston et al., 2002; Stace, 2010). Taxonomic classifications match Stace, (2010). The apomictic microspecies complexes of *Hieracium*, *Rubus* and *Taraxacum*, which are difficult to distinguish morphologically as species (Ellstrand et al., 1996), were represented using aggregate species groupings, while the apomictic genus *Sorbus* was sampled fully. In total, 6,096 individuals were sampled, 4,962 from herbarium specimens and 1,134 from fresh collection of leaf material throughout the UK. Of the 6,096 specimens, 4,272 were sampled and extracted during previous work on DNA barcoding the Welsh flora (de Vere et al., 2012), while 1,824 specimens represent new herbarium and fresh collections to gain coverage for those UK plant species not present in Wales. At least three individuals of each species were targeted for collection.

For samples from herbarium specimens, approximately 2 cm^2 of leaf material was removed where sampling would not detract from the scientific value of the specimen. Further criteria for specimen selection included as follows: being typical representations of the species, having additional taxonomic verification present for the specimen, being collected from geographically distinct locations, and being recently collected. The majority of herbarium sample collection came from specimens housed in the National Museum Wales collections (89%), with additional samples from the Royal Botanic Garden Edinburgh, National Museums Liverpool, Dublin Botanic Garden, Bangor University and Aberystwyth University.

For freshly collected material, approximately 2 cm^2 of undamaged leaf or flower material was sampled and placed into silica gel to dry. Regional floras, online databases from the Botanical Society of Britain and Ireland, and knowledge from local recorders were used to locate species for sampling. Herbarium vouchers were created for all freshly collected material, with the exception of threatened species, and entered collections at the

National Botanic Garden of Wales and National Museum Wales. Where collection of a voucher specimen was prohibited, a photograph was used instead.

2.3.2. DNA extraction and amplification

For freshly collected leaf material, Plant DNeasy kits (Qiagen) were used following the manufacturer's protocol using leaves dried in silica gel. For herbarium samples, the Plant DNeasy protocol was modified to improve success, following de Vere et al., (2012). This used a buffer of 400 µl AP1 from the Qiagen kit, 80 µl DTT (0.75 mg/ml) (Melford Laboratories, UK) and 20 µl proteinase K (1 mg/ml) (Sigma), from which 400 µl were added to the leaf material before disruption with a TissueLyser II (Qiagen) with 3 mm tungsten carbide beads. The incubation phase using the modified AP1 buffer was then extended to 1 hour at 65°C. The final elution stage with the AE buffer was extended to 15 minutes.

For the UK specimens not represented by the Welsh database, *rbcL* amplification was carried out using *rbcL*a-F and *rbcL*r590. For *matK,* multiple primer combinations were used, following de Vere et al., (2012), beginning with universal primer combinations, *matK*-390F with *matK*1326R, and *matK*-2.1a with *matK*-3Fkim-R. If this failed, order specific primers were then used. All of the specimens were additionally amplified for the ITS2 region using the ITS2F and ITS3R primers (Chiou et al., 2007; Yao et al., 2010). Only one primer pair was attempted to avoid producing sequences from different copies of ITS2, a multicopy nuclear marker (Yao et al., 2010).

PCR amplification was carried out in a 20 µl reaction, using 10 µl of Biomix (Bioline), 0.4 µl of forward primer (10 µM) and 0.4 µl of reverse primer (10 µM), 0.8 µl of BSA (1 mg/ml), 6.4 µl of molecular grade H_2O, and 2 µl of template DNA. PCR conditions for *rbcL* and *matK* were 95°C for 2 min, followed by 95°C for 30 sec, 50°C for 90 sec, and 72°C for 40 sec, for 45 cycles, followed by 72°C for 5 min and 30°C for 10 sec. The ITS2 PCR cycle was as follows: 94°C for 5 min, followed by 94°C for 30 sec, 56°C for 30 sec and 72 °C for 45 sec for 40 cycles, and then 72°C for 10 min. Samples were run on 1% agarose gels and successfully amplified samples were sent for forward and reverse Sanger sequencing to Macrogen Europe (Amsterdam, Europe) and to the Royal Botanic Gardens of Edinburgh on an ABI3730XL sequencer (Applied Biosystems).

2.3.3. Sequencing

For each returned sequence, the sequences were quality trimmed (with 25 bp window segments where more than 2 bp showing a quality value of less than 20 were removed), the primers were removed, and the contigs then assembled. Low quality sequences were removed, and for *rbcL* and *matK* sequences with stop codons were also removed. ITS2, as a non-coding region, could have stop codons present. Each contig was manually checked for base call disagreements and manually edited as needed. All sequence assembly and editing was completed using Sequencher v 5.0 (Gene Codes Corporation).

Sequences were verified to ensure their identification was correct. Quality control included comparing sequences from multiple individuals of species and creating neighbour-joining trees. Any species which were misplaced within the tree were investigated to check their identity. Sequences were also checked against available records on GenBank using BLAST.

The reference library, for all three loci, was deposited on to the Barcode of Life Database (BOLD) and GenBank (accessions: JN890545-JN896265; KX165423-KX167996; MK924423-MK926404). Each sequence on the BOLD database is available with the collection information, including location, collector, date collected, and a scan of the herbarium voucher.

2.3.4. Species recoverability

The ability to successfully retrieve a sequence from a species was summarised overall and by plant order. To assess the effect of year of sample collection and plant order on the successful recovery of a sequence, a binomial generalised linear model was fit with the proportion of successfully recovered sequences as the response variable. This was restricted to plant orders with ten or more species attempted. The effect of year of sample collection, plant order and the interaction between the two were included as explanatory variables. Each marker was fitted as a separate model. Model selection was based on the lowest Akaike information criterion score.

The success of sequence recoverability was also examined for the herbarium material separately. Year of collection was divided into nine classes for specimens from 1912 to

2010. Specimens from either side of the range were excluded, due to small sample size in the age class. The relationship between year of collection and sequence recovery was assessed with Spearman's rank correlation, for each marker. All analysis was completed in R v. 3.5.2 (R Development Core Team, 2011).

2.3.5. Species discrimination

The ability of the DNA barcode markers to discriminate to different levels of taxonomic identification was evaluated using BLAST searches which queried each sequence in turn against the database. Other methods of discrimination include looking for a barcode gap or for the presence of monophyletic groups in phylogenetic neighbour-joining trees. For the Welsh plant flora, similar results in the discrimination ability were found between using BLAST, barcode gaps or monophyletic groups (de Vere et al., 2012). As BLAST identification is a common method in DNA metabarcoding studies when assigning taxonomic information to unknown sequences (Deiner et al., 2017), the discrimination ability of the UK native reference library was assessed using BLAST. To allow comparison between the three markers, the BLAST database was restricted to the plant species which had multiple sequences for all three markers, giving 634 plant species. Each sequence was matched against a database which excluded the query sequence, and the level of discrimination assessed for a species, genus, family or order level match.

2.3.6. Creation of a DNA barcode reference library for UK native, naturalised and horticultural species

Certain applications of DNA barcoding may only require native species for identification but for other applications sequence coverage is required for all plants potentially growing within the UK, such as with pollinator foraging studies. To gain this additional coverage, in March 2018, all available *rbcL* and ITS2 plant records were downloaded from NCBI GenBank. The *matK* region was not targeted as, while it can contribute phylogenetic information, it is not often utilised in DNA metabarcoding studies due to the need for multiple primer pairs to gain taxonomic range, and its longer sequence length.

A python script was written (*creatingselectedfastadatabase.py*; available at https://github.com/colford/nbgw-plant-illumina-pipeline) to filter the available sequences

from GenBank using a list of species known to grow in the UK. This used the list of native species of the UK (Stace, 2010) and naturalised and alien species (505 species, Preston et al., 2002). In order to represent the horticultural plants potentially available in the UK, horticultural planting records and plant surveys from a botanic garden were used. In total, this represented 5,586 plant species, covering the UK native, naturalised and horticultural species and included angiosperms and gymnosperms. A limited number of ferns and bryophyte species were included. Firstly, species were extracted from GenBank which matched the species list at species level. For the plants in the list not represented at species level within the GenBank records, a second extraction matching at the genus level was completed. The coverage of the native plants included the sequences generated here for the UK native species reference library. The discrimination ability of the created reference library was then assessed for the 1,732 species which were present in both the *rbcL* and ITS2 libraries with multiple sequences. A BLAST database for each marker was created using the sequences, and each record was queried against the database in turn, with the query sequence removed from the database.

2.4. Results

2.4.1. Recoverability within the UK native species

For the 1,482 native and archaeophyte flowering plants and conifers of the UK, a total of 10,042 barcode sequences were recovered with *rbcL*, *matK*, and ITS2 (Table 2). This represented 4,359 sequences for *rbcL* covering 95% of species, 98% of genera and 99% of families (Figure 2a). For *matK,* 3,116 sequences were recovered across 89% of species, 92% of genera and 92% families, while 2,567 ITS2 sequences, representing 74% of species, 79% genera, 82% families were created (Figure 2a). All three markers were obtained for 1,016 species (69% of the UK flora), while 97% of plant species, 98% of genera and 100% of families were represented with at least one marker (Table 2).

Table 2: Summary statistics for the DNA barcode database for the UK flora of 1,482 species.

	rbcL	***matK***	**ITS2**	**All Markers**
Number of species successfully DNA barcoded	1414 (95%)	1314 (89%)	1101 (74%)	1016 (69%)
Species with more than one individual DNA barcoded	1288 (87%)	1009 (68%)	778 (52%)	634 (43%)
Mean (SD) number of DNA barcodes per species	3.1 (1.3)	2.4 (1.3)	2.3 (1.4)	8.1 (3.5)
Range of individuals DNA barcoded per species	1-12	1-22	1-15	3-49
Total number of DNA barcodes	4359	3116	2567	10,042

When looking at sequence success at the family level, a sequence recovery of above 50% of specimens was seen for 84% plant families in *rbcL,* 50% of families with *matK*, and 32% families in ITS2. Malvaceae was the most consistently successful family with over 80% sequence recovery across all three markers. Looking at the families with ten or more specimens, *rbcL* was able to recover sequences from all families. ITS2 was unable to recover any sequences from four plant families (Araceae, Ruppiaceae, Zosteraceae and Iridaceae), all of which were monocots. There were four plant families (Pinaceae, Elatinaceae, Hypericaceae, and Polygalaceae) that *matK* was unable to recover sequences from.

Recoverability was tested at the order level for plant orders with ten or more species (Figure 3a). For *rbcL* both the year of sample collection (Likelihood ratio; $LR_{1,20}$ = 808.77, p < 0.001) and plant order ($LR_{1,20}$ = 322.16, p < 0.001) were found to significantly predict the success of sequence recovery, with a significant interaction also found between plant order and year ($LR_{1,20}$ = 93.40, p < 0.001), suggesting that plant orders have different sensitivity in the ability to recover a sequence from older samples with different patterns of DNA degradation. The same pattern was seen for both *matK*, with year ($LR_{1,20}$ = 1027.34 , p < 0.001), plant order ($LR_{1,20}$ = 257.53, p < 0.001) and the interaction ($LR_{1,20}$ = 106.02, p < 0.001), and for ITS2 with year ($LR_{1,20}$ = 576.00, p < 0.001), plant order ($LR_{1,20}$ = 377.51, p < 0.001) and the interaction ($LR_{1,20}$ = 91.02, p < 0.001).

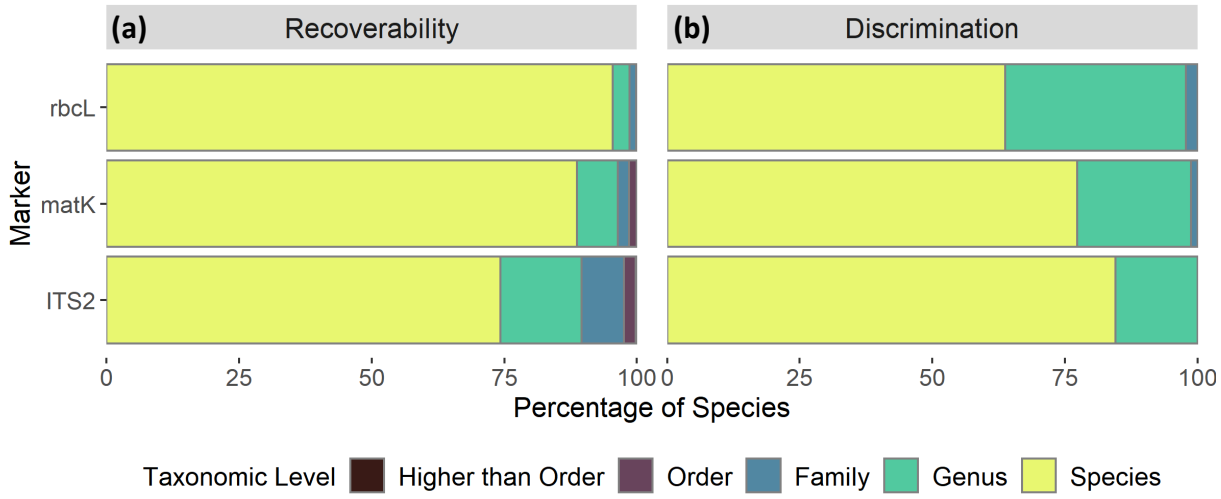


Figure 2: The overall level of taxonomic representation in the database for the species of the UK flora. Recoverability **(a)** shows the level of representation for the native species of the UK flora (n = 1,482) in the reference library. Discrimination **(b)** shows the taxonomic resolution achieved using BLAST for those plant species in the reference library which were represented by all three markers more than once (n = 634).

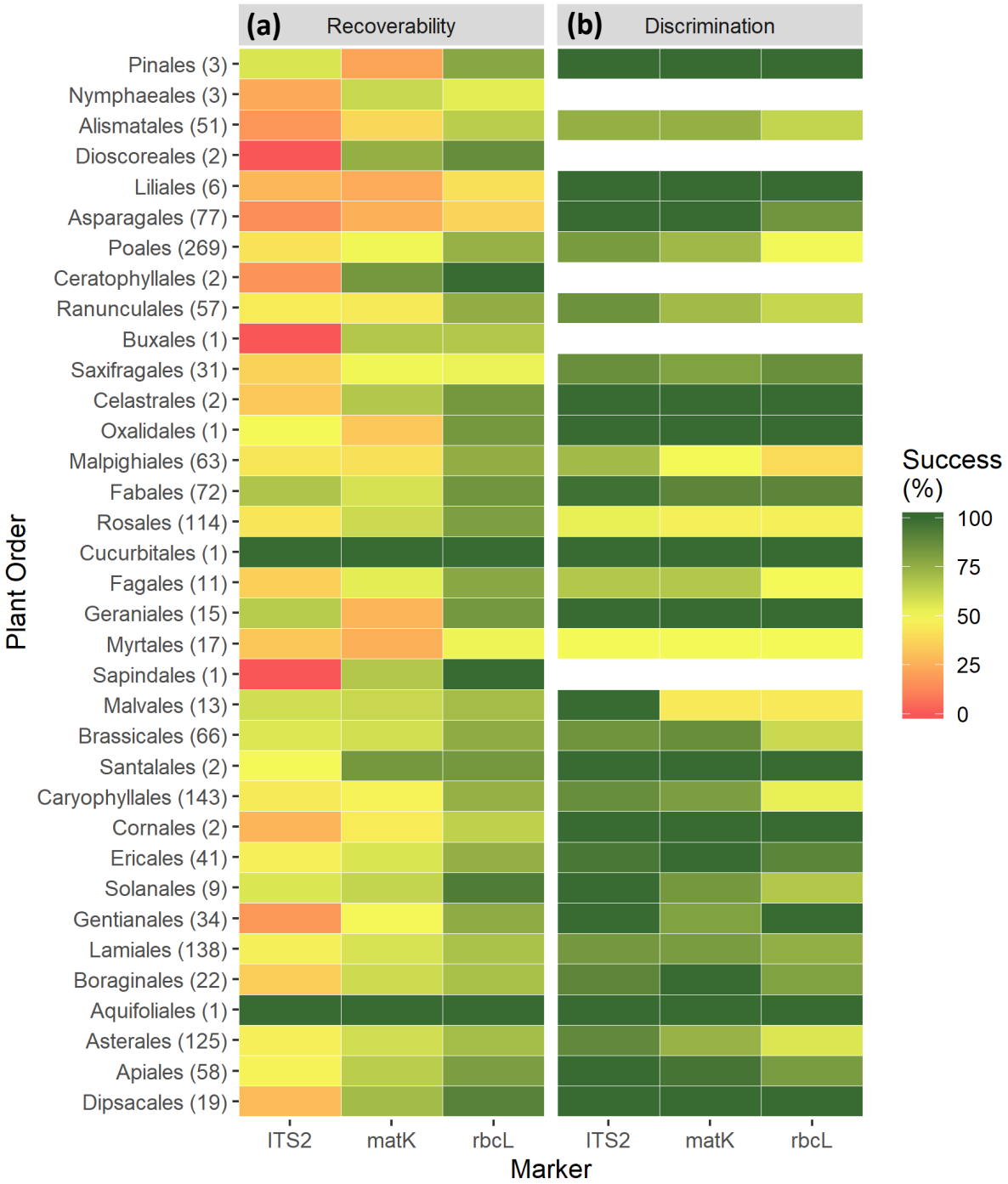


Figure 3: Species level recoverability **(a)** and discrimination **(b)** by plant order for each marker. Recoverability shows the percentage of specimens in each plant order which were successfully sequenced. The discrimination level was assessed for plant species which were represented by all three markers more than once, showing the percentage of species in each plant order which were identified to species level. Number of species represented by an order is shown in brackets. The mode number of specimens per species was 3.

2.4.2. Discrimination within the UK native species

Discrimination ability was assessed for 634 of the sequenced species, which had multiple individuals sequenced for all markers (Figure 2b). For *rbcL,* 64% of species were returned as a species level match, 98% were a genus level match and 100% were a family level match. For *matK* there was greater species discrimination with 77% of species returned at species level, 99% at genus and 100% at family. ITS2 had the highest level of species discrimination with 85% to species and 100% returned to genus.

There were some similar patterns across the three markers for discrimination ability, with reduced species level discrimination in Myrtales in distinguishing *Epilobium* species, as well as the species rich Rosales, with *rbcL* and *matK* distinguishing 47% of species to species level, and ITS2 distinguishing 52% (Figure 3b). ITS2 showed increased discrimination ability compared to *matK* and *rbcL* in Malpighiales and Malvales. With *rbcL* the lowest species level discrimination achieved in a plant order was 39% (Malpighiales), while in *matK* the lowest was 44% (Malvales), and for ITS2 it was 50% (Myrtales).

2.4.3. Comparison between recoverability from herbarium and fresh material

In total, 6,096 specimens were collected, 4,962 from herbarium specimens and 1,134 from fresh collections. For all three markers, freshly collected leaf material was significantly more likely to yield a successful DNA barcode compared to herbarium material. Looking at *rbcL*, overall 72% of specimens yielded a sequence, with 86% success from fresh samples, and 68% from herbarium samples (chi-squared test, with Yates correction; $x^2 = 145.85$, d.f = 1, $p < 0.001$). A lower sequence recoverability was found for *matK,* with 51% of specimens working overall, 73% for fresh material and 46% for herbarium ($x^2 = 277.19$, d.f = 1, $p < 0.001$). ITS2 showed the lowest overall recoverability at 42% of samples (65% for fresh material and 37% for herbarium specimens ($x^2 = 305.01$, d.f = 1, $p < 0.001$), but only one primer pair was attempted to avoid sequencing different versions of the multi-copy marker.

Looking at the success of herbarium specimens alone, a significant negative correlation between the success of sequence recovery and the age of the herbarium specimen was found, with greater success of sequencing for more recently collected specimens (Figure 4) (Spearman's rank correlation coefficient for *rbcL* r_s = 0.973, p < 0.001; for *matK*, r_s = 0.954, p < 0.001; for ITS2, r_s = 0.982, p < 0.001). The collection age of herbarium specimens ranged from 1868 to 2011.

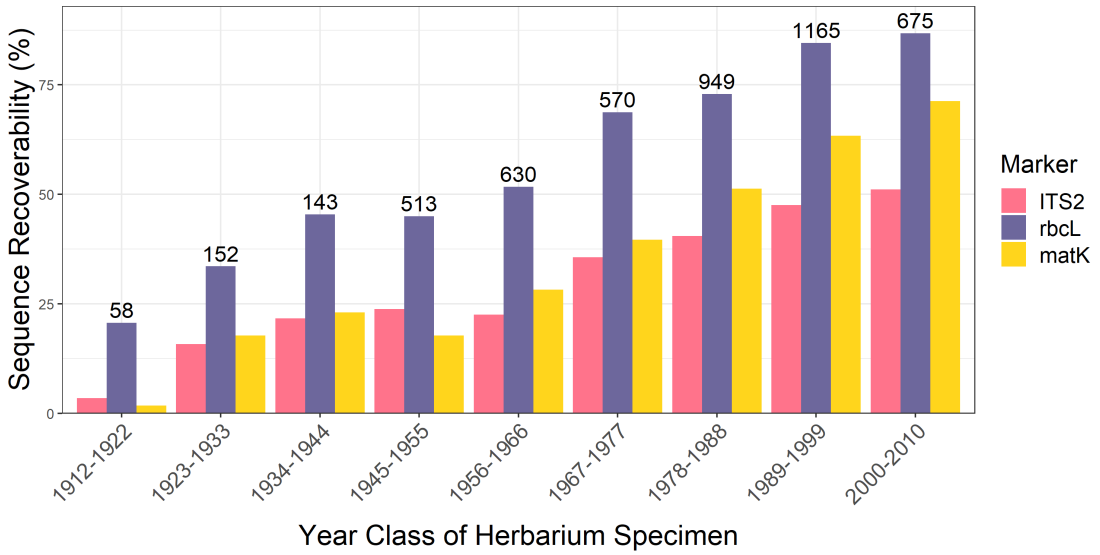


Figure 4: Effect of herbarium specimen age on sequence recoverability. The success of 4,962 herbarium samples was assessed over nine year classes, between 1908-2007. The sample size of each age class is annotated above the bar. There was a significant negative correlation for all three markers, between success of sequence recoverability and age of the herbarium specimen. (Spearman's rank correlation coefficient for *rbcL* r_s = 0.973, p<0.0001; for *matK*, r_s =0.954, p<0.0001; for ITS2, r_s =0.982, p<0.0001)

2.4.4. GenBank representation of UK naturalised, alien and horticultural species

For the 5,586 UK native, naturalised, alien and horticultural species targeted from GenBank records, 32,319 sequences were extracted in total for *rbcL*, representing 57% of species, with genus level coverage at 96% (Figure 5a). For ITS2, the coverage was lower for both species and genus, with 52% species level coverage and 84% at genus level, from 27,476 sequences (Table 3). The species level representation and genus level representation were summarised across plant order for *rbcL* and ITS2 (Figure 6). Genus level representation was good across all orders for *rbcL,* with the lowest being 83% in Dioscroreales. Genus level representation ranged from 0% to 100% for ITS2 at the order level. Concentrating on orders with over ten species, genus level representation was low in the monocot orders Commelinales (25%), Arecales (29%), as well as the species rich Asparagales (68%). Representation was also low in the eudicot order Proteales (35%), and in the ferns Polypodiales (56%).

Table 3: Summary statistics for the DNA barcode database targeting 5,586 plant species which grow within the UK.

	rbcL	**ITS2**
Number of NCBI records extracted at species level	28,879	24,321
Number of unique species extracted at species level	3,308 (57%)	2,909 (52%)
Number of NCBI records extracted at genus level	3,440	3,155
Number of unique genera extracted at genus level	220	192
Number of plant families represented	235	209
Total records in reference database	32,319	27,476

2.4.5. Discrimination within UK native, naturalised and horticultural species

The discrimination ability of the database was assessed using BLAST for 1,732 plant species which were present in both the *rbcL* and ITS2 databases with multiple sequences (Figure 5b). Using BLAST, *rbcL* was able to discriminate to species for 58.9% of species, to genus 90.9%, to family 98.8%, to order 99.4% and to higher than order for 0.6% of species. ITS2 was able to discriminate to species more often, achieving this for 81.6% of species present in the database. Discrimination to genus for ITS2 occurred for 96.4% of species, family for 99.1%, to order for 91.3% of sequences and higher than order 0.7%. Both *rbcL* and ITS2 show a small decrease in species level and genus level discrimination with the native and non-native species compared to native plants alone. The ability to discriminate to species and the ability to discriminate to genus were summarised by plant order for the UK native and non-native species (Figure 6). Genus level discrimination was generally good across the plant orders for ITS2 (>85%), although species in the monocot plant orders Commelinales and Arecales, as well as the magnoliids orders Magnoliales and Laurales performed poorly in achieving genus level discrimination. For *rbcL* genus level discrimination was also high across orders (>85%), with poor performance in orders with less than ten species in the database (Canellales, Buxales, Santalales).

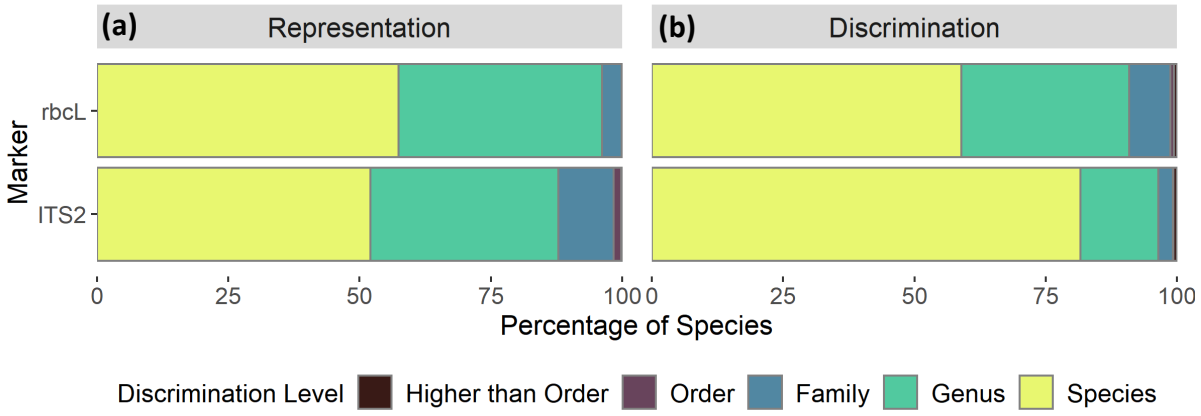


Figure 5: The overall taxonomic representation and relative discrimination ability for the list of UK archaeophytes, naturalised, alien and horticultural species present in GenBank. **(a)** The taxonomic representation in the reference library for the 5,586 targeted plant species. **(b)** The level of discrimination ability for 1,732 of the successfully extracted species UK represented by both *rbcL* and ITS2 with multiple individuals.

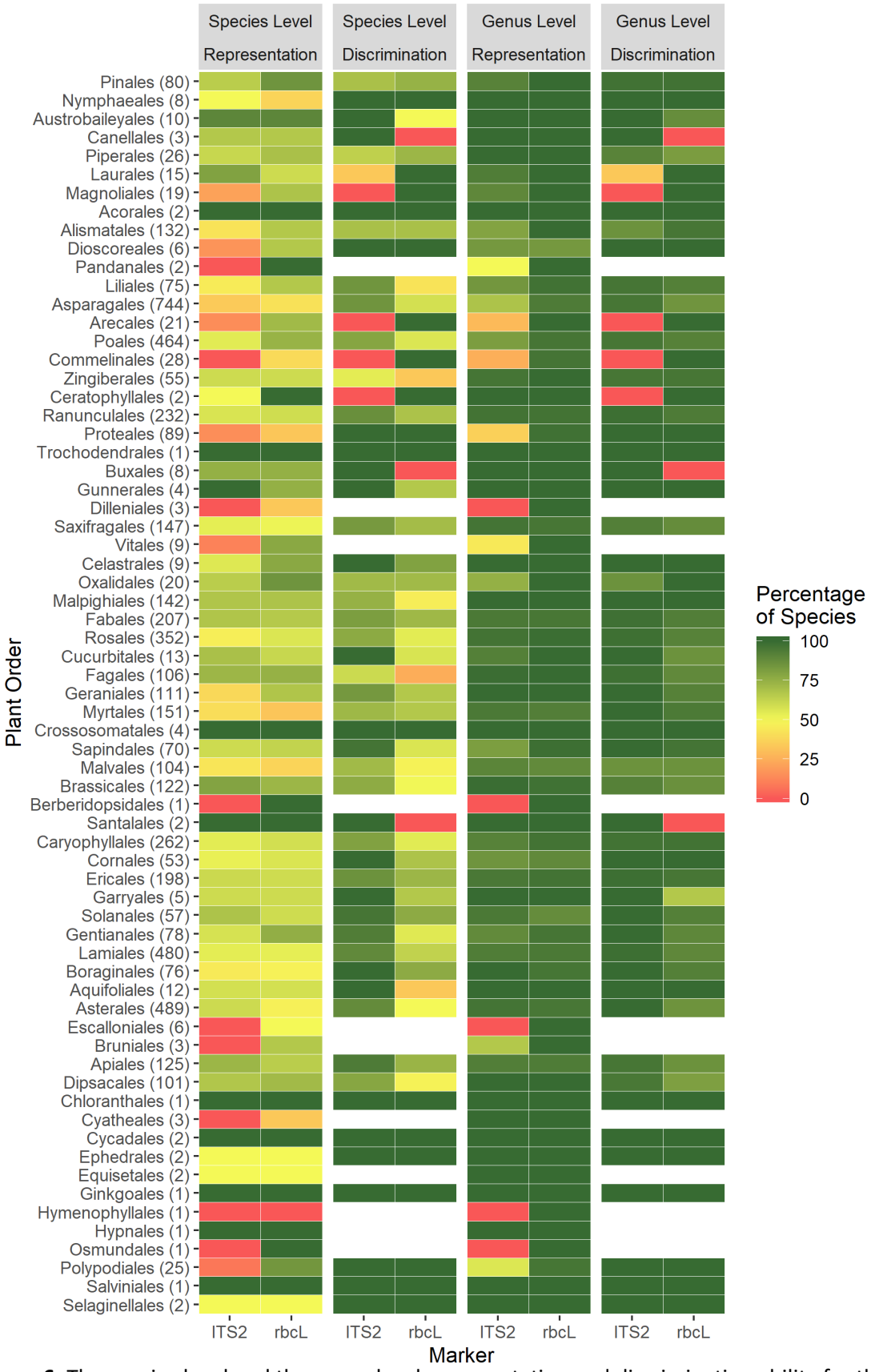


Figure 6: The species level and the genus level representation and discrimination ability for the UK native, naturalised and horticultural species reference library. Representation shows the percentage of species in each plant order which were available from GenBank. The discrimination level was assessed for the plant species which were represented by both *rbcL* and ITS2 markers more than once, showing the percentage of species in each plant order which were identified. Number of species represented by an order is shown in brackets.

2.5. Discussion

This DNA barcode reference library for the flowering plants of the UK provides an important resource for applications of species identification. Presented here is a comprehensive and high-quality DNA barcode reference library for the native UK flora, for three loci, complete with metadata and herbarium voucher images for each specimen. This is supplemented with a curated library of DNA barcodes downloaded from GenBank, which gives representation for the non-native plant species, which may be naturalised, alien or horticultural.

Freshly collected material performed better in terms of successful sequence recovery, compared to herbarium material, across all three markers. When looking at the success of herbarium material alone, the sequence recovery was strongly related to the age of the specimen with more recently collected specimens having increased success. Similar patterns of effect of specimen age were seen during a Canadian plant barcoding effort from herbarium samples (Kuzmina et al., 2017), however they found ITS2 was less affected by specimen age, which was hypothesised to be due to its shorter length. That pattern was not observed here, with ITS2 performing poorly with herbarium material compared to freshly collected material.

While using herbarium material can increase the lab and processing time involved in gaining a DNA barcode, this increase is mitigated by the relative effort involved in collecting, identifying and processing new plant specimens to create high-quality DNA barcodes. However, certain considerations are required when using herbarium material. As evidenced here, sample collection from herbaria should focus on younger specimens. In addition, quality control checks should monitor for potential contamination from the specimen sampling stage, including the presence of algae which is difficult to detect when sampling (de Vere et al., 2012). Herbaria around the world have been recognised as a potential source of efficiently capturing the associated taxonomic expertise housed within (Dormontt et al., 2018; Kuzmina et al., 2017; Xu et al., 2015), as well as answering a broad range of questions beyond core species identification and taxonomy (James et al., 2018). There is great potential for current and new herbarium specimens to incorporate data which would benefit these applications, with samples stored with future DNA analysis in

mind (Heberling and Isaac, 2017). This would involve the creation of a "secondary voucher", one intended for destructive analysis (Kageyama et al., 2014). For DNA analysis this would be leaf material stored in silica gel, associated with the primary traditional voucher specimen, but the creation of secondary specimens have scope for use for other types of analysis, e.g. chemical (Kageyama et al., 2014). In addition, museum collections could be targeted for other species groups, as shown with bee species in the UK (Creedy et al., 2019).

The relative levels of recovery and discrimination seen in *rbcL*, *matK* and ITS2 show the need for a balance between taxonomic universality and the level of species discrimination gained. For the UK flora, ITS2 performed poorly compared to *rbcL* and *matK* in its species recovery, suggesting that the primers are less universal than the other markers (Chen et al., 2010; China Plant BOL Group et al., 2011). In the wider database of UK natives and non-natives, ITS2 performed poorly in certain monocot orders. Similar results were seen with the Canadian flora, with some of the lower sequence recovery from ITS2 attributed to less conserved primer sites in certain monocot families, combined with intragenomic variation from multiple copies leading to successful amplification but unreadable chromatograms from Sanger sequencing (Kuzmina et al., 2017). Within our wider database, the fern order Polypodiales showed low representation and discrimination. In Kuzmina et al., (2017), they also found difficulty in the accurate return of fern species sequences for ITS2, with sequences annotated on BLAST actually originating from contaminants from fungi or angiosperms. Overall, all three of the markers performed well with the UK flora when looking at the ability to discriminate to at least genus level (98-100%), reflecting similar patterns with the Canadian flora (>90% across *rbcL*, *matK*, and ITS2) (Braukmann et al., 2017).

With DNA metabarcoding applications, the mixed source samples used will often provide low amounts of DNA as well as poorer quality template, such as from soil, honey (Hawkins et al., 2015), or faecal samples (Moorhouse-Gann et al., 2018). The longer length of *matK* combined with the number of primer combinations required to gain taxonomic coverage makes it generally unusable for metabarcoding, where the sequencing length is often limited due to the sequencing platform and the degraded quality of the DNA source. However, *matK* and *rbcL* have been shown to be useful in the creation of phylogenetic

trees for community analysis (Lim et al., 2014). The *matK* region, in comparison to *rbcL*, also evolves faster and so can offer increased resolution within phylogenies (Hilu et al., 2003).

Understanding what is potentially available in the study system also can allow for increased discrimination ability. When looking at the ability to discriminate within the Welsh flora, concentrating on a geographically restricted set of species improved the levels of discrimination (de Vere et al., 2012), which is helpful in study sites with well characterised species availability. The success of species discrimination can be associated with the level of species diversity or generic diversity in a study area, as found when looking at different biogeographic regions of the Canadian flora (Braukmann et al., 2017). More floristically diverse areas were found to improve species discrimination when compared to areas with a lower number of species but a higher than average number of congeners, such as in the Arctic communities (Braukmann et al., 2017). Intraspecific variation has also been found to increase with the geographical scale of sampling (Bergsten et al., 2012), meaning the global application of DNA barcoding is limited without geographical representation, and highlights the importance in associating location metadata with DNA barcode sequences.

By augmenting the native species records with sequence data from GenBank, we have gained additional species level representation in several areas. Firstly, for non-native plant species present in the UK, secondly for the species which were not successfully DNA barcoded, and thirdly additional representation for native species already present in the database. In order to maximise the potential genus level coverage, we also extracted records for genera without any prior representation in the database. This will have resulted in the presence of species which are not grown in the UK, in order to gain representation for the genus, which needs to be taken into consideration during application. For all applications of DNA barcoding, thorough validation of any identification of the unknown sequences being queried is required, which is supported by understanding what is available in the reference library in detail. In addition, by understanding the current coverage present within GenBank, species lists can be created for further DNA barcode creation from herbarium specimens and, for horticultural species, living collections within botanic gardens.

2.5.1. Conclusions

The DNA barcode reference library presented here represents a high-quality database that is publicly available and able to facilitate the wide-ranging applications which require plant identification. We have shown the effective use of herbarium collections in efficiently using taxonomic expertise to build a robust DNA barcode library. We also provide a method for creating restricted references libraries, suitable for use in the identification of DNA metabarcoding data. Understanding the representation and ability of DNA barcode libraries is vital to effectively utilise them in the identification of unknown sequences.

2.6. References

Arrizabalaga-Escudero, A., Merckx, T., García-Baquero, G., Wahlberg, N., Aizpurua, O., Garin, I., Goiti, U., Aihartza, J., 2019. Trait-based functional dietary analysis provides a better insight into the foraging ecology of bats. *Journal of Animal Ecology* 1365-2656.13055.

Bell, K.L., de Vere, N., Keller, A., Richardson, R.T., Gous, A., Burgess, K.S., Brosi, B.J., 2016. Pollen DNA barcoding: current applications and future prospects. *Genome* **59**, 629–640.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P., 2012. The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology* **61**, 851–869.

Braukmann, T.W.A., Kuzmina, M.L., Sills, J., Zakharov, E. V., Hebert, P.D.N., 2017. Testing the efficacy of DNA barcodes for identifying the vascular plants of Canada. *PLoS ONE* **12**, 1–19.

Brennan, G.L., Potter, C., de Vere, N., Griffith, G.W., Skjøth, C.A., Osborne, N.J., Wheeler, B.W., McInnes, R.N., Clewlow, Y., Barber, A., Hanlon, H.M., Hegarty, M., Jones, L., Kurganskiy, A., Rowney, F.M., Armitage, C., Adams-Groom, B., Ford, C.R., Petch, G.M., Creer, S., 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution* **3**.

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., Labra, M., 2015. A DNA barcoding approach to identify plant species in multiflower honey. *Food Chemistry* **170**, 308–315.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12794–7.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., Leon, C., 2010. Validation of the ITS2 region as a novel DNA

barcode for identifying medicinal plant species. *PLoS ONE* **5**, 1–8.

China Plant BOL Group, Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., Chen, Z.-D., Zhou, S.-L., Chen, S.-L., Yang, J.-B., Fu, C.-X., Zeng, C.-X., Yan, H.-F., Zhu, Y.-J., Sun, Y.-S., Chen, S.-Y., Zhao, L., Wang, K., Yang, T., Duan, G.-W., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences* **108**, 19641–19646.

Chiou, S.J., Yen, J.H., Fang, C.L., Chen, H.L., Lin, T.Y., 2007. Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. *Planta Medica* **73**, 1421–1426.

Creedy, T.J., Norman, H., Tang, C.Q., Qing Chin, K., Andujar, C., Arribas, P., O'Connor, R.S., Carvell, C., Notton, D.G., Vogler, A.P., 2019. A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. *Molecular Ecology Resources* DOI: 10.1111/1755-0998.13056.

De Boer, H.J., Ghorbani, A., Manzanilla, V., Raclariu, A.C., Kreziou, A., Ounjai, S., Osathanunkul, M., Gravendeel, B., 2017. DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade. *Proceedings of the Royal Society B: Biological Sciences* **284**.

de Vere, N., Jones, L.E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S., Ford, C.R., 2017. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports* **7**, 42838.

de Vere, N., Rich, T.C.G., Ford, C.R., Trinder, S.A., Long, C., Moore, C.W., Satterthwaite, D., Davies, H., Allainguillaume, J., Ronca, S., Tatarinova, T., Garbett, H., Walker, K., Wilkinson, M.J., 2012. DNA Barcoding the Native Flowering Plants and Conifers of Wales. *PLoS ONE* **7**, e37945.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant

communities. *Molecular Ecology* **26**, 5872–5895.

Dormontt, E.E., van Dijk, K., Bell, K.L., Biffin, E., Breed, M.F., Byrne, M., Caddy-Retalic, S., Encinas-Viso, F., Nevill, P.G., Shapcott, A., Young, J.M., Waycott, M., Lowe, A.J., 2018. Advancing DNA Barcoding and Metabarcoding Applications for Plants Requires Systematic Analysis of Herbarium Collections—An Australian Perspective. *Frontiers in Ecology and Evolution* **6**, 1–12.

Ellstrand, N.C., Whitkus, R., Rieseberg, L.H., 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5090–3.

Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S., Hänfling, B., 2018. Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution* **8**, 6330–6341.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., Baillie, L., Adams-Groom, B., 2015. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. *PLOS ONE* **10**.

Heberling, J.M., Isaac, B.L., 2017. Herbarium specimens as exaptations: New uses for old collections. *American Journal of Botany* **104**, 963–965.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270**, 313–321.

Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T.A.B., Rohwer, J.G., Campbell, C.S., Chatrou, L.W., 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* **90**, 1758–1776.

Hollingsworth, P.M., 2011. Refining the DNA barcode for land plants. *Proceedings of the*

National Academy of Sciences **108**, 19451–19452.

Hollingsworth, P.M., Li, D., van der Bank, M., Twyford, A.D., 2016. Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150338.

James, S.A., Soltis, P.S., Belbin, L., Chapman, A.D., Nelson, G., Paul, D.L., Collins, M., 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research: Global. *Applications in Plant Sciences* **6**, 1–8.

Kageyama, M., Monk, R.R., Bradley, R.D., Edson, G.F., Baker, R.J., 2014. The changing significance and definition of the biological voucher. In: Museum Studies: Perspectives and Innovations. pp. 257–264.

Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W., Pringle, R.M., 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences* **112**, 8019–8024.

Kress, W.J., Erickson, D.L., Swenson, N.G., Thompson, J., Uriarte, M., Zimmerman, J.K., 2010. Advances in the Use of DNA Barcodes to Build a Community Phylogeny for Tropical Trees in a Puerto Rican Forest Dynamics Plot. *PLoS ONE* **5**, e15409.

Kuzmina, M.L., Braukmann, T.W.A., Fazekas, A.J., Graham, S.W., Dewaard, S.L., Rodrigues, A., Bennett, B.A., Dickinson, T.A., Saarela, J.M., Catling, P.M., Newmaster, S.G., Percy, D.M., Fenneman, E., Lauron-Moreau, A., Ford, B., Gillespie, L., Subramanyam, R., Whitton, J., Jennings, L., Metsger, D., Warne, C.P., Brown, A., Sears, E., Dewaard, J.R., Zakharov, E. V., Hebert, P.D.N., 2017. Using Herbarium-Derived DNAs to Assemble a Large-Scale DNA Barcode Library for the Vascular Plants of Canada. *Applications in Plant Sciences* **5**, 1700079.

Landi, M., Dimech, M., Arculeo, M., Biondo, G., Martins, R., Carneiro, M., Carvalho, G.R., Brutto, S. Lo, Costa, F.O., 2014. DNA barcoding for species assignment: The case of Mediterranean marine fishes. *PLoS ONE* **9**.

Li, M., Cao, H., But, P.P.H., Shaw, P.C., 2011. Identification of herbal medicinal materials using DNA barcodes. *Journal of Systematics and Evolution* **49**, 271–283.

Lim, J., Crawley, M.J., de Vere, N., Rich, T., Savolainen, V., 2014. A phylogenetic analysis of the British flora sheds light on the evolutionary and ecological factors driving plant invasions. *Ecology and Evolution* **4**.

Lucas, A., Bodger, O., Brosi, B.J., Ford, C.R., Forman, D.W., Greig, C., Hegarty, M., Jones, L., Neyland, P.J., de Vere, N., 2018. Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding. *Scientific Reports* **8**.

Moorhouse-Gann, R.J., Dunn, J.C., De Vere, N., Goder, M., Cole, N., Hipperson, H., Symondson, W.O.C., 2018. New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. *Scientific Reports* **8**, 1–15.

Preston, C.D., Pearman, D.A., Dines, T.D., 2002. *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands*. Oxford University Press, Oxford.

R Development Core Team, R., 2011. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

Stace, C., 2010. New Flora of the British Isles, Third. ed. Cambridge University Press.

Stoeckle, M.Y., Gamble, C.C., Kirpekar, R., Young, G., Ahmed, S., Little, D.P., 2011. Commercial teas highlight plant DNA barcode identification successes and obstacles. *Scientific Reports* **1**, 1–7.

Thomsen, P.F., Sigsgaard, E.E., 2019. Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution* **9**, 1665–1679.

Walker, K.J., Preston, C.D., 2006. Ecological predictors of extinction risk in the flora of lowland England, UK. *Biodiversity and Conservation* **15**, 1913–1942.

Xu, C., Dong, W., Shi, S., Cheng, T., Li, C., Liu, Y., Wu, P., Wu, H., Gao, P., Zhou, S., 2015. Accelerating plant DNA barcode reference library construction using herbarium specimens: Improved experimental techniques. *Molecular Ecology Resources* **15**, 1366–1374.

Yan, D., Mills, J.G., Gellie, N.J.C., Bissett, A., Lowe, A.J., Breed, M.F., 2018. High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biological Conservation* **217**, 113–120.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., Chen, S., 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* **5**.

CHAPTER 3

Temporal and spatial patterns of honeybee foraging in a diverse floral landscape

3.1. Abstract

Honeybees contribute both directly and indirectly to humans; through honey, wax and propolis, and as the pollinator of both wild and crop plants. The increased rate of honeybee colony loss has caused worldwide concern, caused by the interacting effects of habitat loss and fragmentation, agrochemicals, pests and diseases, and climate change. Understanding which plants honeybees use throughout the season is vital to being able to provide recommendations for suitable forage. Here, a Botanic Garden and an adjacent National Nature Reserve, which includes organic farmland, have been used to assess foraging by honeybees in order to build a temporal and spatial picture of plant use. Each month, from April to September, all plants in flower throughout the study site were recorded and honey was sampled from six hives. Of the six hives, three were set in an apiary within the Botanic Garden with close range access to horticultural habitat, while the other three hives were placed 1 km away within the Nature Reserve. We used DNA metabarcoding to survey which plants honeybees use by assessing the pollen biodiversity within honey. Only a small proportion of available flowering plants were visited by honeybees within a diverse landscape. In total, 136 plant taxa were found in the honey from April to September with *rbcL,* but only 16 of these plants represented more than 10% of the DNA sequences returned for each month. The greatest proportion of DNA comes from native plants, including *Rubus* spp., *Trifolium repens*, *Hedera helix*, and *Taraxacum officinale*, supplemented with lower levels of horticultural species.

3.2. Introduction

Pollination is a key ecosystem service and required for a diverse food supply, with 75% of globally important food crops pollinated by insects (Klein et al., 2007). However, pollinators are under stress from multiple interacting factors; the spread of pests and diseases, climate change, and anthropogenic pressures including agricultural intensification leading to changes in land-use and the increased application of pesticides (Goulson et al., 2015; Potts et al., 2010). Nutritional stress from the loss and fragmentation of suitable foraging habitat has been suggested as one of the major drivers of population decrease, meaning an understanding of the floral preferences in foraging is essential to mitigate habitat declines and support pollinator populations. Due to their ease of management, compared with other pollinators, honeybees are important to crop pollination as well as providing honey and wax products (Potts et al., 2016). There is concern, therefore, over honeybee colony loss due to increasing pressure from the reduction in the quantity and diversity of suitable foraging habitat coupled with exposure to agrochemicals and spreading pest and diseases (Goulson et al., 2015; Naug, 2009).

Foraging is energetically demanding. Honeybees forage for nectar, pollen, and water, with nectar being the main energy source, providing sugars which are processed into honey for long-term storage. It has been estimated that a colony can collect 120 kg of nectar, 20 kg of pollen, and 20 L of water per year, with most nectar and pollen consumed during the spring and summer months (Seeley, 1995). Pollen is the main source of protein while also supplying lipids, sterols, vitamins and minerals, and is vital during brood rearing for healthy growth and development of the larvae (Brodschneider and Crailsheim, 2010; Pernal and Currie, 2001). Nectar from plants can vary highly in the concentration, the proportion of sugar types, and the volume of production (Nicolson, 2007), while amino acid composition and overall protein content can differ between pollen. The diversity and quantity of pollen used by bees has been shown to affect development, longevity and behaviour of the hive (Maurizio and Hodges, 1950; Scofield and Mattila, 2015; Standifer, 1967).

The foraging choices bees are making have been reported to be influenced by the surrounding vegetation, the availability of the resource (Beekman and Ratnieks, 2000),

flowering phenology, interspecific competition (Balfour et al., 2013), and the hive requirements. The importance of these influences can also differ between nectar and pollen foraging. Nectar foraging was found not to vary with changes in the levels of honey storage within the colony, with no increase in nectar intake rates when honey stores were decreased (Fewell and Winston, 1996). In contrast, pollen intake rates increased under low pollen storage conditions via increased foraging activity and increased pollen load size (Fewell and Winston, 1992). Foraging distances can change with the surrounding landscape, with longer distances of 10 km found in environments where patch size and quality are variable (Beekman and Ratnieks, 2000; Steffan-Dewenter and Kuhn, 2003). The foraging distance also has been found to vary during the season, with shorter foraging distances in spring compared with summer (Couvillon et al., 2015, 2014). Honeybees have been shown to generally prefer a high concentration of sugar in nectar (Waller, 1972), but this does not dictate what they will collect in the field, where foraging is also dependent on what is available in the locality (Wright et al. 2018).

In order to support pollinator populations, sowing flower-rich field margins as well as laying and maintaining hedgerows have been encouraged in order to increase the abundance and diversity of floral resources throughout the year in agricultural landscapes (Goulson et al., 2015). Additionally, gardens have been recognised as an important resource for pollinators, especially in urban landscapes (Baldock et al., 2019). As well as habitat diversity, plant availability is also affected by seasonal change. Certain crop species can provide mass-flowering over only a restricted period which makes semi-natural habitats vital during times of low floral abundance in the wider landscape (Requier et al., 2015).

Honeybees have been described as super-generalists when foraging but have been shown to still be selective in the pollen and nectar they choose. de Vere et al., (2017) showed that in spring, honeybees use a small proportion of what is available to them in a landscape, collecting from a small number of nectar and pollen rich sources to provide most of their nutritional needs. Knowing the plants honeybees are actually using throughout the season is vital to providing appropriate floral resources to support healthy honeybee colonies.

Melissopalynology, the traditional method of pollen analysis of honey, uses morphological identification of the pollen grains by microscopy (Von Der Ohe et al., 2004), and has been used to examine the botanical and geographical source of honey (Anklam, 1998), as well as honeybee foraging (Coffey and Breen, 1997). However, it requires a high level of expertise to identify the different pollen types and can be limited in the level of identification within certain plant groups due to an overlap in morphology (Galimberti et al., 2014). Alternatively, DNA metabarcoding, using high-throughput sequencing with a standardised DNA region, can be an efficient method to identify multiple species within a community, reducing the need for time-consuming specialised identification (Deiner et al., 2017). DNA metabarcoding is being used to answer questions in an increasing number of ecological applications, with plants identified from the pollen contained within honey (de Vere et al., 2017; Hawkins et al., 2015), from pollen loads collected from traps on the entrance to the hives (Danner et al., 2017; Richardson et al., 2015), as well as from the bodies of different pollinators (Lucas et al., 2018; Potter et al., 2019). As with melissopalynology, the capability of DNA metabarcoding in identifying species is only possible with a reference database of the potential species available in the study system, with comprehensive coverage of taxa in the reference database being crucial to the quality of the results.

Here we investigate honeybee foraging throughout the season, using hives set within a complex landscape, including a Botanic Garden and a National Nature Reserve, managed as an organic farmland. The National Botanic Garden of Wales contains over 5,500 plant taxa within its horticultural plantings, inventoried using a botanical garden collection management database (IRIS BG). All of the UK flowering plant species have been DNA barcoded and this resource was augmented with curated reference sequences from GenBank to create a comprehensive reference database of available plant species for the DNA metabarcoding work (Chapter 2).

We used the DNA barcode region *rbcL* to characterise the plant composition from honey on a monthly basis (April to September), while also surveying the habitat around the apiaries to identify which plants are in flower and available to the honeybees. Honey was sampled from six hives. Three of the hives were set in an apiary within the Botanic Garden and with close range access to horticultural habitat, while the other three hives were

placed 1 km away within the Nature Reserve, still within foraging distance of the Botanic Garden, with access to grassland, broadleaved woodland, hedgerows and linear features. We present a honey DNA metabarcoding approach to monitor honeybee foraging in a complex landscape with high floristic diversity, investigating how season, diversity of the landscape, and apiary location influence differences in the floral composition of the honey.

3.2.1. Aims and objectives

- Which plants are honeybees choosing throughout the season (April to September) when offered a diverse floral resource?
- Are there differences in foraging between hives placed in close proximity to horticultural plants, compared with hives situated further away, but within foraging distance?
- How does the abundance of a plant within the landscape relate to the abundance found within the honey?
- Is there any pattern to the characteristics of the forage plants found throughout the season, in terms of native status, growth form and associated habitat?

3.3. Methods and Analysis

3.3.1. Floral surveys

The National Botanic Garden of Wales survey area was divided into 279 survey zones, circling two apiaries, one contained within the Botanic Garden and one in the Nature Reserve (Figure 7). Each zone was classified into four main habitat types, broadleaved woodland, grassland, hedgerow and linear features, and horticultural. Horticultural habitat represents areas of the Botanic Garden which are planted, including native and non-native plant species. Grassland habitats are found within both the Botanic Garden and the Nature Reserve and include both semi-improved grassland and species rich meadows, either managed by grazing or cutting. Hedgerow and linear features include hedgerow habitats and scrub field margins. For the horticultural habitat, the survey zones tended to represent distinct flowerbeds, while for the non-planted habitats the zones were split into the main habitat type (Figure 7). Each of the zones were mapped using QGIS v. 3.6.1 and R v. 3.5.1.

Floral surveys were carried out monthly from May 2016 to September 2016 and again from April 2017 to September 2017. Surveys took place over seven to fourteen days and a list of the plant species in flower was recorded for each zone within the survey area. During the 2017 survey, to give a measure of abundance within the site, the percentage cover of each species within each zone was recorded and scaled to the total available area of the zone.

3.3.2. Location of apiaries

During the same time period as the floral surveys, honey was sampled from a total of six *Apis mellifera* hives split between two apiaries with differing close-range forage. Three of the sampled hives were placed in the Botanic Garden with close range access to horticultural habitat, including a systematics garden with plant families arranged according to the APG III classification system (Chase et al., 2009). The remaining three sampled hives were placed one kilometre away from the horticultural apiary and habitat, within a National Nature Reserve, which is managed as an organic farm (Figure 7). The surrounding landscape of the surveyed habitat is an agricultural landscape, with improved

grassland used for grazing sheep and cattle. The six honeybee colonies were managed in the same way, with inspections each week to find the queen or evidence of egg laying, perform checks for pests or diseases, and prevent swarming. Standard national hives were used.

3.3.3. Honey sampling and DNA extraction

Using a sterile 50 ml tube to crush a section of comb and release the honey, approximately 30 ml of freshly capped honey was collected from the comb within each hive. Any wax was removed using sterile forceps and 10 g of honey was weighed out for DNA extraction using a modified version of the DNeasy Plant Mini extraction kit (Qiagen). Firstly, the 10 g of honey was made up to 30 ml with molecular biology grade water and incubated in a water bath at 65 °C for 30 minutes. Samples were then centrifuged (Sorvall RC-5B) for 30 minutes at 15,000 rpm, the supernatant was discarded, and the pellet was resuspended in 400 µL of buffer made from a mix of 400 µL AP1 from the DNeasy Plant Mini Kit (Qiagen), 80 µL proteinase K (1 mg/ml) (Sigma) and 1 µL RNase A (Qiagen). This was incubated again for 60 minutes at 65 °C in a water bath and then disrupted using a TissueLyser II (Qiagen) for 4 minutes at 30 Hz with 3 mm tungsten carbide beads. The remaining steps were carried out according to the manufacturer's protocol, excluding the use of the QIAshredder and the second wash stage. The OneStep PCR Inhibitor Removal Kit (Zymo Research) was used to purify the DNA extract and then was diluted 1 in 10.

3.3.4. PCR and library preparation

Illumina MiSeq paired end indexed amplicon libraries were created using a two-step PCR protocol, as in de Vere *et al.*, (2017). Two libraries were prepared using *rbcL* and ITS2. Samples were first amplified using the template specific primers with 5' complementary overhangs to the index primers. The first PCR had a total volume of 20 µl: 2 µl template DNA, 10 µl of 2x Phusion Hot Start II High-Fidelity Mastermix (New England Biolabs UK), 0.4 µl (0.25 µM) forward and reverse primers, and 7.2 µl of PCR grade water. Samples from this first PCR were assessed by gel electrophoresis on 1% agarose. The first PCR was completed three times and pooled before entering a bead clean up.

For the *rbcL* barcode region, the samples were first amplified using the template specific primers *rbcL*af and *rbcL*r506 with 5' overhangs complementary to Nextera XT index

primers. Thermal cycling conditions for the first PCR for *rbcL* were: 98 °C for 3 min, 95 °C for 2 minutes; 95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 40 seconds (40 cycles); 72 °C for 5 minutes, 30 °C for 10 seconds.

Forward universal tail and *rbcL*af:
[**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**][ATGTCACCACAAACAGAGACTAAAGC]

Reverse universal tail and *rbcL*r506:
[**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**][AGGGGACGACCATACTTGTTCA]

For the ITS2 barcode region, initial amplification used the template specific primers ITS2F and ITS3R, with universal tails designed to attach custom indices in the second round PCR. Thermal cycling conditions for the first ITS2 PCR were: 98 °C for 3 min, 94 °C for 5 minutes; 94 °C for 30 seconds, 56 °C for 30 seconds, 72 °C for 40 seconds (40 cycles); 72 °C for 10 minutes, 30 °C for 1 minute.

Forward universal tail, 6N sequence and ITS2F:
[**ACACTCTTTCCCTACACGACGCTCTTCCGATCT**]NNNNNN[ATGCGATACTTGGTGTGAAT]

Reverse universal tail and ITS3R:
[**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**][GACGCTTCTCCAGACTACAAT]

The pooled products from the first PCR were then purified following Ilumina's 16S Metagenomic Sequencing Library Preparation protocol using Agencourt AMPure XP beads (Beckman Coulter) with a 1:0.6 ratio of product to beads.

For *rbcL*, the purified PCR product from round one was followed by a second round of amplification to anneal sample specific Illumina Nextera indices. This index PCR stage used a total volume of 25 µl (12.5 µl of 2x Phusion Hot Start II High-Fidelity Mastermix, 2.5 µl of Nextera XT i7 Index Primer, 2.5 µl of Nextera XT i5 Index Primer, 5 µl of PCR grade water, and 2.5 µl of purified first-round PCR product).

For ITS2, the second-round amplification annealed custom unique and identical i5 and i7 indices to each sample (Ultramer, IDT). This index PCR stage used a final volume of 25 µl reaction (12.5 µl of 2x Phusion Hot Start II High-Fidelity Mastermix, 1 µl of i7 Index Primer

and i5 Index Primer, 6.5 µl of PCR grade water, and 5 µl of purified first-round PCR product).

Thermal cycling conditions for both *rbcL* and ITS2 were: 98 °C for 3 min; 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s (8 cycles); 72 °C for 5 min, 4 °C for 10 min. To check the success of the index PCR a 1% gel was run, comparing with the cleaned-up product from the first PCR. The index PCR product was then purified following the PCR clean-up 2 section of the Illumina protocol, using a 1:0.8 ratio of product to AMPure XP beads.

The purified products of the index PCR were quantified using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and pooled at equal concentrations, producing the final library. Negative controls using PCR grade water were amplified and sequenced alongside honey samples using the Illumina MiSeq. Sequence data will be available at the NCBI Sequence Read Archive (SRA).

3.3.5. Bioinformatic analysis

Sequence data was processed using a modified bioinformatic analysis pipeline first developed in de Vere *et al.*, (2017) (https://github.com/colford/nbgw-plant-illumina-pipeline). Raw reads were trimmed to remove low quality regions (Trimmomatic v. 0.33), paired, and then merged (FLASH v. 1.2.11), with merged reads shorter than 450 bp discarded. Identical reads were dereplicated within samples and then clustered simultaneously at 100% identity across all samples (vsearch v. 2.3.2), with singletons (sequence reads that occurred only once across all the samples) then discarded.

A custom reference database was created for sequence identification (Chapter 2). The species list for the Botanic Garden was generated using the list of native species of the UK (Stace, 2010), naturalised and alien species (505 species, Preston et al., 2002), planting records from the IRIS BG horticultural database at the National Botanic Garden of Wales, and survey data records from the 2016 and 2017 floral surveys. This represented 5,586 species. All available *rbcL* and ITS2 plant records were downloaded from NCBI GenBank and the total species list was used to extract relevant records using the script *creatingselectedfastadatabase.py* (available at https://github.com/colford/nbgw-plant-illumina-pipeline). For plants on the species list not represented at species level within

GenBank, a second extraction was completed for records at the genus level. In the reference database created, species level coverage for *rbcL* was 57%, and coverage at genus level was 96%. For ITS2 the species level coverage was 52% with 84% coverage at genus level.

The sequence data from the honey samples were compared against the reference database using blastn, using the script vsearch-pipe.py. The top twenty BLAST hits were then summarised using the script vsearch_blast_summary.py. If the top bit scores of a sequence matched to a single species, then the sequence was identified to that species. If the top bit scores matched to different species with the same genus, then the result was attributed to the genus level. If the top bit score belonged to multiple genera within the same family, then a family level designation was made. Sequences that returned families from different clades were considered to be chimeric and excluded. These computed identifications were then checked manually for botanical veracity, in terms of the phenology of the plants and their presence within the study site.

The number and overlap of plant taxa found by the *rbcL* and ITS2 barcode regions was compared. For the plant taxa which were found by both markers, the proportion of read abundance for *rbcL* and ITS2 was compared. Spearman's rank correlation was used to test the correlation. Rarefaction curves were generated using the 'rarefy' function from the R package *vegan* to assess whether the sequencing depth was adequate for both markers.

3.3.6. Assigning plant taxa information

Habitat and native status were assigned for all plant taxa identified at genus and species level from the DNA metabarcoding. Taxa identified at family level were not categorised. The habitat type assigned to plants matches the categories of habitat found in the survey site. For plant taxa which were not primarily represented by one habitat type, overlapping categories were created, e.g. grassland-horticultural. The habitat type categories are broadleaved woodland, grassland, hedgerow and linear features, and horticultural, coupled with woodland-horticultural, hedgerow-horticultural and grassland-horticultural. Native status was similarly categorised with native, naturalised, horticultural and native-horticultural. Native represents plants native to the UK as defined by Stace (2010),

including archaeophytes. Native-horticultural represents those taxa which were available in both a native form and a horticultural form in the survey area in equal amounts.

3.3.7. Statistical analysis

A model-based approach for the multivariate metabarcoding abundance data was used (Brennan et al., 2019). A generalised linear model was run using the manyglm function in the R package *mvabund* (Wang et al., 2012), to examine the abundance and composition of plant taxa within the honey through the season and by location. To standardise the read count across samples, the proportion of sequences returned was calculated for each sample. The proportion of sequences was the response variable and the effects of time and location of the apiaries were the explanatory variables. The data best fit a negative binomial distribution, with the large number of zero values for taxa across the samples resulting in a strong-mean variance relationship. To use a model with a negative binomial distribution and retain the sequence proportion information, the proportion of sequences was multiplied by a factor of 1,000, and the values were changed to integers as in Brennan et al., (2019). The 'manyglm' function allows a generalized linear model to be fitted to each plant taxa individually and includes an assumption for the mean-variance relationship. Model selection was based on the Akaike information criterion score and tested using the 'dropterm' function in R. Samples collected in April 2017 were excluded from the model analysis to allow comparison between the sampling years.

Non-metric multidimensional scaling (NMDS) ordination was used to visualise seasonal changes in the composition of the honey over time, based on the proportion of reads returned for each plant taxa. Ordinations were carried out using the metaMDS function in the *vegan* package in R (Dixon, 2003), using Bray-Curtis dissimilarity indices.

The diversity of the surrounding landscape was calculated from the survey data based on the available plant genera in each habitat through the season using Shannon's diversity index. Spearman's rank correlations with Bonferroni corrections for multiple testing were run to investigate any correlation between metabarcoding abundance of the plant taxa and the available abundance within the landscape, throughout the season. All statistical analyses were performed using R 3.5.1 (R Development Core Team, 2011).

3.4. Results

3.4.1. Characterising available plant species and habitat

The surveyed area around the two apiaries covered 62 ha and was classified into four main habitat types, broadleaved woodland (13%), grassland (70%), hedgerow and linear features (8%) and planted horticultural areas (9%) (Figure 7). The horticultural habitat showed a higher level of generic richness and diversity (assessed using Shannon Diversity Index) compared to the native habitat, throughout the surveyed season (Figure 8). The total number of plant records generated was 19,343. This represented 1,895 unique species records and 762 unique genera.

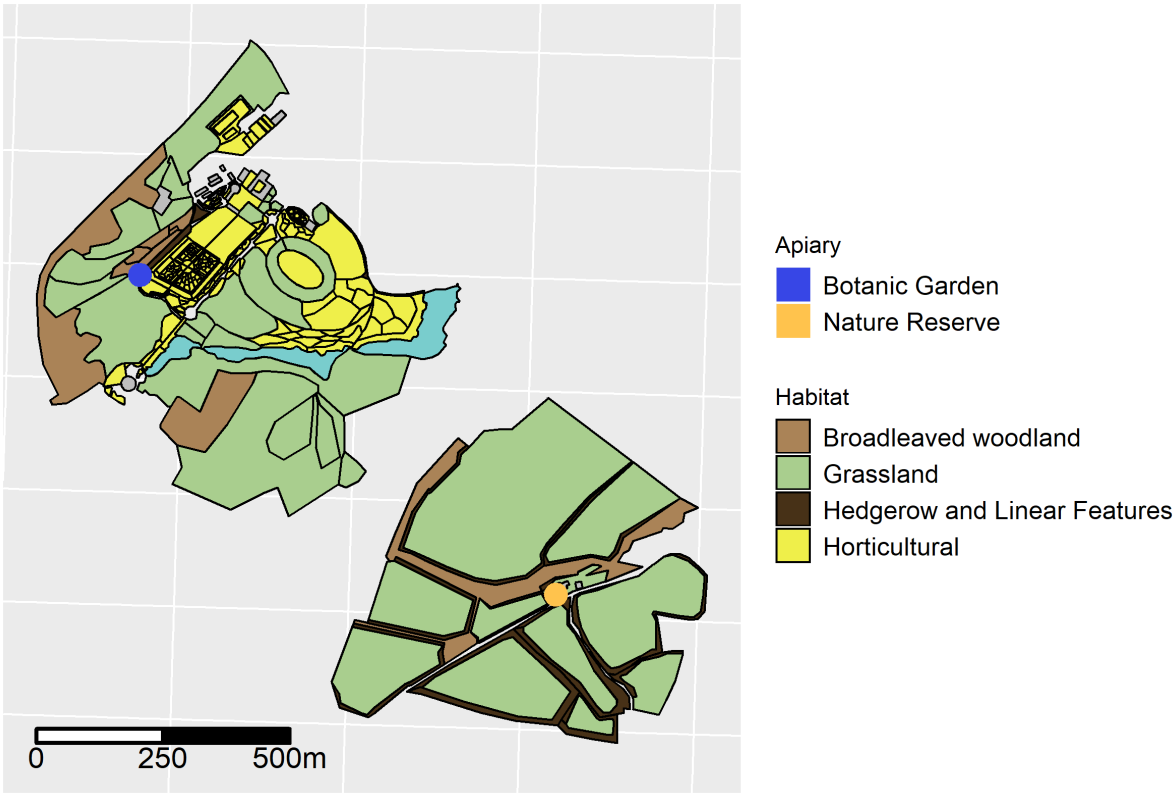


Figure 7: Illustrating the different close-range habitat types within the survey area and the two apiaries set in the study site. Grassland consisting of predominantly semi-improved natural grassland is the greatest area covered. Maps were created in QGIS v 3.6.1 and R v 3.5.1 from OS data © Crown Copyright (2018) licensed under the Open Government Licence.

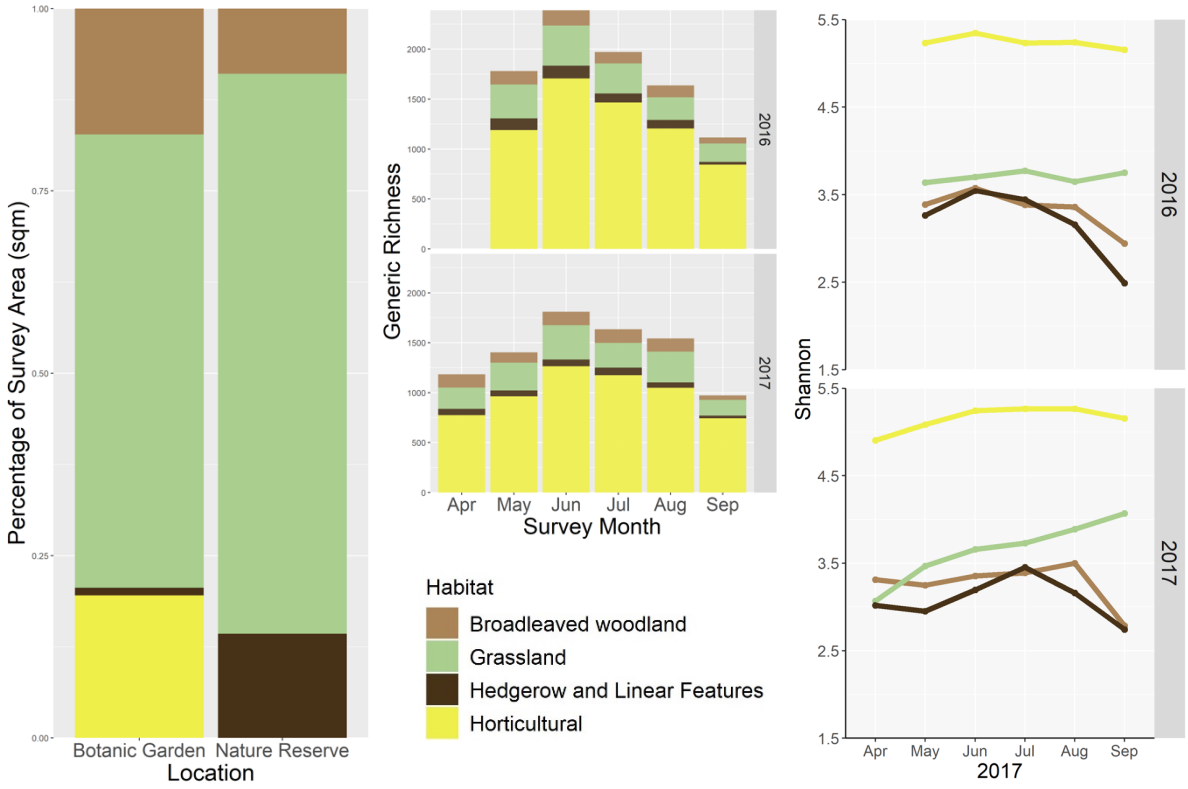


Figure 8. Characterising the composition of the close-range foraging area (500 m) for the two apiaries over time. The horticultural habitat found in the Botanic Garden is richer and more diverse in terms of available flowering plant genera but represents less than 10% of the total surveyed areas.

3.4.2. Sequencing

The *rbcL* sequencing run yielded a total of 11,916,038 returned read pairs from 66 samples. After the quality control, trim, pair and merge a total of 8,214,091 reads remained. The sequences were then clustered at 100% identity within and across all samples. Singletons were discarded, leaving 6,952,520 remaining sequences which entered the identification pipeline. During identification, 263,941 low quality sequences were discarded, leaving 6,688,579 sequences which entered analysis. Of the 66 honey samples, three returned less than 100 sequences and were excluded from further analysis. The mean sequence number returned for each sample was 106,525 (SD = 42,025) and ranged from 33,971 to 217,408.

For ITS2, 589,872 high quality sequences progressed through identification. The number of sequences returned for a sample ranged from 1,181 to 35,282, with one sample

returning zero sequences after quality control. Rarefaction curves were generated to examine the sequencing depth for *rbcL* and ITS2, with plateaus being observed for both markers (**Error! Reference source not found.**).

In total *rbcL* identified 136 plant taxa, while ITS2 found 37. Of the 37 taxa identified by ITS2, 34 were also found by *rbcL*. As the majority of the taxa identified by ITS2 were represented in the *rbcL* results, the ITS2 results were not included in the main analysis. A correlation was found between the proportion of reads for the shared taxa between *rbcL* and ITS2, when including the most abundantly returned taxon *Rubus* spp. (Spearman's rank correlation, $r_s = 0.674$, $p = 0.0008$) and without *Rubus* spp. ($r_s = 0.625$, $p = 0.003$) (Figure 9).

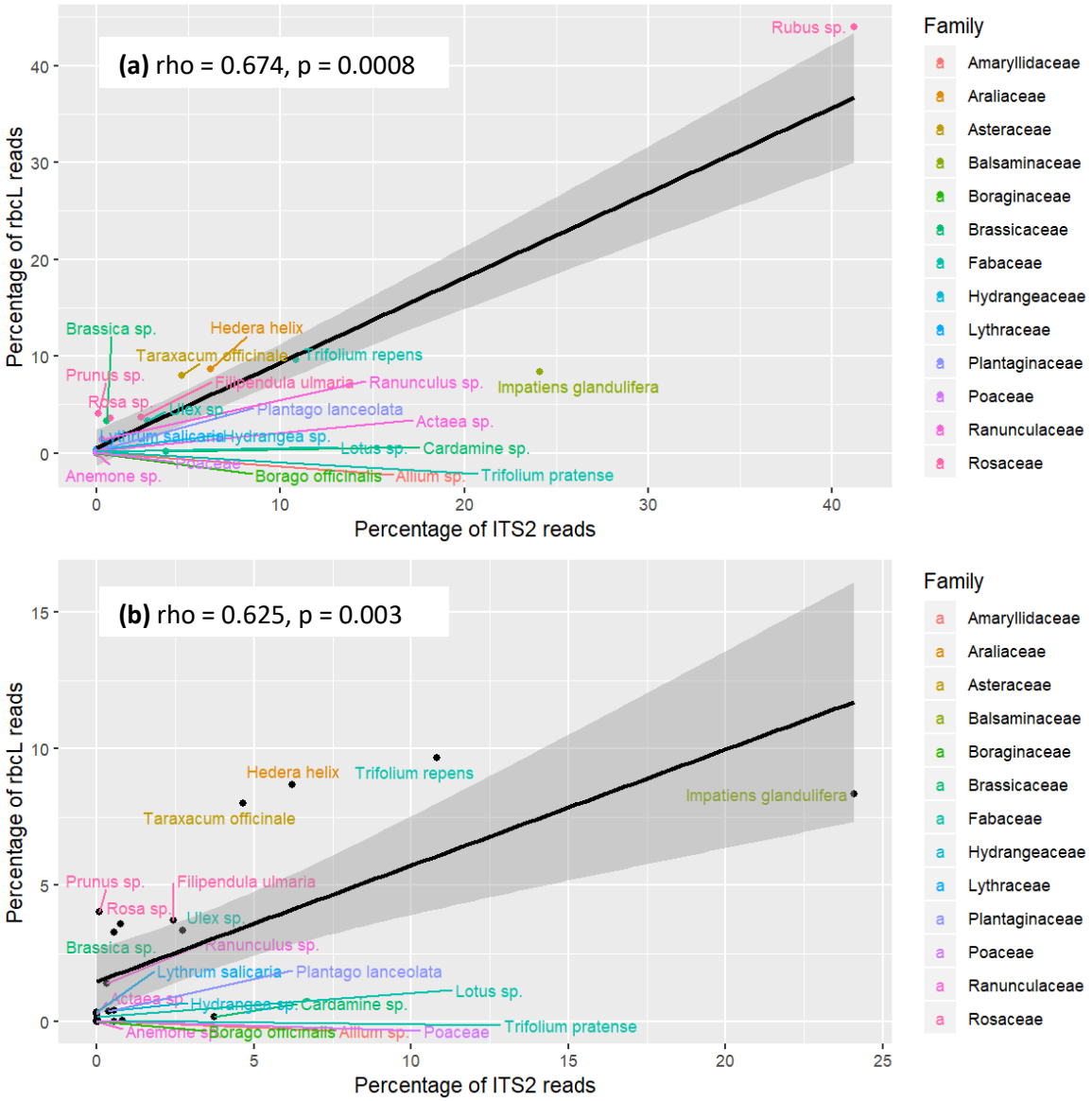


Figure 9: Examining the relationship between the proportion of reads returned for *rbcL* and ITS2, for the taxa that were found using both markers. A significant relationship was found tested using Spearman's rank correlation both with *Rubus* spp. present **(a),** and when excluded **(b)**.

3.4.3. Plant taxa found in the honey

In total, using the *rbcL* plant DNA barcode region, 136 plant taxa were identified across all samples, from May to September for 2016 and April to September for 2017 (Table 4). Of the sequencing reads returned, 26% were assigned to species, 60% were to genus level, 8% were matched to a tribe and 6% to a family. Only 17 taxa were returned at over 1% of all sequences, representing 88% of the total sequences. The plant taxa found were labelled with four categories of abundance, those representing over 10% of sequences were designated major plant taxa, between 1% and 10% secondary taxa, between 0.01% and 1% minor taxa, and below 0.01% were occasional (Table 4).

Table 4: List of plant taxa found in the honey using the *rbcL* DNA barcode marker. Habitat was classified as broadleaved woodland (BW), grassland (G), hedgerow and linear features (H), horticultural (Ht), woodland-horticultural (BWHt), grassland-horticultural (GHt), and hedgerow-horticultural (HHt). Status was grouped into native (Nv), naturalised (Nr), horticultural (Ht) and native-horticultural (NvHt). Plant taxa were classified into plant growth type, herb (H), shrub (S), or tree (T). Taxa returned at family level were not categorised (NC). Looking on a monthly basis, over 10% of sequences are categorised as major plant taxa, between 1% and 10% are secondary taxa, between 0.01% and 1% are minor taxa, and below 0.01% are occasional.

					2016					2017						
Family	Binomial	Habitat	Status	Type	May	June	July	Aug	Sep	Apr	May	Jun	Jul	Aug	Sep	Total (%)
Rosaceae	*Rubus* spp.	H	Nv	S	0.013	60.180	81.607	47.971	7.874	0.001	0.004	30.786	38.102	30.361	17.618	30.388
Fabaceae	*Trifolium repens*	G	Nv	H	0.001	3.460	6.690	1.620	1.301		0.000	13.717	23.164	13.105	0.871	6.669
Rosaceae	Maleae (*Crataegus/Malus/ Cotoneaster*)	HHt	NvHt	T	54.289	0.842	0.001	0.001	0.000	0.841	21.080	2.234	1.012	0.075	0.002	6.154
Araliaceae	*Hedera helix*	BW	Nv	S	0.004	0.006	0.002	0.024	44.305	0.372	0.006	0.005	0.002	0.021	21.920	5.992
Balsaminaceae	*Impatiens glandulifera*	H	Nr	H	0.026	1.435		26.725	21.804	1.436			0.000	4.205	7.808	5.768
Asteraceae	*Taraxacum officinale*	G	Nv	H	8.970	0.006	0.002	0.001	0.004	10.549	29.970	4.780	0.003	1.562	20.096	5.528
Asteraceae	*Cirsium/Hypochaeris* spp.	G	Nv	H	0.004	2.580	1.705	0.000	0.001	0.001	0.000	25.021	16.238	5.261		5.093
Salicaceae	Salicaceae	NC	NC	NC	0.003	0.489	1.974	0.001	11.087	0.002	0.012	0.117	0.002	16.934	5.991	3.385
Rosaceae	*Prunus* spp.	Ht	Ht	T	5.090	0.018	0.001	0.002	0.003	20.512	1.999	0.484	0.004	0.002	0.002	2.791
Rosaceae	*Filipendula ulmaria*	G	Nv	H	0.024	0.173	3.875	9.150	4.938	0.000	0.000	0.000	0.002	8.191	0.001	2.558
Rosaceae	*Rosa* spp.	Ht	Ht	S	0.001	18.171	0.095	0.032	0.007	0.000	0.000	12.781	0.012	0.020	0.003	2.471
Sapindaceae	*Acer* spp.	BWHt	NvHt	T	11.833	0.005		0.000	0.001	1.278	9.257	0.562	0.001	6.766		2.345
Salicaceae	*Salix* spp.	BW	Nv	T	0.229	0.001	0.001	0.001	0.002	21.174	0.135	0.273	0.003	0.011	0.002	2.336
Fabaceae	*Ulex* spp.	G	Nv	S	0.832	0.007	0.001	0.001	0.002	16.475	7.510	0.003	0.003	0.001	0.001	2.307
Brassicaceae	*Brassica* spp.	Ht	Ht	H	2.753	0.000	0.000	0.001	0.000	18.766	0.187	0.001	0.001	0.001	0.000	2.258
Caprifoliaceae	*Weigela* spp.	Ht	Ht	S							16.309					1.002
Rosaceae	Maleae (Jul17)	GHt	NvHt	NC	0.045	0.000		0.000	0.000	0.001	0.019	0.004	8.051	0.000		1.001
Theaceae	*Camellia* spp.	Ht	Ht	S	0.080	0.002		0.005	0.064	0.001	0.001	0.000	0.003	0.000	15.412	0.978
Euphorbiaceae	Euphorbiaceae *(inc. Mercurialis)*	NC	NC	NC	0.000	0.000	0.000	0.000	0.000	1.100	4.961	0.000	0.000	5.180	1.063	0.976
Ranunculaceae	*Ranunculus* spp.	G	Nv	H	0.371	0.010	0.088			0.000		2.856	5.424			0.963
Iridaceae	Iridaceae1 (Jul17)	NC	NC	NC									7.369	0.000		0.912
Rosaceae	*Sorbus* spp.	BW	Ht	T	9.341	0.866	0.000	0.000	0.001	0.000		0.211	0.000		0.000	0.835
Fagaceae	*Quercus* spp.	BW	Nv	T	0.032	0.001		0.001	0.007		8.442	0.005	0.000	0.002	0.000	0.523
Adoxaceae	*Sambucus/Viburnum* spp.	Ht	Ht	S	0.044	1.602			0.001	0.011	0.070	3.825		0.000	0.005	0.463
Apiaceae	Apiaceae1 (Jun-Jul)	NC	NC	NC		4.294	0.002	0.000	0.000		0.000	0.057	0.442	0.002		0.374
Onagraceae	*Epilobium angustifolium*	G	Nv	H	0.001		0.252	0.827	0.243	1.755		0.000			0.001	0.317
Myrtaceae	*Eucalyptus/Myrtus* spp.	Ht	Ht	T	0.000		0.078	0.494	2.440				0.000			0.305
Hypericaceae	*Hypericum* spp.	GHt	NvHt	H		0.022	1.076	0.918	0.077					0.763		0.289
Hydrangeaceae	*Hydrangea* spp.	Ht	Ht	S		0.002	0.001		0.001	0.001	0.000	0.000		3.008	0.037	0.284
Poaceae	*Dactylis glomerata*	G	Nv	H			0.000		0.291	0.000		0.001		0.000	4.006	0.281
Plantaginaceae	*Plantago lanceolata*	G	Nv	H	2.071	0.006	0.000		0.000	0.001		0.001		0.000	1.779	0.279
Apiaceae	Apiaceae2 (Aug-Sept)	NC	NC	NC		0.015		2.039	0.390				0.010	0.125		0.235
Lythraceae	*Lythrum salicaria*	G	Nv	H			0.002	0.389	1.934	0.001						0.235
Fabaceae	*Otholobium* spp.	Ht	Ht	H	0.000		0.000	2.476				0.000		0.000		0.220
Asteraceae	Cynareae	GHt	NvHt	NC	0.000	0.939	0.515	0.524	0.208	0.022				0.012		0.200
Ranunculaceae	*Actaea* spp.	Ht	Ht	H				1.884	0.308	0.000				0.002	0.002	0.199
Ranunculaceae	*Clematis* spp.	Ht	NvHt	S	0.160			1.437	0.128	0.016				0.055	0.348	0.182
Asteraceae	*Rudbeckia* spp.	Ht	Ht	H										1.877		0.176
Cornaceae	*Cornus* spp.	Ht	Ht	T	2.039	0.007										0.165
Asteraceae	*Asteraceae2 (Aug)*	NC	NC	NC		0.000								1.742		0.163
Onagraceae	*Epilobium* spp.	G	Nv	H	0.000			1.095	0.614						0.038	0.163
Brassicaceae	*Cardamine* spp.	G	Nv	H	0.015	0.000				0.840	0.001	0.497	0.000			0.136
Fabaceae	*Lotus* spp.	G	Nv	H	0.000		1.101			0.005	0.001	0.000				0.130
Malvaceae	*Tilia* spp.	Ht	Ht	T			0.003	1.363					0.000		0.025	0.123
Elaeagnaceae	*Hippophae rhamnoides*	Ht	Ht	S	0.000					1.051		0.000				0.113
Sapindaceae	*Aesculus hippocastanum*	BW	Ht	T	0.706	0.307	0.000		0.010	0.183	0.002	0.030		0.037		0.106
Papaveraceae	*Papaver* spp.	Ht	Ht	H		1.040	0.048		0.000			0.000		0.221		0.102

Lamiaceae	Mentheae	GHt	NvHt	H	0.001		0.000			0.783							0.085
Asteraceae	Asteraceae6 (Apr)	NC	NC	NC	0.017				0.000	0.752	0.011	0.001					0.083
Bignoniaceae	*Campsis* spp.	Ht	Ht	S					0.775	0.003							0.080
Onagraceae	*Oenothera* spp.	Ht	Ht	H		0.392	0.071	0.107	0.135	0.044		0.001	0.000	0.002			0.065
Ranunculaceae	*Anemone nemorosa*	BW	Nv	H						0.591		0.000					0.064
Asteraceae	Asteraceae	NC	NC	NC	0.000	0.123	0.047	0.123	0.083	0.002	0.006	0.061	0.096	0.007			0.053
Boraginaceae	*Nemophila* spp.	Ht	Ht	H		0.721											0.053
Hypoxidaceae	Hypoxidaceae	NC	NC	NC		0.001		0.000					0.009	0.000	0.786		0.050
Arecaceae	*Chamaerops/Trachycarpus* spp.	Ht	Ht	T		0.052						0.523	0.000				0.050
Convolvulaceae	*Convolvulus/Calystegia* spp.	H	Nv	H			0.000		0.023			0.000		0.010	0.724		0.049
Fabaceae	*Lupinus* spp.	Ht	Ht	H		0.663				0.001	0.001		0.000				0.049
Oleaceae	*Ligustrum* spp.	Ht	Ht	S		0.001	0.354	0.000	0.055	0.000		0.000					0.047
Asteraceae	Asteraceae 1 (Sept17)	NC	NC	NC	0.001	0.000				0.000		0.000			0.750		0.047
Ranunculaceae	*Ficaria* spp.	G	Nv	H	0.000					0.280		0.101					0.039
Asparagaceae	*Hosta* spp.	Ht	Ht	H	0.465										0.002		0.038
Rutaceae	Rutaceae	NC	NC	NC	0.003				0.000	0.333		0.000		0.000			0.036
Poaceae	Poaceae	NC	NC	NC	0.139	0.007	0.062		0.131	0.002		0.000			0.042		0.035
Ranunculaceae	*Helleborus* spp.	Ht	Ht	H						0.316		0.001	0.001				0.034
Rosaceae	Rosaceae	NC	NC	NC	0.007	0.105	0.087	0.008	0.010	0.059	0.000	0.012	0.001	0.010	0.025		0.030
Fabaceae	*Trifolium pratense*	G	Nv	H		0.001						0.334	0.001				0.030
Plantaginaceae	*Veronica* spp.	GHt	NvHt	H				0.317									0.028
Iridaceae	Iridaceae2 (Sept17)	NC	NC	NC	0.000	0.000									0.449		0.028
Asparagaceae	*Hyacinthoides non-scripta*	BW	Nv	H	0.004					0.006		0.301					0.027
Solanaceae	*Nicotiana* spp.	Ht	Ht	H		0.000								0.279			0.026
Asteraceae	Asteraceae5 (Jun)	NC	NC	NC								0.292		0.000			0.026
Ericaceae	*Arbutus unedo*	Ht	Ht	T				0.001	0.247						0.000		0.026
Anacardiaceae	*Cotinus coggygria*	Ht	Ht	T		0.329											0.024
Asteraceae	Asteraceae 3 (Jun-Jul)	NC	NC	NC		0.064	0.045					0.071	0.040	0.007			0.022
Caryophyllaceae	*Silene* spp.	H	Nv	H		0.281											0.021
Asteraceae	*Asteraceae 9 (Jun)*	NC	NC	NC		0.241	0.004	0.003	0.002				0.000				0.019
Rosaceae	Potentilla spp.	GHt	NvHt	S		0.238											0.017
Campanulaceae	Campanula spp.	Ht	Ht	H		0.001		0.175	0.003			0.002					0.016
Menyanthaceae	*Menyanthes trifoliata*	G	Nv	H	0.033	0.000				0.114	0.010	0.003		0.000			0.016
Ranunculaceae	*Caltha palustris*	G	Nv	H						0.144							0.015
Cistaceae	*Cistus/Helianthemum* spp.	Ht	Ht	S		0.204				0.005							0.015
Asteraceae	Asteraceae 8 (Sept16)	NC	NC	NC			0.001		0.142					0.001			0.015
Oleaceae	Oleaceae 2 (May16)	NC	NC	NC	0.166												0.013
Asteraceae	Asteraceae 4 (Jul)	NC	NC	NC			0.111		0.001				0.000	0.000			0.013
Urticaceae	*Urtica* spp.	G	Nv	H			0.093										0.011
Fabaceae	*Vicia* spp.	G	Nv	H				0.117									0.010
Oxalidaceae	Oxalidaceae	NC	NC	NC									0.000		0.156		0.010
Fagaceae	*Castanea sativa*	Ht	Ht	T				0.013	0.083								0.010
Scrophulariaceae	*Verbascum/Buddleja* spp.	Ht	Ht	S					0.087	0.000							0.009
Ranunculaceae	*Thalictrum* spp.	Ht	Ht	H										0.088			0.008
Oleaceae	Oleaceae 3 (Sept16+17)	NC	NC	NC			0.002		0.050						0.028		0.007
Lamiaceae	*Callicarpa* spp.	Ht	Ht	S				0.064									0.006
Ranunculaceae	Ranunculaceae	NC	NC	NC	0.005		0.000	0.003	0.002	0.012		0.030		0.005	0.001		0.005
Lardizabalaceae	Lardizabalaceae	NC	NC	NC				0.006	0.015	0.027				0.001	0.000		0.005
Francoaceae	*Francoa* spp.	Ht	Ht	H					0.048								0.005
Iridaceae	Iridaceae 3 (Sept16)	NC	NC	NC					0.039								0.004
Asteraceae	Asteraceae 7 (May)	NC	NC	NC	0.046						0.000	0.000					0.004
Apiaceae	*Eryngium* spp.	Ht	Ht	S										0.039			0.004
Ranunculaceae	*Anemone* spp.	Ht	Ht	H				0.033	0.001					0.001	0.006		0.003
Oleaceae	*Fraxinus* spp.	BW	Nv	T	0.019					0.017							0.003
Ranunculaceae	*Aconitum* spp.	Ht	Ht	H	0.000			0.036				0.000		0.001			0.003
Ranunculaceae	*Aquilegia* spp.	Ht	Ht	H	0.018	0.000		0.011	0.000			0.005	0.000	0.000			0.003
Brassicaceae	Brassicaceae	NC	NC	NC						0.022		0.005					0.003
Lamiaceae	*Lavandula* spp.	Ht	Ht	H		0.037											0.003
Aquifoliaceae	*Ilex* spp.	BW	Nv	S	0.029												0.002
Pinaceae	*Pinus* spp.	Ht	Ht	T	0.029												0.002
Commelinaceae	*Tradescantia* spp.	Ht	Ht	H		0.032											0.002
Oleaceae	Oleaceae1 (Apr17)	NC	NC	NC						0.019							0.002
Cannabaceae	*Humulus lupulus*	Ht	Ht	H				0.001	0.008	0.010							0.002
Magnoliaceae	*Magnolia* spp.	Ht	Ht	T						0.015							0.002
Rosaceae	*Chaenomeles* spp.	Ht	Ht	S	0.018					0.001							0.002
Brassicaceae	*Hesperis* spp.	Ht	Ht	H	0.005					0.008							0.001
Berberidaceae	*Berberis* spp.	Ht	Ht	S	0.001		0.000		0.001	0.009		0.001					0.001
Rosaceae	*Dasiphora* spp.	Ht	Ht	S			0.001	0.001	0.011					0.000			0.001
Cucurbitaceae	*Cucurbita* spp.	Ht	Ht	H	0.014												0.001
Buxaceae	*Buxus* spp.	Ht	Ht	S		0.013											0.001
Amaryllidaceae	*Galanthus* spp.	Ht	Ht	H						0.009							0.001
Tropaeolaceae	*Tropaeolum* spp.	Ht	Ht	H										0.006			0.001
Boraginaceae	*Borago officinalis*	Ht	Ht	H			0.001		0.000	0.003							0.000
Amaryllidaceae	*Narcissus* spp.	Ht	Ht	H						0.004							0.000
Betulaceae	*Betula* spp.	BWHt	NvHt	T		0.006											0.000
Ericaceae	*Rhododendron* spp.	Ht	Ht	S						0.002							0.000
Iridaceae	Iridaceae 4 (Apr17)	NC	NC	NC						0.002							0.000
Malvaceae	Malvaceae	NC	NC	NC						0.001				0.000	0.000		0.000
Paeoniaceae	*Paeonia* spp.	Ht	Ht	H		0.002											0.000
Asparagaceae	*Convallaria* spp.	Ht	Ht	H						0.001							0.000
Rhamnaceae	*Frangula alnus / Rhamnus cathartica*	BW	Nv	S						0.001							0.000
Amaryllidaceae	*Allium* spp.	Ht	Ht	H				0.001									0.000
Celastraceae	*Euonymus* spp.	Ht	Ht	S	0.001												0.000
Fabaceae	Fabaceae	NC	NC	NC					0.000			0.001	0.000	0.000			0.000
Ulmaceae	*Ulmus* spp.	Ht	Ht	T						0.001							0.000
Lamiaceae	*Glechoma* spp.	G	Nv	H						0.001							0.000
Grossulariaceae	*Ribes* spp.	Ht	Ht	S						0.001							0.000
Solanaceae	Solanaceae (Apr17)	NC	NC	NC						0.001							0.000
Moraceae	*Morus* spp.	Ht	Ht	S											0.001		0.000

3.4.4. Phenological progression of plant taxa in the honey

Time, represented by the month the honey sample was collected, was found to significantly predict changes in honey composition through the foraging season (Figure 10, Figure 11; $LR_{1, 89} = 428$, $P = 0.003$), along with the year (2016, 2017) the sample was collected ($LR_{1, 89} = 277$, $P = 0.003$). Across the two years sampled, 16 taxa were classified as major in at least one month (Figure 10). In April 2017, the top taxa returned were *Salix* spp., followed by *Prunus* spp., two key early flowering tree genera. The *Crataegus, Malus*, *Cotoneaster* spp. group in the tribe Maleae was top in May 2016 and second to *Taraxacum officinale* in May 2017. In both years *Acer* spp. were also returned highly in May. *Rubus* spp. were the top taxon found in June, July and August for both years, while in September 2016 and 2017 *Hedera helix* was highest, known to be a key late-flowering species. Other major forage taxa included *Trifolium repens* and the *Cirsium, Hypochaeris* spp. grouping, both found predominantly in the June, July and August samples, as well as *Rosa* spp. which were found both years in June. *Impatiens glandulifera* was a major forage found mostly in August and September.

While the years sampled were similar in terms of number of different taxa returned (110 in 2016 and 113 in 2017) there was variation in the abundance of the taxa (Table 4, Figure 10). Looking on a monthly basis between the years, a similar number of taxa were returned as major and secondary forage with 28 taxa in 2016 and 30 taxa in 2017. However, 14 taxa were designated major forage in 2017 while only seven were found in 2016. The major taxa consistent between years were *Rubus* spp., the *Crataegus*, *Malus*, *Cotoneaster* spp. group, *Hedera helix*, a member of the Salicaceae family, and *Rosa* spp. The majority of taxa found to be major in one year were classed as secondary in another, as with *Trifolium repens,* where in 2016 the months in which it appears were mostly dominated by *Rubus* spp. The exceptions were *Salix* spp. and *Ulex* spp., two genera returned as major in the April 2017 honey, which was not sampled in 2016. Only one taxon was found highly in 2017 and not at all in 2016: the horticultural genus *Weigela* spp.

There was a greater difference between the years for the 28 plant taxa designated secondary forage (Figure 11), with only three of the 28 plants considered secondary forage for both 2016 and 2017: *Filipendula ulmaria,* the *Sambucus* and *Viburnum* spp.

grouping, and *Plantago lanceolata*. The *Sambucus* and *Viburnum* spp. grouping was found consistently in June, while *F. ulmaria* was found in July, August and September for 2016 and mostly found in August for 2017. *P. lanceolata,* which flowers throughout the season, was particularly distinct between the two years being found in May 2016 and September 2017.

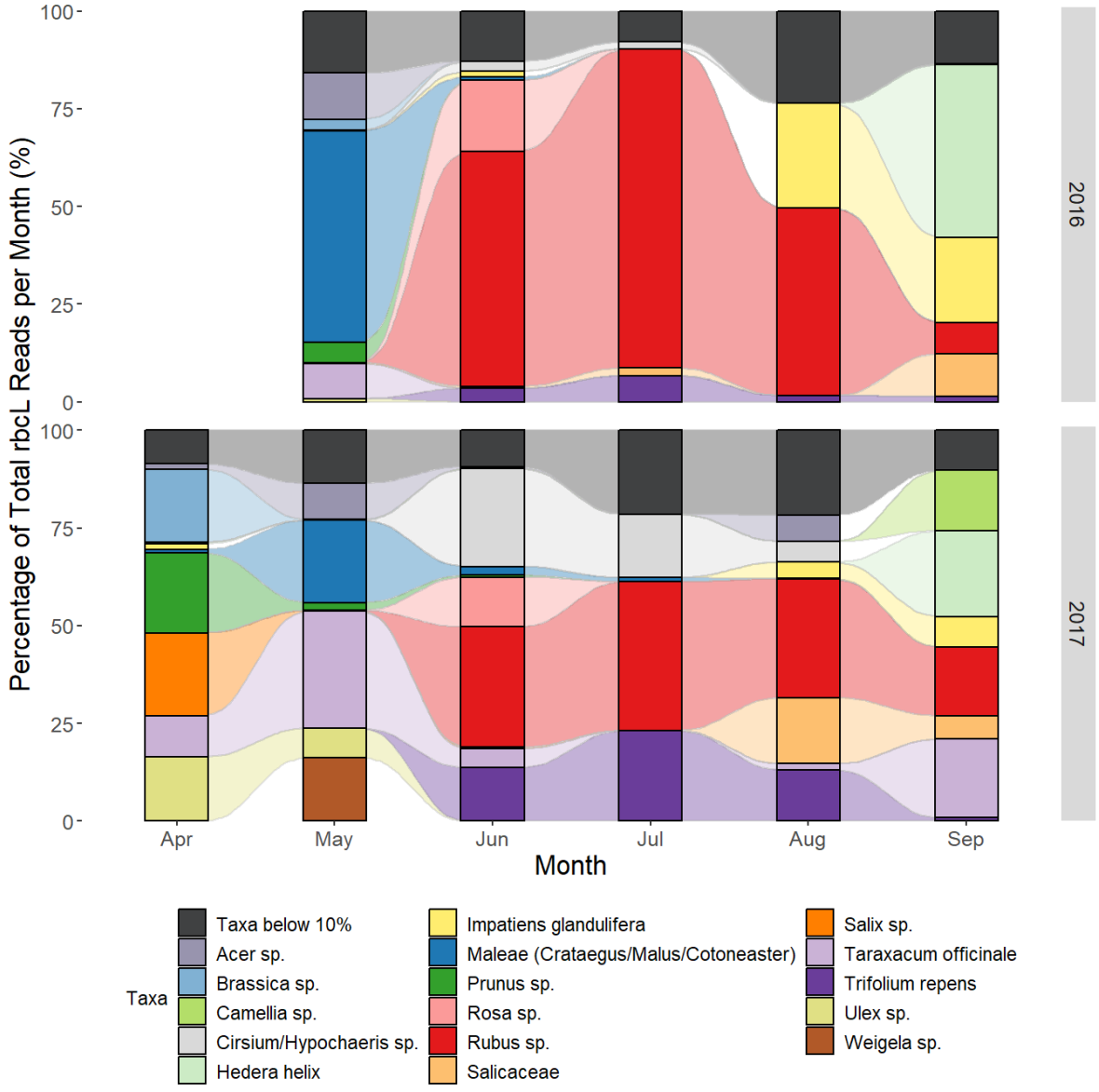


Figure 10: Major plant forage found within the honey from April to September for 2016 and 2017. Plant taxa are labelled when the sequencing reads returned were over 10% for at least one month.

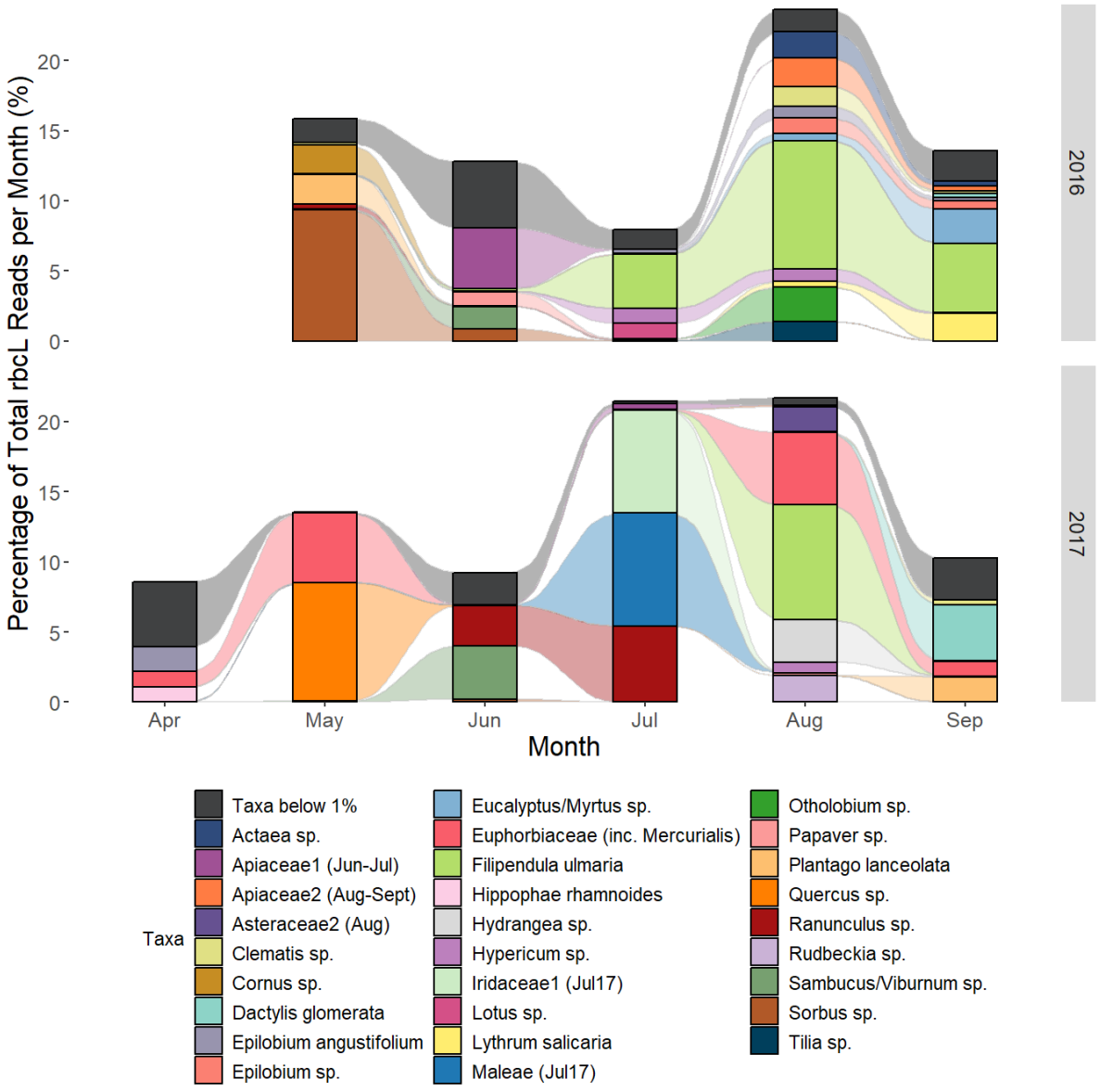


Figure 11: Secondary plant forage found within the honey from April to September for 2016 and 2017. Plant taxa are labelled when the sequencing reads returned were between 1% and 10% for at least one month.

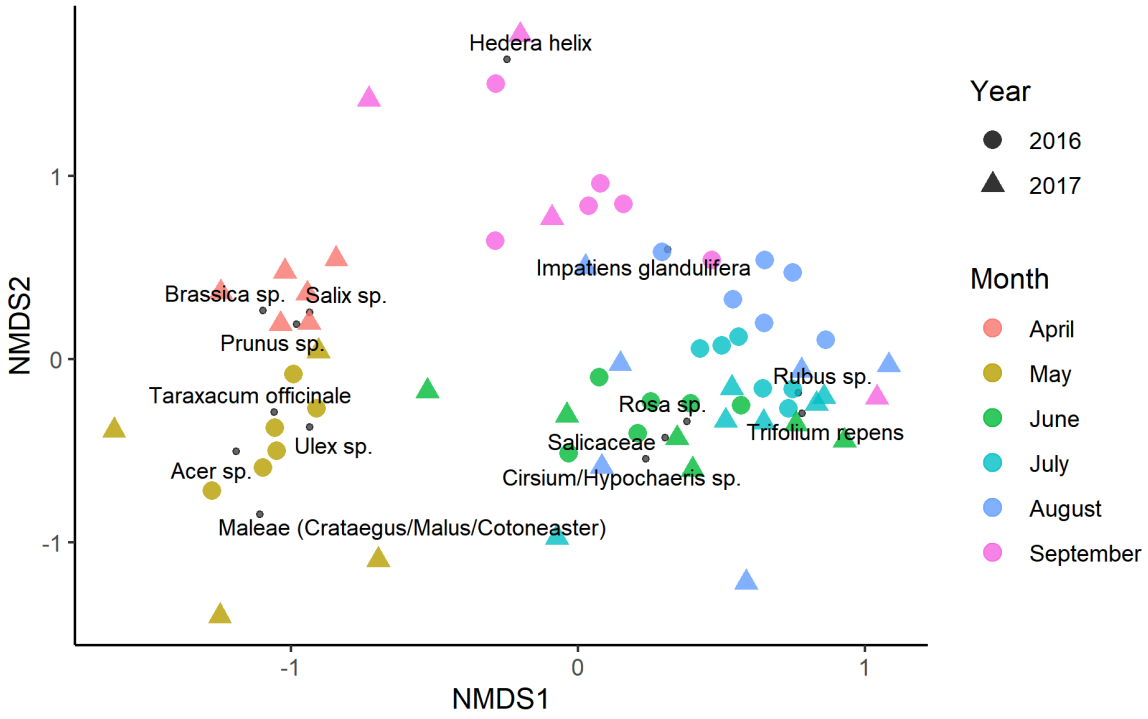


Figure 12: Non-metric multidimensional scaling (NMDS) ordination of the honey samples. Colour indicates the month of collection and shape indicates the year. Plant taxa found in over 10% of the reads for each month are plotted separately indicating the top taxa driving the changes of month on community composition of the honey.

Non-metric multidimensional scaling (NMDS) ordination supports that the samples from each month are most similar to each other, with increasing dissimilarity as the season progresses (Figure 12). This is seen in the plot, with the April and May samples clustering closer than the September samples, likely driven by the continued low level presence of plant taxa in later samples which are associated more strongly with samples from earlier in the season (Figure 12). There is good correspondence with the flowering phenology, with April and May samples clustering together, and key plants at this time being the Maleae tribe representing *Crataegus*, *Malus*, and *Cotoneaster* spp., as well as *Salix* spp., *Taraxacum officinale, Prunus* spp., *Brassica* spp., *Acer* spp. and *Ulex* spp. Of these, all show peak flowering during either April or May, excepting *Ulex* spp. which can flower throughout the season. In June, July and August the top plants are *Rubus* spp., *Rosa* spp., *Trifolium repens* and the *Cirsium, Hypochaeris* spp. grouping. The spread in samples from September is being driven by differing proportions of *Hedera helix* and *Impatiens glandulifera* found in each hive. Out of the ten hive samples for September 2016 and

2017, one returned *Rubus* spp. and *Trifolium repens* as its top species and this clusters with the June and July samples.

3.4.5. Hive level foraging differences

The majority of the top 16 major forage taxa were also present across all six hives each year, with 15 of the taxa collected by all hives in 2016 and 14 in 2017. The exceptions were *Weigela* spp. which was collected by only one hive in May 2017, *Impatiens glandulifera* which was found in three hives in 2017, and *Brassica* spp. which was found in five hives in 2016 (Figure 13). As with comparing between the years, a greater difference was seen between the hives when looking at the secondary forage plants, with only 27% of these taxa present in all hive samples in 2016, and 15% in 2017. Variation in the phenological trend of the plants was also usually found on an individual hive basis. For example, in September, while the autumn flowering *Hedera helix* and *I. glandulifera* were the top species in six hives, one hive sample in September was found to contain primarily *Taraxacum officinale*, a possible result of honey collected earlier in the season still being present in the hive.

When comparing the three hives located in the Botanic Garden and the three hives located in the Nature Reserve, the location of the hives was found to drive the community composition of pollen in the honey ($LR_{1, 89} = 174$, $P = 0.003$) and this is supported by the total unique plant genera ($LR_{1, 89} = 124254$, $P = 0.003$) and plant diversity ($LR_{1, 89} = 282056$, $P = 0.003$), in the surrounding surveyed habitat of each location. Of the top taxa, Salicaceae showed a consistent difference between the proportion of reads returned for the two locations, with the Nature Reserve hives having a greater abundance of Salicaceae than the Botanic Garden hives for both 2016 and 2017. The *Chamaerops* and *Trachycarpus* spp. grouping was a horticultural plant used at low levels and only found in the Botanic Garden hives in both 2016 and 2017.

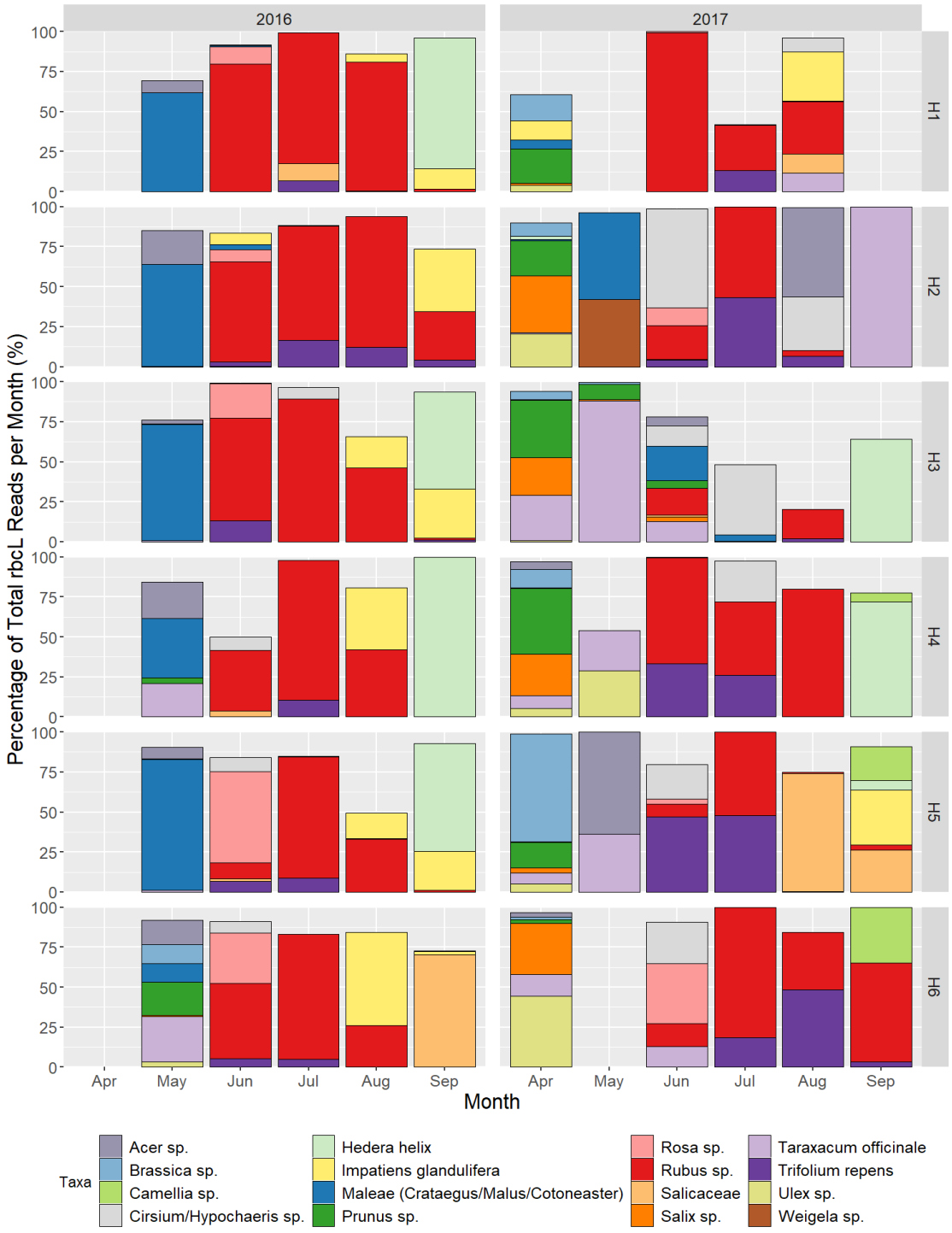


Figure 13: Illustrating the hive level foraging choices in major plants (>10% of reads return in one month) across the season for each hive sampled. Hives one to three were located within the Botanic Garden and hives four to six were located in the Nature Reserve.

3.4.6. Comparison with available floral resources

Overall, of the 100 plant taxa identified to genus or species level in the honey, 86 were recorded within the survey area during 2016 or 2017. Of the 14 plants that were not recorded, the majority were tree or grass species known to be planted or present in the site. On a monthly basis, disparity between the genera found in the honey, and the number of those genera that were recorded as flowering in that month is mostly due to the presence of low levels of plants carrying over in the honey samples. For example, in the April honey samples, late-flowering plants were found from the end of the previous season, such as *Hedera helix*.

Table 5. The plant genera used compared to plant genera available in the survey area, for hives located in the Botanic Garden and hives located in the Nature Reserve. The plant genera not flowering in the survey area but found in the honey were explained by low levels of plants in the honey from different points in the season. The comparison was limited to plants identified to at least genus level with the DNA.

	Botanic Garden		Nature Reserve		Overall	
	Total genera in honey	Genera in honey found in survey area during month	Total genera in honey	Genera in honey found in survey area during month	Total Genera in flower	Genera utilised
2016						
May	28	26 (93%)	24	24 (100%)	383	9%
Jun	28	26 (93%)	18	14 (78%)	405	8%
Jul	17	11 (65%)	17	9 (53%)	339	6%
Aug	23	10 (43%)	13	12 (92%)	326	7%
Sep	29	18 (62%)	16	11 (69%)	302	9%
2017						
Apr	29	21 (72%)	27	19 (70%)	270	13%
May	8	7 (88%)	10	8 (80%)	311	5%
Jun	21	12 (57%)	21	16 (76%)	346	7%
Jul	16	9 (56%)	16	8 (50%)	326	4%
Aug	26	22 (85%)	10	5 (50%)	340	8%
Sep	6	3 (50%)	16	11 (69%)	311	7%

After removing these low-level taxa whose presence is explained by a previous month's forage, the plant genera known to be in flower within the survey area were compared with the plant genera found in the honey. The results show that the honeybees were only using a small proportion of what was available throughout the season (Table 5). April

2017 has the highest proportion of plant genera available that were also found in the honey with 13%, whilst the lowest was July 2017 with 4% of genera available used by the honeybees.

3.4.7. Relationship to status, form, and habitat of plant taxa

Of the 136 taxa returned, 104 taxa at species, genus and tribe level were classified according to status, form and habitat of the plant (Table 4). In terms of the proportion of sequence reads returned, native plants represented 64% of total reads, horticultural plants were 13% of reads, native-horticultural were 10% and the single naturalised plant, *Impatiens glandulifera,* was 6% of total reads returned (Figure 14).

For plant form, overall 44% of sequences returned were shrubs, 32% were herbs and 16% were trees. While trees contributed lowest to the total forage, when the variation in plant form over the months is examined, trees were used mostly in the early foraging season of April and May (Figure 15), and comparatively much less as the season progressed, where foraging was split between herbs and shrubs. Early flowering tree species used by the honeybees here include *Salix*, *Prunus*, *Acer*, *Sorbus*, *Quercus* and *Cornus* spp.

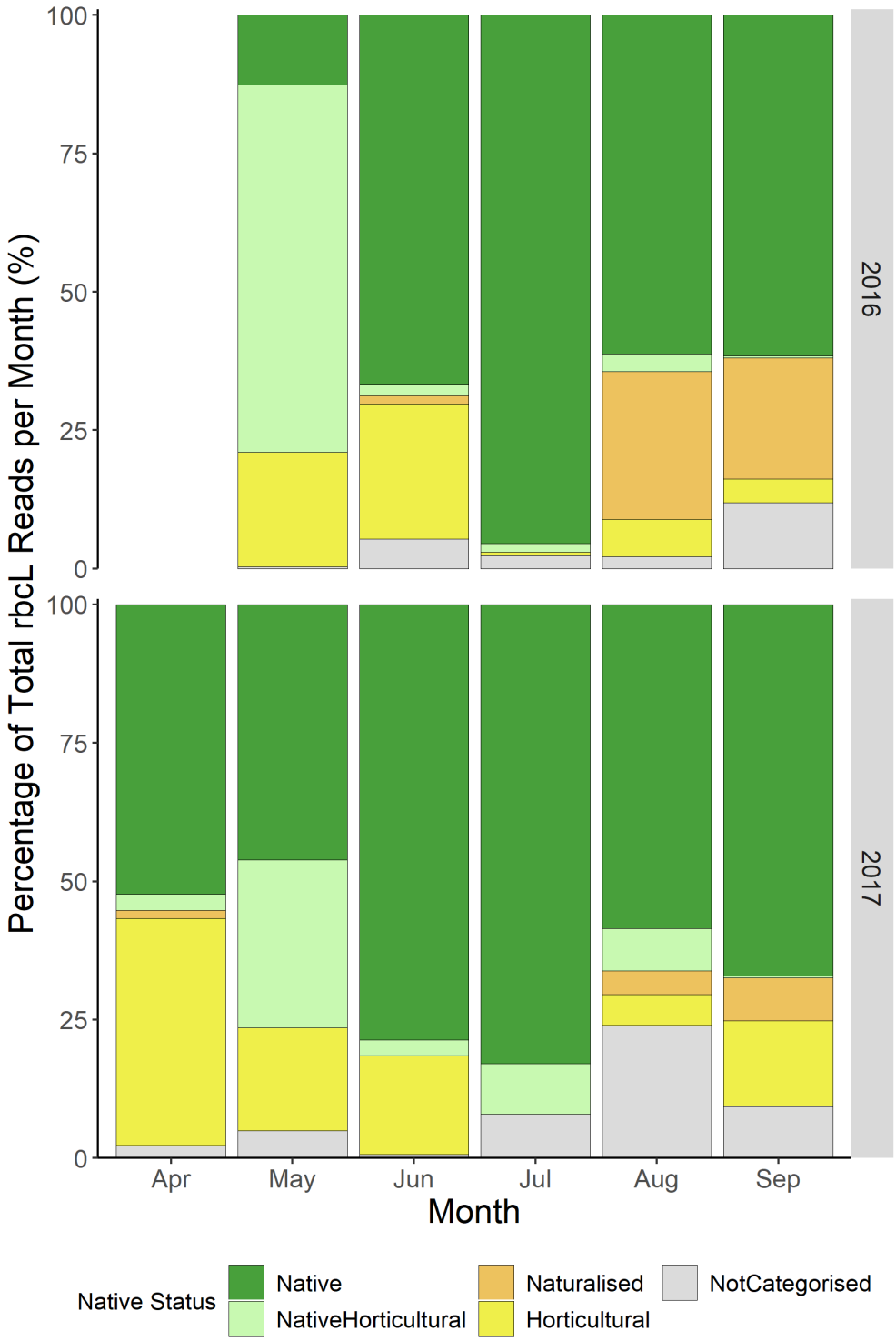


Figure 14: Proportion of *rbcL* sequence reads returned from the honey samples characterised by the native status of the plant within the locality of the survey area. Taxa returned at family level were not categorised.

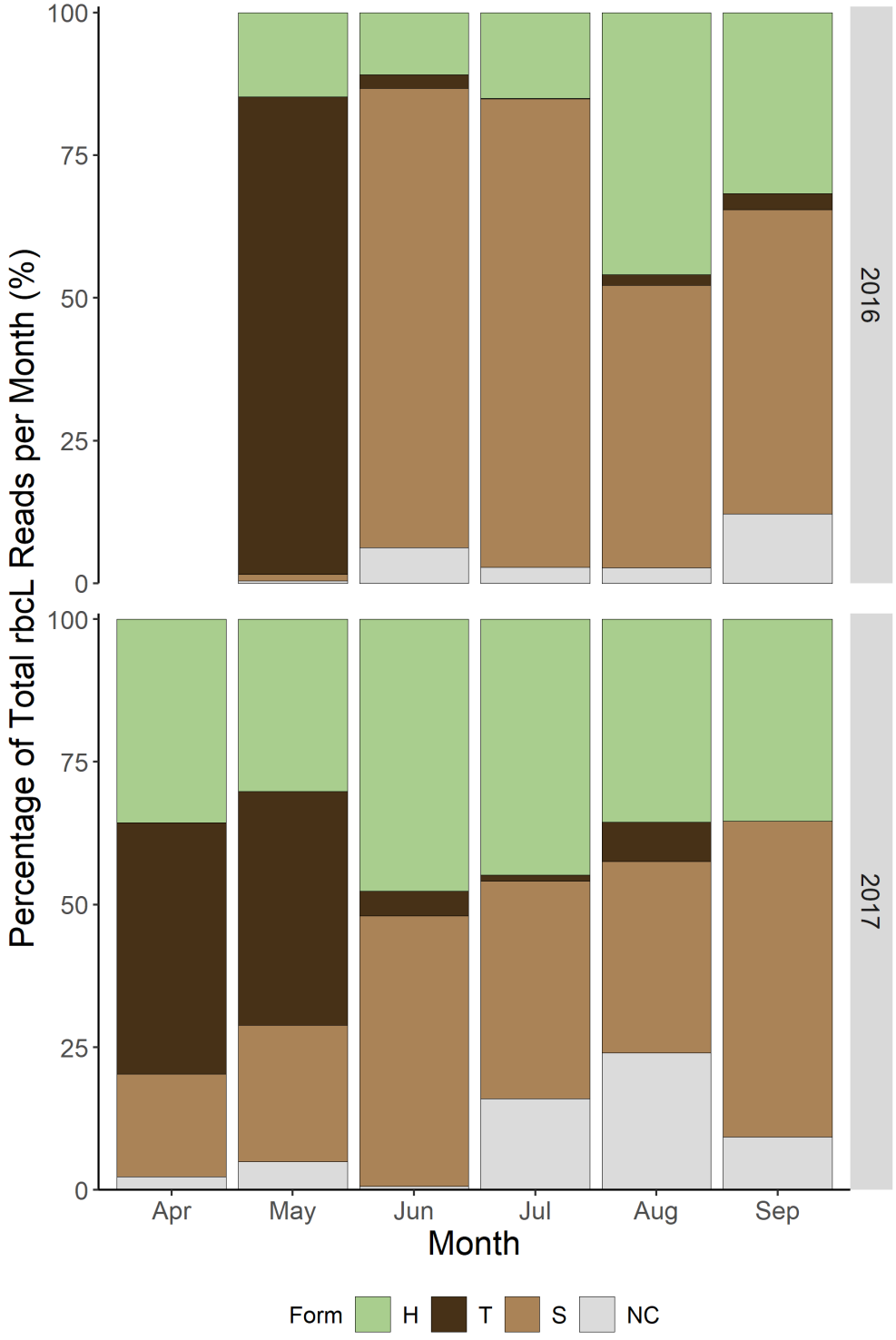


Figure 15. Proportion of *rbcL* sequence reads returned from the honey samples with plant taxa classified into herbs (H), trees (T) and shrubs (S). Taxa returned at family level were not categorised (NC).

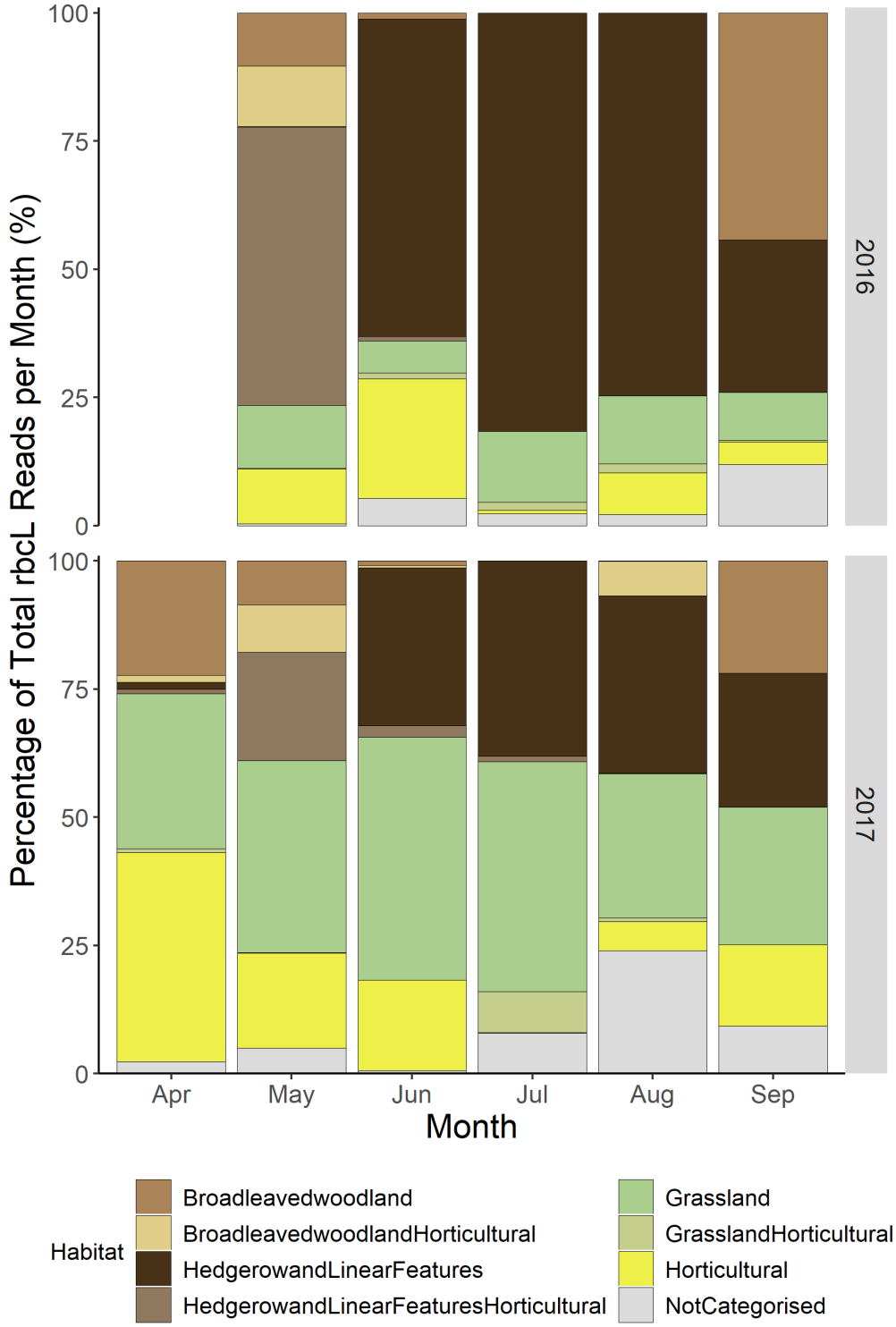


Figure 16: Proportion of *rbcL* sequence reads returned from the honey samples characterised by the main habitat type of the plant taxa within the locality of the survey area. Taxa returned at family level were not categorised.

For the main habitat of the plant, hedgerow and linear features accounted for 30% of the total sequences returned, driven by *Rubus* spp. and *I. glandulifera,* while grassland was 25% of reads, broadleaved woodland was 16% of reads, and horticultural was 12% (Figure 16). Of the plant taxa equally likely to be found in a native habitat and horticultural habitat, 6% of sequences were classified as hedgerow-horticultural, 2% woodland-horticultural, and 2% grassland-horticultural.

3.4.8. Relationship between taxa abundance in the landscape and metabarcoding abundance

In 2017, the area of abundance for each plant species was recorded for the available plant species in order to examine the relationship with the proportion of reads found within the honey. The relationship between abundance for the taxa found in both the survey area and in the honey was examined for each month (Figure 17). A significant correlation was not found from April to September between abundance in the local landscape and abundance in the honey (Figure 17).

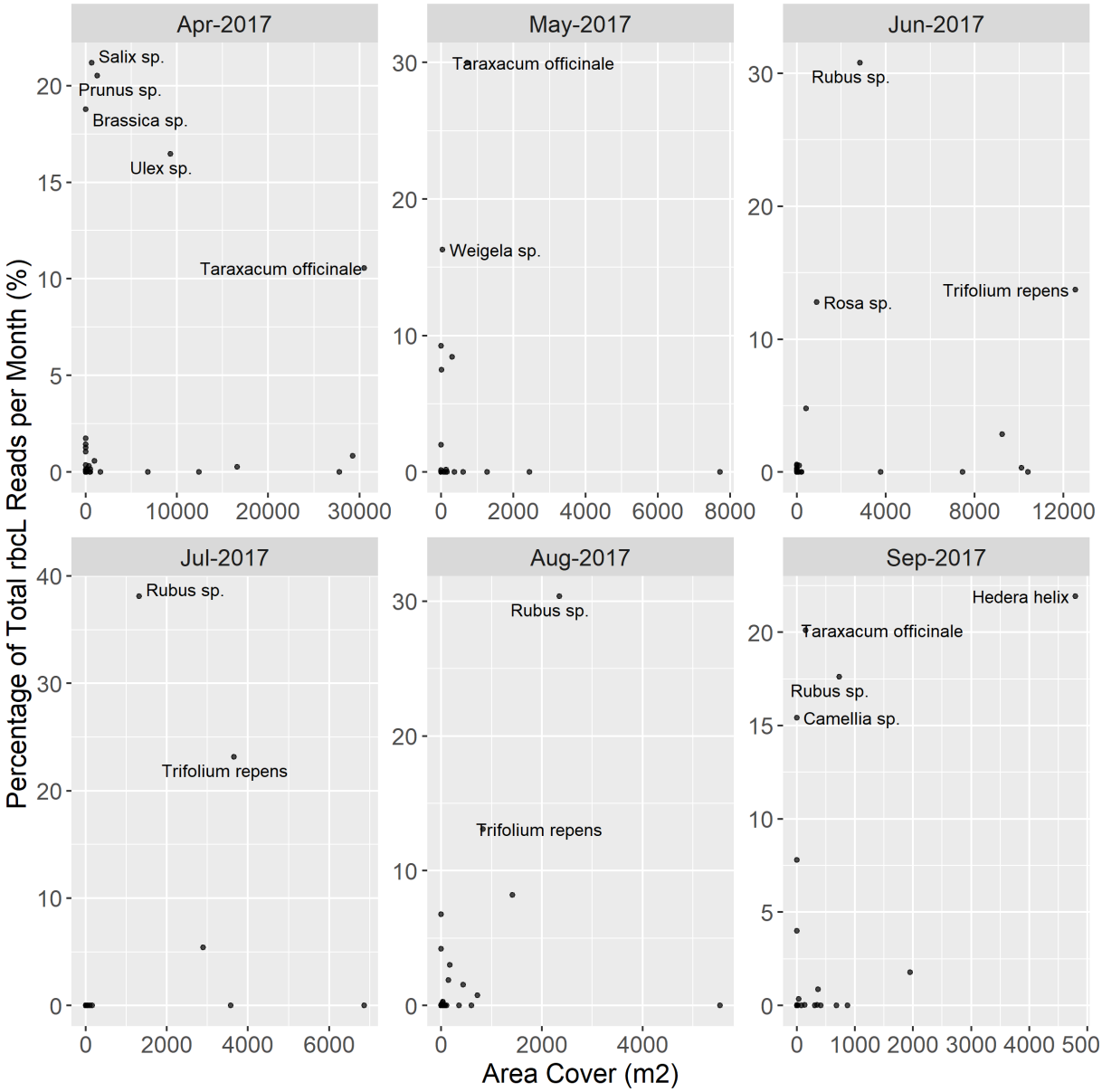


Figure 17. Examining the relationship between abundance within the survey site and abundance in the honey. Only plant taxa which were identified at a species or genus level were included. No significant correlation was found for any of the months using Spearman's rank correlation with Bonferroni correction for multiple testing: April (Spearman's rank correlation coefficient: r_s = 0.199, p=1, May (r_s = -0.253, p=1), June (r_s= 0.235, p=0.971), July (r_s = 0.367, p=0.509), August (r_s = 0.414, p=0.090), September (r_s = 0.426, p=0.302).

3.5. Discussion

The honeybees within this study had access to a high diversity of plants, both native and horticultural but only used a small proportion of the floral resource available. The honeybees visited a wide variety of horticultural plants, but at lower levels compared to the native and near native plant species, indicating that honeybees with a diverse plant offering are still reliant on plants from predominantly native habitats to supply most of their nutritional needs. In Baude *et al.*, (2016) nectar productivity for 270 nationally important plant nectar species was modelled, finding that 26 plant genera produce over 95% of nectar at a national land cover scale for native and agricultural habitats. Of those 26 genera, 14 were represented at species or genus level in the honey collected here, including the most abundant taxa: *Rubus* spp., *Trifolium repens*, Maleae (*Crataegus monogyna)* and *Hedera helix*.

The month of sampling was the biggest predictor of change in the composition of the plant taxa found within the honey. Similar results in terms of the abundance of plants found within the honey and foraging phenology have been seen in studies in the UK and Ireland using microscopic techniques to identify the pollen in honey and pollen loads. In Coffey and Breen (1997), plants identified in freshly collected nectar from hives in Ireland are described, showing a similar foraging phenology in spring for April and May as seen here, with *Salix* spp., *Ulex* type and *Prunus/Pyrus* type identified from the pollen morphology. Similarly, in June and July *Trifolium repens* and *Rubus spp*. were the main providers of pollen and nectar with *Filipendula ulmaria* supplying additional pollen during July and August, matching the results shown in this study. In Percival, (1947) *Rubus fruticosus* and *Trifolium repens* were identified from pollen loads as the major forage in June, July and August. For September, *Impatiens glandulifera* and *Hedera helix* were also found to be important forage in Coffey and Breen (1997). *Calluna vulgaris* was also identified as yielding large quantities of pollen, which was not found in the honey at the Botanic Garden despite being present at low quantities in the survey area. *H. helix* has been noted as an important autumn resource for honeybees and other pollinators, for both nectar and pollen (Garbuzov and Ratnieks, 2014).

The flowering phenology of the plants found in the honey matched that of the survey area well, with spring species found in the April-May samples (e.g. *Salix* spp., the Maleae tribe, *Crataegus* spp*., Malus* spp., *Cotoneaster* spp.), and key late-flowering species present in the September samples (e.g. *Hedera helix, Impatiens glandulifera*). However, species associated with different seasons were found at lower levels in different months, most notably in April 2017: where low levels of *H. helix* and *I. glandulifera* can be explained by the carried over presence of honey stores from the previous year. Higher levels of discordant taxa compared to the season were usually found at the hive level. For example, in honey from September 2017, one hive returned majority *Taraxacum officinale* reads, a species which can be found flowering from March till October but is most associated with mass flowering in April and May. Samples were taken of what was observed to be the most freshly capped honey, but honeybees are known to move honey around the hive (Eyer et al., 2016; Wright et al., 2018) potentially explaining the presence of plant species identified earlier in the season in samples from later months.

The year of sampling was also found to have a significant effect on the plant composition of the honey, indicating the importance of multiple years of sampling to build a stronger picture of the seasonality of forage. The plants used most frequently were more consistently present between the years, with changes in the relative abundance within the honey. There was more variety in the plants used at a lower level. Many environmental factors will affect both the availability of the nectar and the concentration of sugar from plants, such as temperature, humidity, or precipitation (Corbet, 2003; Corbet et al., 1979; Nicolson et al., 2007), which could lead to variation in the reward for a forager between seasons. Among beekeepers, *Crataegus monogyna* is considered inconsistent in the nectar flow offered for honey production between seasons (Howes, 1945). Compared to other woody Rosaceae species, *C. monogyna* has a shorter flowering period, possibly making it more vulnerable to being missed by the honeybees due to inclement weather (Gyan and Woodell, 1987).

Across all the hives and months, *Rubus* spp. alone accounted for 30% of the total reads returned, with *Trifolium repens* second at 7%. *Rubus* spp. have a long flowering period, with *Rubus fruticosus* flowering from June until September (Gyan and Woodell, 1987). *T. repens* has a similar flowering period. In 1946, in England, *Trifolium repens* was recorded

as being the main pollen source brought back to the hive (Synge, 1947). While, Baude *et al.*, (2016), found that *T. repens* contributes the most to nectar supply on a national scale. However, the use of *T. repens* in pasture has decreased in the UK since the increasing application of inorganic fertilisers and herbicides after 1940, possibly affecting its availability as a foraging resource (Balfour et al., 2018; Robinson and Sutherland, 2002). *Rubus* spp., as an abundant source of nectar and pollen during the same flowering period may be offering an alternative resource.

The majority of plant taxa found in the honey were also recorded in the survey area and represent the closest available instance of the plant within a small foraging range. Honeybees have been shown to forage 10 km, with the exact distance seeming to depend on the proximity of rewarding forage (Beekman and Ratnieks, 2000). Steffan-Dewenter and Kuhn (2003) found that foraging distances for pollen collection were significantly larger in simple landscapes compared to complex. The most notable plant not to be found in the survey area was *Impatiens glandulifera*, a highly invasive species that commonly occurs along rivers and streams. It has been shown that the rate of sugar production in *I. glandulifera* is higher than other plant species associated with the same habitat (Chittka and Schürkens, 2001), in addition to having the largest measured nectar sugar content per flower per day out of 270 nationally important UK nectar plant species (Baude et al., 2016). Floral abundance has been found to have a positive effect on insect visitation, including for *Apis mellifera*; with the relationship between visitation and total sugar available being weak at low floral abundance and positive at high floral abundance (Fowler et al., 2016). Foraging has a high energetic cost, and a high sugar reward therefore may attract honeybees into further foraging distances but only if the flower abundance is high.

No relationship was found here between the abundance of the plant in the landscape through the season (taken as the total percentage cover within the survey area each month) and the DNA metabarcoding abundance, indicating that the abundance of the plant within the landscape is not the main driver for selection for forage. When foraging honeybees make decisions based on additional factors as well as abundance, such as the distance of the forage, the time of day, and whether targeting nectar or pollen. By incorporating data on other drivers of nectar and pollen foraging with the abundance

data, a more refined picture of the landscape could be developed. For nectar this would involve estimates of floristic abundance in addition to percentage cover, as well as measurements of nectar production and concentration, or pollen quality. While many common UK native plants have been measured in this way for nectar (Baude et al., 2016), data is lacking on the horticultural species, and it has been shown that horticultural modifications can reduce the nectar reward or attractiveness to pollinators (Comba et al., 1999; Garbuzov et al., 2015). Incorporating data on the pollen reward of a plant would provide further resolution on the preferences of the honeybees. The colony level requirements are more likely to be driven by pollen needs than nectar, with the colony responding to pollen deficiencies in the hive by increasing the amount of pollen returned (Pernal and Currie, 2001), while nectar foraging is regulated with the availability of nectar rather than the stores in the hive (Seeley, 1995). Both nectar plants and pollen plants were represented in the honey. The majority of plants can provide both nectar and pollen, and while honeybee foragers can specialise in collecting either resource, they can also collect both at the same time (Wright et al., 2018).

In de Vere et al., (2017), honeybee foraging in April and May was characterised using DNA metabarcoding within the Botanic Garden. The results presented here are consistent with these spring results, continuing the trend that honeybees use a small percentage of the total genera available to them throughout the season. While the plants used are taxonomically diverse, there are a small number of core species which are forming the majority of the honeybees' nutritional needs. Honeybees have been referred to as super-generalists within plant-pollinator networks, however, as evidenced here, they are still making foraging decisions in the selection of plants within a system and not using everything available to them. Similarly, in a meta-analysis of plant-pollinator interaction networks, frequent visitation was found to be restricted to a minority of plant species (Hung et al., 2018). The plant species other generalist and specialist pollinators are selecting within the system is therefore a key area for further research, to establish a full understanding of the pollinator-plant assemblies. While diversity of forage will provide honeybees with nutritional variety and contingency against environmental variation in nectar and pollen availability, any siting of hives should consider their access to the highly abundant species which they frequently target.

The honey collected here in spring was found to contain more tree species present compared with later in the season, consistent with spring foraging patterns found in de Vere et al., (2017). A similar pattern is seen in Balfour *et al.*, (2018) with UK insect-pollinated plant species classified into trees, shrubs and herbs, with trees peaking in spring, shrubs in early summer and herbs in July. Beekeepers have long discussed a "June gap" where the availability of nectar from native floral sources is said to be lacking (Coffey and Breen, 1997; Crane, 1976; Percival, 1947), and Balfour et al., (2018) highlighted that this June gap may be occurring in between the flowering peaks of tree and herbaceous plants. The spring tree species found here, such as *Salix* spp., *Acer* spp., *Quercus* spp., and *Cornus* spp. all provide pollen, vital to the healthy growth and development of the larvae and the colony early in the season (Brodschneider and Crailsheim, 2010).

While the importance of garden habitats to pollinators has been shown in an urban areas (Baldock et al., 2019) and areas which are intensively farmed (Samnegård et al., 2011), we found that in a landscape with horticultural, semi-natural and native habitats, native plants made up the majority of taxa found. When comparing between trial plots with native, near-native and exotic plants, Salisbury et al., (2015) found that a greater floral resource resulted in an increase in pollinator visits, with a greater abundance of pollinators on native and near-native plants compared with exotic plants. Horticultural species here are used at low levels throughout the season. The habitat types of plants found in the honey emphasise the disproportionate contribution of plant taxa which are found within hedgerow and linear features, compared to the area covered in the study site. Hedgerows have been named as a potential way to efficiently increase the available nectar in a landscape, due to their high nectar productivity within a small area (Baude et al., 2016).

Here, the relative contribution of each plant taxon to the composition of the honey was analysed. The sequencing results of DNA metabarcoding are commonly considered semi-quantitative, with biases arising throughout the process (Alberdi et al., 2019; Zinger et al., 2019). A common approach to representing the sequence read numbers returned is to calculate the relative read abundance (Brennan et al., 2019; Deagle et al., 2019; Richardson et al., 2018). While the variance in the ability of the read count to act as a measure of the abundance of the biomass in the study system can be high, it has been

shown to provide a more accurate representation of diet than methods such as just using presence or absence of a taxon (Deagle et al., 2019; Lamb et al., 2019). A multi-locus approach has also been proposed as a method of improving the estimates of abundance returned by pollen (Richardson et al., 2018). Here, all honey samples were additionally amplified and sequenced with the ITS2 nuclear marker. There was a significant correlation between the abundance of reads for the plant taxa found by both *rbcL* and ITS2. This provides further support for the abundances returned by *rbcL*, and it being a good measure for comparing between those taxa which were used abundantly and those which were utilised at lower levels. However, the results found here with ITS2 did not provide further coverage of the plant community or allow for improved species discrimination, and so were excluded from further analysis. The rarefaction curves indicated that the sequencing depth was adequate for both *rbcL* and ITS2 to detect the diversity present in the samples. This indicates that the divergence in the number of taxa found by ITS2 may be a result of the stochasticity of the PCR.

3.5.1. Conclusions

Here, the foraging of honeybees was examined in a diverse landscape over multiple seasons. Using DNA metabarcoding we were able to discover the plants that are most important to the nutritional demands of the colony throughout the foraging season. The location of the hives explained a portion of the differences in plant communities found within the honey, but time and therefore floral phenology was the biggest predictor of floral composition. The major forage plants for both hive locations, those within the Botanic Garden and those within the Nature Reserve, were found to be the same. These major forage plants found in the honey throughout the season were characterised by being native plants, found in hedgerow and linear features and grassland habitats, while the plants foraged at a lower level throughout the season included horticultural plants. There are implications both for the management of habitat in the landscape for honeybees and the siting of hives. Only a small proportion of the plants found in the honey are being used abundantly, with both the quantity and quality of the floral resource likely impacting honeybee foraging. While the horticultural plants may be supplying the honeybees with the nutritional diversity that they require, any high quantity of hives should be placed considering their access to the semi-natural and native habitats which supply the majority of their nutritional needs.

3.6. References

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M., Gilbert, M.T.P., 2019. Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources* **19**, 327–348.

Anklam, E., 1998. A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chemistry* **63**, 549–562.

Baldock, K.C.R., Goddard, M.A., Hicks, D.M., Kunin, W.E., Mitschunas, N., Morse, H., Osgathorpe, L.M., Potts, S.G., Robertson, K.M., Scott, A. V., Staniczenko, P.P.A., Stone, G.N., Vaughan, I.P., Memmott, J., 2019. A systems approach reveals urban pollinator hotspots and conservation opportunities. *Nature Ecology & Evolution* **3**, 363–373.

Balfour, N.J., Garbuzov, M., Ratnieks, F.L.W., 2013. Longer tongues and swifter handling: Why do more bumble bees (*Bombus* spp.) than honey bees (*Apis mellifera*) forage on lavender (*Lavandula* spp.)? *Ecological Entomology* **38**, 323–329.

Balfour, N.J., Ollerton, J., Castellanos, M.C., Ratnieks, F.L.W., 2018. British phenological records indicate high diversity and extinction rates among late-summer-flying pollinators. *Biological Conservation* **222**, 278–283.

Baude, M., Kunin, W.E., Boatman, N.D., Conyers, S., Davies, N., Gillespie, M.A.K., Morton, R.D., Smart, S.M., Memmott, J., 2016. Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature* **530**, 85–88.

Beekman, M., Ratnieks, F.L.W., 2000. Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology* **14**, 490–496.

Brennan, G.L., Potter, C., de Vere, N., Griffith, G.W., Skjøth, C.A., Osborne, N.J., Wheeler, B.W., McInnes, R.N., Clewlow, Y., Barber, A., Hanlon, H.M., Hegarty, M., Jones, L., Kurganskiy, A., Rowney, F.M., Armitage, C., Adams-Groom, B., Ford, C.R., Petch, G.M., Creer, S., 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution* **3**.

Brodschneider, R., Crailsheim, K., 2010. Nutrition and health in honey bees. *Apidologie* **41**, 278–294.

Chase, M.W., Fay, M.F., Reveal, J.L., Soltis, D.E., Soltis, P.S., Peter, F., Anderberg, A.A., Moore, M.J., Olmstead, R.G., Rudall, P.J., Kenneth, J., 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**, 105–121.

Chittka, L., Schürkens, S., 2001. Successful invasion of a floral market. *Nature* **411**, 653–653.

Coffey, M.F., Breen, J., 1997. Seasonal variation in pollen and nectar sources of honey bees in Ireland. *Journal of Apicultural Research* **36**, 63–76.

Comba, L., Corbet, S.A., Barron, A., Bird, A., Collinge, S., Miyazaki, N., Powell, M., 1999. Garden Flowers: Insect Visits and the Floral Reward of Horticulturally-modifed Variants. *Annals of Botany* **83**, 73–86.

Corbet, S.A., 2003. Nectar sugar content: estimating standing crop and secretion rate in the field. *Apidologie* **34**, 1–10.

Corbet, S.A., Unwin, D.M., Prŷs-Jones, O.E., 1979. Humidity, nectar and insect visits to flowers, with special reference to *Crataegus*, *Tilia* and *Echium*. *Ecological Entomology* **4**, 9–22.

Couvillon, M.J., Riddell Pearce, F.C., Accleton, C., Fensome, K.A., Quah, S.K.L., Taylor, E.L., Ratnieks, F.L.W., 2015. Honey bee foraging distance depends on month and forage type. *Apidologie* **46**, 61–70.

Couvillon, M.J., Schurch, R., Ratnieks, F.L.W., 2014. Waggle dance distances as integrative indicators of seasonal foraging challenges. *PloS ONE* **9**.

Crane, E., 1976. *Honey: a comprehensive survey*. Heinemann, London.

Danner, N., Keller, A., Härtel, S., Steffan-Dewenter, I., 2017. Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected

pollen. *PLoS ONE* **12**, 9–12.

de Vere, N., Jones, L.E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S., Ford, C.R., 2017. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports* **7**, 42838.

Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L., Kartzinel, T.R., Eveson, J.P., 2019. Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology* **28**, 391–406.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* **26**, 5872–5895.

Dixon, P., 2003. Computer program review VEGAN , a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927–930.

Eyer, M., Neumann, P., Dietemann, V., 2016. A look into the cell: Honey storage in honey bees, *Apis mellifera*. *PLoS ONE* **11**, 1–20.

Fewell, J.H., Winston, M.L., 1992. Colony state and regulation of pollen foraging in the honey bee, *Apis mellifera* L. *Behavioral Ecology and Sociobiology* **30**, 387–393.

Fewell, J.H., Winston, M.L., 1996. Regulation of nectar collection in relation to honey storage levels by honey bees, *Apis mellifera*. *Behavioral Ecology* **7**, 286–291.

Fowler, R.E., Rotheray, E.L., Goulson, D., 2016. Floral abundance and resource quality influence pollinator choice. *Insect Conservation and Diversity* **9**, 481–494.

Galimberti, A., De Mattia, F., Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., Casiraghi, M., Labra, M., 2014. A DNA Barcoding Approach to Characterize Pollen Collected by Honeybees. *PloS ONE* **9**.

Garbuzov, M., Ratnieks, F.L.W., 2014. Ivy: An underappreciated key resource to flower-visiting insects in autumn. *Insect Conservation and Diversity* **7**, 91–102.

Garbuzov, M., Samuelson, E.E.W., Ratnieks, F.L.W., 2015. Survey of insect visitation of ornamental flowers in Southover Grange garden, Lewes, UK. *Insect Science* **22**, 700–705.

Goulson, D., Nicholls, E., Botias, C., Rotheray, E.L., 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **347**.

Gyan, K.Y., Woodell, S.R.J., 1987. Flowering Phenology, Flower Colour and Mode of Reproduction of *Prunus spinosa* L. (Blackthorn); *Crataegus monogyna* Jacq. (Hawthorn); *Rosa canina* L. (Dog Rose); and *Rubus fruticosus* L. (Bramble) in Oxfordshire, England. *Functional Ecology* **1**, 261.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., Baillie, L., Adams-Groom, B., 2015. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. *PLOS ONE* **10**.

Howes, F.N., 1945. *Plants and Beekeeping*. Faber & Faber, London.

Hung, K.L.J., Kingston, J.M., Albrecht, M., Holway, D.A., Kohn, J.R., 2018. The worldwide importance of honey bees as pollinators in natural habitats. *Proceedings of the Royal Society B: Biological Sciences* **285**.

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences* **274**, 303–313.

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G., Taylor, M.I., 2019. How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* **28**, 420–430.

Lucas, A., Bodger, O., Brosi, B.J., Ford, C.R., Forman, D.W., Greig, C., Hegarty, M., Jones, L., Neyland, P.J., De Vere, N., 2018. Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding. *Scientific Reports* **8**.

Maurizio, A., Hodges, F.E.D., 1950. The Influence of Pollen Feeding and Brood Rearing on the Length of Life and Physiological Condition of the Honeybee Preliminary Report. *Bee World* **31**, 9–12.

Naug, D., 2009. Nutritional stress due to habitat loss may explain recent honeybee colony collapses. *Biological Conservation* **142**, 2369–2372.

Nicolson, S.W., Nepi, M., Pacini, E., 2007. *Nectaries and Nectar*. Springer.

Percival, M., 1947. Pollen collection by *Apis mellifera*. *New Phytologist* **46**, 142–165.

Pernal, S.F., Currie, R.W., 2001. The influence of pollen quality on foraging behavior in honeybees (*Apis mellifera* L.). *Behavioral Ecology and Sociobiology* **51**, 53–68.

Potter, C., de Vere, N., Jones, L., Ford, C.R., Hegarty, M.J., Hodder, K.H., Diaz, A., Franklin, E.L., 2019. Pollen metabarcoding reveals broad and species-specific resource use by urban bees. *PeerJ* **7**.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25**, 345–353.

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., Vanbergen, A.J., 2016. Safeguarding pollinators and their values to human well-being. *Nature* **540**, 220–229.

Preston, C.D., Pearman, D.A., Dines, T.D., 2002. *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands*. Oxford University Press, Oxford.

R Development Core Team, R., 2011. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

Requier, F., Odoux, J.-F., Tamic, T., Moreau, N., Henry, M., Decourtye, A., Bretagnolle, V., 2015. Honey bee diet in intensive farmland habitats reveals an unexpectedly high flower richness and a major role of weeds. *Ecological Applications* **25**, 881–890.

Richardson, R.T., Curtis, H.R., Matcham, E.G., Hua Lin, C., Suresh, S., Sponsler, D.B., Hearon, L.E., Johnson, R.M., 2018. Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. *Molecular Ecology* **28**, 686–697.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015. Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. *Applications in Plant Sciences* **3**.

Robinson, R.A., Sutherland, W.J., 2002. Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology* **39**, 157–176.

Salisbury, A., Armitage, J., Bostock, H., Perry, J., Tatchell, M., Thompson, K., 2015. Enhancing gardens as habitats for flower-visiting aerial insects (pollinators): should we plant native or exotic species? *Journal of Applied Ecology* **52**, 1156–1164.

Samnegård, U., Persson, A.S., Smith, H.G., 2011. Gardens benefit bees and enhance pollination in intensively managed farmland. *Biological Conservation* **144**, 2602–2606.

Scofield, H.N., Mattila, H.R., 2015. Honey bee workers that are pollen stressed as larvae become poor foragers and waggle dancers as adults. *PloS ONE* **10**.

Seeley, T.D., 1995. *The Wisdom of the Hive*. Harvard University Press.

Stace, C., 2010. New Flora of the British Isles, Third. ed. Cambridge University Press.

Standifer, L.N., 1967. A comparison of the protein quality of pollens for growth-stimulation of the hypopharyngeal glands and longevity of honey bees, *Apis mellifera* L. (Hymenoptera: Apidae). *Insectes Sociaux* **14**, 415–425.

Steffan-Dewenter, I., Kuhn, A., 2003. Honeybee foraging in differentially structured landscapes. *Proceedings of the Royal Society B: Biological Sciences* **270**, 569–575.

Synge, A.D., 1947. Pollen collection by honeybees (*Apis mellifera*). *Journal of Animal Ecology* **16**, 122–138.

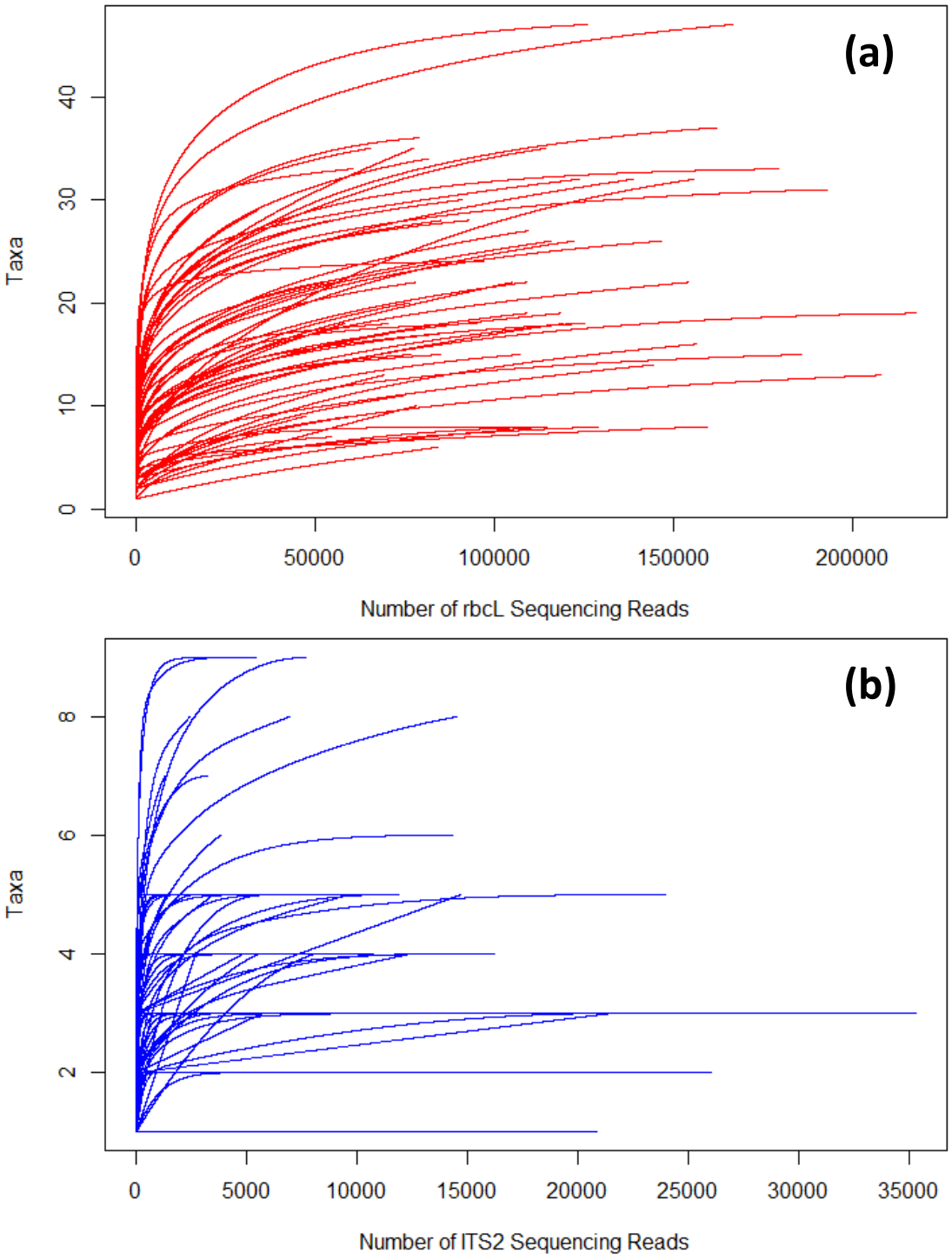
Von Der Ohe, W., Persano Oddo, L., Piana, M.L., Morlot, M., Martin, P., 2004. Harmonized methods of melissopalynology. *Apidologie* **35**, S18–S25.

Wang, Y., Naumann, U., Wright, S.T., Warton, D.I., 2012. mvabund - an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* **3**, 471–474.

Wright, G.A., Nicolson, S.W., Shafir, S., 2018. Nutritional Physiology and Ecology of Honey Bees. *Annual Review of Entomology* **63**, 327–344.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., De Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., Kauserud, H., Orlando, L., Pansu, J., Pawlowski, J., Tedersoo, L., Thomsen, P.F., Willerslev, E., Taberlet, P., 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* **28**, 1857–1862.

3.7. Supplementary Figures



Supplementary Figure 1: Rarefaction curves for **(a)** *rbcL* and **(b)** ITS2. The curves indicate a plateau for both markers being reached with the sequencing depth achieved.

CHAPTER 4

Agricultural intensification, shifts in crop use and invasive species induce nationwide change in honeybee foraging

4.1. Abstract

Decreasing floral resources as a result of habitat loss and fragmentation is one of the key factors in the decline of pollinating insects worldwide, with the resulting impact on food supply and biodiversity receiving increasing global concern. Understanding which plants pollinators use on a wide scale is vital to inform the provision of appropriate floral resources to mitigate pollinator declines. Here we show that post 1950 shifts in floral resources at the UK landscape scale are manifested in honeybee forage, demonstrating anthropogenic forcing of change in the foraging behaviour of a globally important pollinator. Using DNA metabarcoding, we analysed 441 UK honey samples from 2017 to assess the most important floral resources used by honeybees. We found 157 plant taxa, with only four taxa found in over 50% of samples: *Rubus* spp., *Trifolium repens*, *Brassica* spp. and the Maleae tribe, including *Crataegus* spp., *Malus* spp. and *Cotoneaster* spp. Comparing with a nationwide survey of 855 honey samples in 1952, we find a decline in the use of *Trifolium repens* and fruit trees and an increase in use of *Rubus* spp., *Brassica* spp. and the non-native invasive *Impatiens glandulifera*. The changes reflect the availability of forage in the landscape, with shifts coming from the intensification of agriculture such as the reduction in clover leys, changes in crop cultivation with the increase in oilseed rape and the spread of the invasive species *Impatiens glandulifera*. Improved grasslands are the most widespread habitat type in the UK and so changes in management of these areas have the potential to greatly influence floral resource availability, with concomitant benefits for honeybee forage availability and pollinator health.

4.2. Introduction

The decline in both wild and managed pollinators has received considerable concern (IPBES, 2016; Potts et al., 2010a, 2010b; Vanbergen, 2013), with one of the key drivers of decreases being the loss and fragmentation of habitat from land use change (Carvell et al., 2006; Ollerton et al., 2014). The loss of temperate floral resources has been influenced by many factors, including agricultural intensification in the form of increased application of pesticides, herbicides and inorganic nitrogen fertilizers, the removal of hedgerows, and the decline in semi-natural, species rich grasslands (Green, 1990; Petit et al., 2003; Robinson and Sutherland, 2002; Stevens et al., 2003).

Across Europe the dominant land use is agriculture, with over half of the European landscape being agriculturally managed (Batáry et al., 2015). It is therefore a focus of conservation efforts to prevent the loss of associated biodiversity (Batáry et al., 2015; Benton et al., 2003). Post-war farming practices have seen increasing intensification, with a corresponding decrease in landscape diversity (Robinson and Sutherland, 2002). The increasing use of machinery has resulted in larger farms, with 1% of farms covering greater than 200 ha in 1949, compared with over 6% in 1999 (Robinson and Sutherland, 2002). To allow access for this increasing mechanisation and to increase the size of fields, widespread hedgerow removal started from the 1960s (Barr and Gillespie, 2000; Robinson and Sutherland, 2002), with a significant decline in hedgerows between 1984 and 1990 (Petit et al., 2003).

Agricultural intensification has also seen the increased application of agrochemicals to improve yields. Since the 19th century, crop rotation and tillage was used to maintain the fertility of the land and control weeds, by rotating between cereal crops and non-cereal crops including legumes (Robinson and Sutherland, 2002). By using herbicides instead, weeds can be killed before the crop is sown. As such, a continuous cereal crop can be maintained, with a decrease in the occurrence of grass leys and fallow fields which allow the presence of non-crop plants (Robinson and Sutherland, 2002). There has been an increase in the dominance of crop monocultures including, since the late 1960s, insect attractive crops such as oilseed rape (*Brassica napus*) (Chamberlain et al., 2000; Schürch et al., 2015; Sutcliffe and Kay, 2000).

High fertiliser input, ploughing, reseeding, high levels of grazing, and herbicide application can all cause species-rich, semi-natural grassland to become improved grassland, with a corresponding reduction in the diversity and availability of forbs (French, 2017). In England and Wales, the proportion of semi-natural grassland present in lowland grasslands was estimated to be only 3% of what was present prior to the war (Fuller, 1987). A major change in post-war management has been the shift from using hay meadows to grass grown for silage. A move to silage production began in the 1960s and by the 1980s was the main form of grass feed production (Chamberlain et al., 2000).

These landscape changes have been monitored in the UK using the Countryside Survey, first undertaken in 1978, with the survey plots revisited in 1990, 1998 and 2007 (Wood et al., 2017). The resulting data has been used to assess national scale changes, looking at aspects such as the declines in forage availability for bumblebees (Carvell et al., 2006) and estimating the changes in hedgerows and field boundaries (Barr and Gillespie, 2000; Petit et al., 2003). In Baude et al., (2016), the overall loss of floral resource since the 1940s was quantified using land cover data, Countryside Survey data, and estimates of nectar production, which together showed the national decline in nectar resource during the 20th century. Baude et al., (2016) found that nectar levels have reduced by 32% between 1930 and 1978, with increases in the past 30 years. However, the overall nectar resource was found to still be lower than the levels before the 1930s.

The honeybee, *Apis mellifera*, is the most commonly managed bee in the world, contributing to the pollination of crops and wild plants, as well as providing honey, propolis and wax products (Potts et al., 2016). Honeybees are well suited to monitoring nectar resource at a landscape scale, for several reasons. They can provide information from a sizeable area due to their large foraging range and their ability to recruit foragers to good food sources (Seeley, 1995; Visscher and Seeley, 1982). As a managed species, in hives with a known location, they are widespread throughout the UK and can be easily sampled, when compared with finding and sampling wild pollinators. The location of the hive provides information on the starting point of foraging. Furthermore, by using honey as a source of information on the plants used, it captures the forage effort of the colony over a long time period, when compared to the pollen from the body of an individual foraging insect.

Traditionally, melissopalynology has been used to characterise the botanical and geographical origin of honey, vital to the regulation of a food product, by using the morphological identification of pollen from the honey under a light microscope (Louveaux et al., 1978; Von Der Ohe et al., 2004). In 1952, 855 honey samples from across the UK and Ireland were characterised using melissopalynology, representing the last time UK honey was examined on such a scale (Deans, 1958, 1957). More recently, DNA metabarcoding has been used to identify species from pollen, a technique which uses high throughput sequencing of amplified regions of DNA from mixed sources. Using honey, DNA metabarcoding has been used to answer ecological questions on the foraging preferences of honeybees as well as a potential method of food regulation (Chapter 3, de Vere et al., 2017; Prosser and Hebert, 2017; Smart et al., 2017). The identification abilities of DNA metabarcoding are reliant on the availability of a high quality, comprehensive reference library for species within the study area. To facilitate this we use a high-quality DNA barcode reference library that includes comprehensive coverage of the UK native flora supplemented with curated sequences for the UK naturalised and horticultural species (de Vere et al., 2012, Chapter 2).

Nectar and pollen supply most of the honeybee's nutritional requirements, with phenological gaps in forage resource resulting in negative health impacts for the colony (Requier et al., 2017). Nectar provides the honeybees with the carbohydrates they need for high energy activities such as flying; while pollen provides a source of protein, fats, sterols and micronutrients (Wright et al., 2018). The protein found in pollen is essential to the healthy growth and development of the larvae during brood rearing (Brodschneider and Crailsheim, 2010). In honeybees, the foraged nectar can also be processed into honey and stored by reducing the water content and hydrolysing the sucrose through the addition of invertase (Eyer et al., 2016). Pollen is also stored within the hives. These stores provide the bees with sources of food when foraging is not possible, such as periods of inclement weather and during winter. For beekeepers, the surplus honey produced by the honeybees is extracted from the hives, and during periods of high productivity supers containing frames will be placed on the hive for the honeybees to fill with honey stores and to prevent swarming (Jones et al., 2011; Manley, 1936).

Here, we characterise the floral source of 441 honey samples from beekeepers across the UK, using DNA metabarcoding with two plant DNA barcode regions, *rbcL* and ITS2. This represents the first UK wide survey of honey samples since 1952. We examine the relationship between the floral composition of the plants in the honey with the location and the time the honey was removed from the hives. By comparing between the major sources of forage found in this UK survey in 2017 and the results of the UK survey from 1952, we can examine if the landscape scale changes in floral resource has led to a change in honeybee foraging.

4.2.1. Aims and objectives

- Characterise the plants honeybees are using across the UK using DNA metabarcoding
    - Are there spatial and temporal patterns to honeybee foraging within the UK?
    - How does the presence of crop species within foraging distance of the hive relate to the floral composition of the honey?
    - Have landscape scale changes in forage availability since the early 20th century affected the use of key honeybee forage plants in the UK?

4.3. Method and Analysis

4.3.1. Honey sampling and DNA extraction

Beekeepers across the UK were asked to supply approximately 30 ml of honey from any date in 2017, reporting the date of sample collection and the location of the apiary, using a grid reference or postcode. The call for samples was publicised through a national gardening TV show, *Gardeners' World,* broadcast in July 2017. Additional recruitment to the survey included 14 talks given at different beekeeping associations in Wales and England, including one at the National Honey Show, and an article in the UK beekeeping magazine, *BeeCraft* (Appendix A). As part of the survey, the individual and overall results will be fed back to beekeepers.

Any wax was removed using sterile forceps and DNA was extracted from 10 g of honey using a modified version of the DNeasy 96 Plant extraction kit protocol (Qiagen). Firstly, the 10 g of honey was made up to 30 ml with molecular biology grade water and incubated in a water bath at 65 °C for 30 minutes. Samples were then centrifuged (Sorvall RC-5B) for 30 minutes at 15,000 rpm, the supernatant was discarded, and the pellet resuspended in 400 µL of buffer made from a mix of 400 µL AP1 from the DNeasy Plant 96 Kit (Qiagen), 80 µL proteinase K (1 mg/ml) (Sigma) and 1 µL RNase A (Qiagen). This was incubated again for 60 minutes at 65 °C in a water bath and then disrupted using a TissueLyser II (Qiagen) for 4 minutes at 30 Hz with 3 mm tungsten carbide beads. The remaining steps were carried out according to the manufacturer's protocol. The extracted DNA was purified using the OneStep PCR Inhibitor Removal Kit (Zymo Research) and diluted 1 in 10.

4.3.2. PCR and library preparation

Illumina MiSeq paired end indexed amplicon libraries were created via a two-step PCR protocol. Two libraries were prepared for the DNA barcode regions, *rbcL* and ITS2. Initial amplification used the template specific primers *rbcL*af and *rbcLr*506, and ITS2F and ITS3R, with universal tails designed to attach custom indices in the second round PCR. To improve clustering on the Illumina MiSeq, a 6N sequence was also added between the forward template specific primer and the universal tail.

Forward universal tail, 6N sequence and *rbcL*af:

[**ACACTCTTTCCCTACACGACGCTCTTCCGATCT**]NNNNNN[ATGTCACCACAAACAGAGACTAAAGC]

Reverse universal tail and *rbcL*r506:

[**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**][AGGGGACGACCATACTTGTTCA]

Forward universal tail, 6N sequence and ITS2F:

[**ACACTCTTTCCCTACACGACGCTCTTCCGATCT**]NNNNNN[ATGCGATACTTGGTGTGAAT]

Reverse universal tail and ITS3R:

[**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**][GACGCTTCTCCAGACTACAAT]

This first PCR used a final volume of 20 µl: 2 µl template DNA, 10 µl of 2x Phusion Hot Start II High-Fidelity Mastermix (Thermo Fisher Scientific), 0.4 µl (2.5 µM) forward and reverse primers, and 7.2 µl of PCR grade water. Thermal cycling conditions for *rbcL* were: 98 °C for 3 min, 95 °C for 2 minutes; 95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 40 seconds (40 cycles); 72 °C for 5 minutes, 30 °C for 10 seconds. Thermal cycling conditions for the first ITS2 PCR were: 98 °C for 3 min, 94 °C for 5 minutes; 94 °C for 30 seconds, 56 °C for 30 seconds, 72 °C for 40 seconds (40 cycles); 72 °C for 10 minutes, 30 °C for 1 minute. The initial PCR was carried out three times and pooled.

The pooled products from the first PCR were purified following Ilumina's 16S Metagenomic Sequencing Library Preparation protocol using Agencourt AMPure XP beads (Beckman Coulter). The purified PCR product from round one was followed by a second round of amplification to anneal custom unique and identical i5 and i7 indices to each sample (Ultramer, IDT). This index PCR stage used a final volume of 25 µl reaction (12.5 µl of 2x Phusion Hot Start II High-Fidelity Mastermix, 1 µl of i7 Index Primer and i5 Index Primer, 6.5 µl of PCR grade water, and 5 µl of purified first-round PCR product). Thermal cycling conditions were: 98 °C for 3 min; 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s (8 cycles); 72 °C for 5 min, 4 °C for 10 min. Following the index PCR, a 1% gel was run to confirm success. The index PCR product was then purified following the PCR clean-up 2 section of the Illumina protocol. The purified products of the index PCR were quantified using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and pooled at equal

concentrations to produce the final library. Positive and negative controls were amplified and sequenced alongside honey samples. Sequence data will be made available at the NCBI Sequence Read Archive (SRA).

4.3.3. Positive control

A positive control was made from a mixture of five individual tropical tree species which had previously been barcoded using *rbcL* and were not present in the survey site. The species *Baccaurea stipulata*, *Colona serratifolia*., *Dillenia excelsa*, *Kleinhovia hospita*, and *Pterospermum macrocarpum* were used, taking 5 µl from each separate DNA extraction and mixing, before following the protocol as with the honey samples.

4.3.4. Bioinformatic analysis

Sequence data was processed using the modified data analysis pipeline first developed in de Vere *et al.*, (2017) (https://github.com/colford/nbgw-plant-illumina-pipeline). Raw reads were trimmed to remove low quality regions (Trimmomatic v. 0.33), paired, and then merged (FLASH v. 1.2.11), with merged reads shorter than 450 bp discarded. Identical reads were dereplicated within samples and then clustered at 100% identity across all samples (vsearch v. 2.3.2), with singletons (sequence reads that occurred only once across all samples) discarded.

A custom reference database was created for sequence identification, representing 5,586 unique plant species. This UK species list was generated using the list of native species of the UK from Stace (2010), 505 naturalised alien species (BSBI), and horticultural species from the IRIS BG database at the National Botanic Garden of Wales. Using the DNA barcoding resource for the UK provided 98% coverage for the native flowering plants and conifers of the UK (de Vere et al., 2012, Chapter 2). In the *rbcL* reference database, species level coverage for the 5,586 plant species was 57%, and coverage at genus level was 96%. In the ITS2 reference database species level coverage was 52% and genus coverage was 84%.

The sequence data from the honey samples were compared against the reference database using blastn, using the script vsearch-pipe.py. The top twenty BLAST hits were then summarised using the script vsearch_blast_summary.py. If the top bit scores of a

sequence matched to a single species, then the sequence was identified to that species. If the top bit scores matched to different species with the same genus, then the result was attributed to the genus level. If the top bit score belonged to multiple genera within the same family then a family level designation was made. Sequences that returned families from different clades were considered to be chimeric and excluded. These automated identifications were then checked manually for botanical veracity.

To investigate the most frequently used honey plants of the UK and compare them with the previous UK-wide survey from 1952, frequency classes were used, using the designations from traditional melissopalynology and what was reported in Deans, (1958, 1957). . The frequency classes were assigned for each honey sample based on the percentage of reads returned for the two DNA regions *rbcL* and ITS2. Plant taxa represented by over 45% of reads were designated *predominant* for that sample; between 15% and 45% were *secondary*; between 1-15% were *important minor* taxa, and less than 1% of reads were classed as *minor* taxa.

4.3.5. Methodological changes from Chapter 3

In Chapter 3, the *rbcL* plant DNA barcode was used alone to characterise the honey, as the ITS2 data did not provide any further taxonomic discrimination or representation compared to what was found using *rbcL* alone. With the UK honey samples, ITS2 did provide additional plant taxa which were not detected by *rbcL,* and so the sequencing data from both markers was included in further analysis. In Chapter 3, the relative abundance of reads from the DNA metabarcoding was used to provide a semi-quantitative measure of foraging and track changes throughout the season, using honey samples collected directly from the hive each month. Here, to provide a robust comparison between the two survey methods of melissopalynology and DNA metabarcoding, frequency classes are used. Further discussion on the results and methodological differences between Chapter 3 and Chapter 4 is provided in Chapter 5.

4.3.6. 1952 Honey Sampling

In 1952, 855 honey samples were characterised from 66 counties across the UK and Ireland using melissopalynology (Deans, 1958, 1957). Samples were obtained via a general appeal and were all collected during the honey season of 1952. For each honey sample, 10 g was sampled, from

which approximately 200 pollen grains were identified using the morphology of the pollen under the microscope, following the protocol established in Maurizio & Hodges, (1951).Each taxon found in the sampled honey was reported in frequency classes, common to melissopalynology analysis: *predominant* (>45% of pollen grains), *secondary* (15-45% of pollen grains) and *important minor* (1-15% of pollen grains). Those grains classified as *minor* (less than 1% of pollen grains) were not reported. The location data for the honey samples are restricted to the county level, and summary data tables were presented for each UK county that returned honey. The number of times each taxon occurred at each level of abundance was then calculated from the county results, with the sum of this giving the frequency of occurrence across all the UK samples. In order to compare taxa between the more taxonomically resolved 2017 DNA and traditional 1952 honey surveys, the pollen taxonomic classification was changed within the DNA results to reflect the grouping achieved with the morphological identification of the pollen. The relationship between the frequency of occurrence for the matched plant taxa between 1952 and 2017 was assessed using Kendall's rank correlation.

The change in proportion of predominant and secondary forage between 1952 and 2017 was examined for the plant taxa that occurred as predominant and secondary forage in more than 1% of samples for both honey surveys. Chi-squared contingency tests were used to assess differences, with Bonferroni correction for multiple testing. All statistical analyses were carried out using R (v. 3.5.2) (R Development Core Team, 2011).

4.3.7. Countryside Survey vegetation plot data frequency changes

Changes in the local frequency of the predominant and secondary plant forage species found in both 1952 and 2017 were assessed using the Countryside Survey data from 1978 and 2007 (Wood et al., 2017). In 1978, the survey looked at 256 1 km squares within which fixed plots were established, representing fields and unenclosed land (200 m^2) as well as linear features including hedgerows, streams and roadsides (10 m^2). In each plot, a list of all vascular plants was recorded. Where possible, squares and plots were then revisited in 2007, representing 236 1 km squares containing 1,577 plots. For these revisited plots, the percentage change in plot frequency for the predominantly found forage species was calculated.

4.3.8. Landscape data

The Land Cover 2017 map was used to characterise habitat in a 2 km radius of the hives while the 2017 CEH Land Cover Plus: Crops map was used to assess the presence and absence of crop species (oilseed rape and field beans) within a 2 km radius of each hive (Rowland et al., 2017). A chi-squared contingency test was used to look at the differences between the presence of the crop species in the honey and the presence and absence of the crop within the landscape. Analyses and maps were generated in R (v. 3.5.2) (R Development Core Team, 2011).

4.3.9. Statistical analysis of the DNA metabarcoding data

To understand how the plant taxa composition within the 2017 honey samples was structured in space and time, the effect of time (measured as the calendar month number in 2017) along with the latitude and the longitude of sampling location were included in a two-tailed generalized linear model using the 'manyglm' function in the package *mvabund* (Wang et al., 2012). Using a negative binomial distribution, the effect of month, latitude and longitude were included as explanatory variables in the model and the proportion of sequences was set as the response variable (proportion data has been previously demonstrated to be an effective method for controlling differences in sequence read numbers (McMurdie and Holmes, 2014)). The proportion of sequences was scaled by 1000 and values were converted to integers so that a negative binomial distribution could be used. The strong mean-variance relationship in the data (Supplementary Figure 2a), supports the use of a negative binomial distribution in the model. The appropriateness of the models was checked by visual inspection of the residuals against predicted values from the models (Supplementary Figure 2b). Analyses were completed in R (v.3.5.2).

4.4. Results

4.4.1. 2017 honey sampling

A total of 441 honey samples were successfully processed with the two markers, *rbcL* and ITS2 from honey samples from beekeepers across the UK. The samples were collected from April to October, with the majority of sampling occurring during July and August (Figure 18). Improved grassland, arable and horticulture made up 56% of the habitat in the locality of the hives (Figure 18). Differences between the land cover of the UK and within the locality of the hive were found, illustrating beekeepers setting their hives in accessible, suitable locations, with suburban and urban areas making up 24% of the land within 2 km, compared with 7% nationally, and bog representing 0.2% of the area around hives compared with 4% nationally. For the countries of the UK, 319 samples were returned from England, 21 from Scotland, 84 from Wales and three from Northern Ireland.

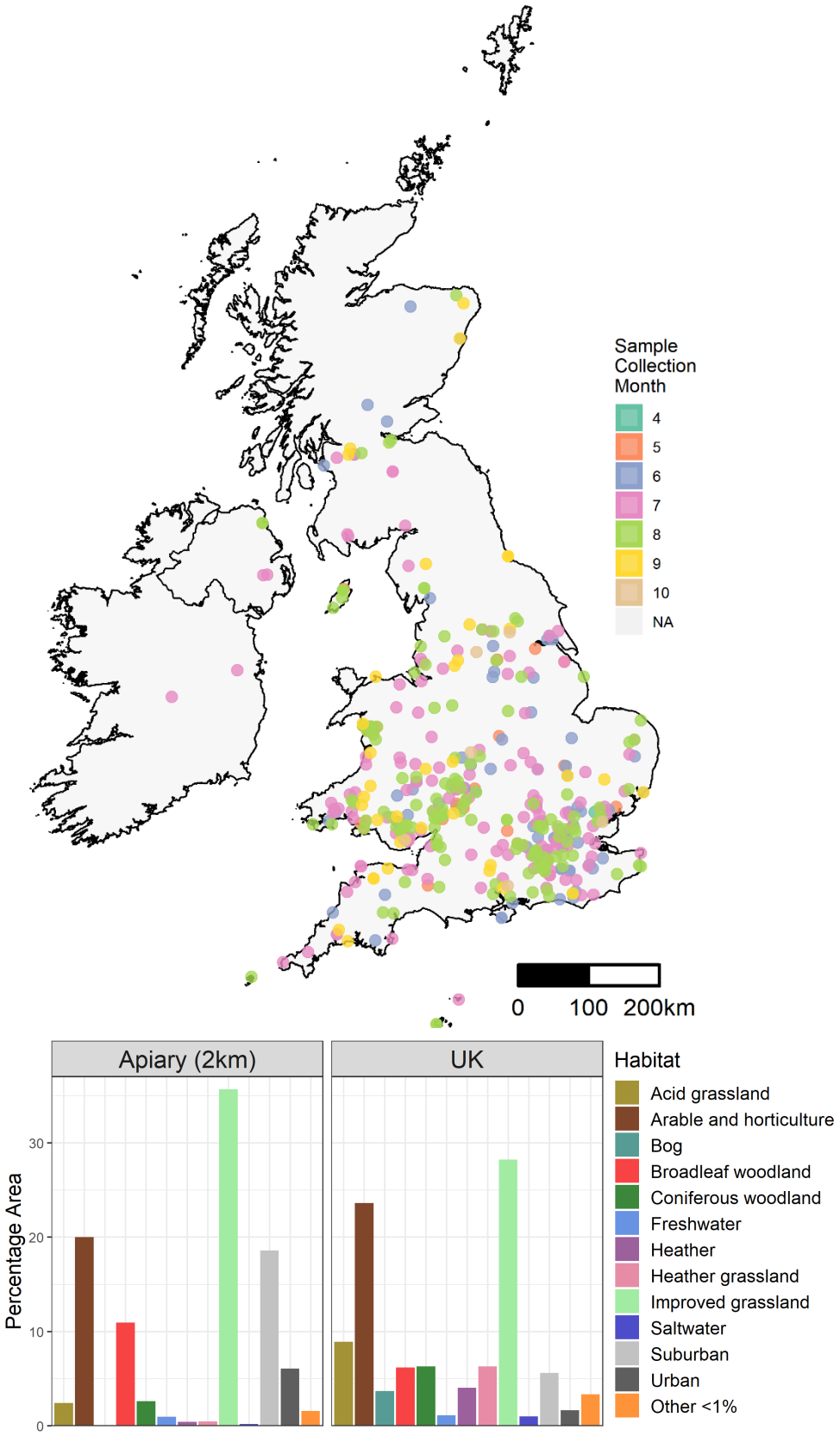


Figure 18: Distribution of sampled honey (n = 441). Colour indicates the month the honey was collected in 2017. The percentage area of habitats is presented for the UK overall and for the subset of a 2 km radius around the honey samples located in Great Britain (n = 424), characterised using the 2015 CEH Land Cover map (NERC CEH).

4.4.2. 2017 honey survey DNA sequencing

Three sequencing runs were completed for 476 honey samples yielding a total of 39,958,235 returned read pairs. After the quality control, trim, pair and merge a total of 23,418,356 reads remained. Sequences were then clustered at 100% identity within and across all samples. Singletons were removed leaving 12,335,190 remaining sequences which entered the identification pipeline. During identification, 2,305,168 low quality sequences were discarded, leaving 10,050,022 sequences which entered analysis, 57% *rbcL* and 43% ITS2. Only honey samples which returned over 100 sequences for both *rbcL* and ITS2 went through for analysis, leaving 441 samples. The mean total sequence number returned for each sample was 22,176 (SD = 14,965) and ranged from 960 to 87,515.

During identification, 121 taxa were identified by *rbcL* and 84 were identified by ITS2. There were 48 overlapping taxa between the two lists. As both *rbcL* and ITS2 provided additional plant taxa information unique to the marker (73 from *rbcL* and 36 from ITS2), the taxa list for each sample was combined with percentages calculated from the total of *rbcL* and ITS2 sequence reads for each sample.

4.4.3. Plants found in the 2017 honey survey

Across the 2017 honey samples, 157 plant taxa were identified using the *rbcL* and ITS2 barcode regions combined. Of these, 21% were identified to species, 69% to genus, 1% to tribe and 8% to family levels. All five plants in the positive control were successfully returned. Only 44 of the plant taxa occurred in over 5% of the honey samples (Figure 19) and only four taxa were identified in over 50% of samples: *Rubus* spp., *Trifolium repens*, *Brassica* spp. and the Maleae tribe, including *Crataegus* spp., *Malus* spp. and *Cotoneaster* spp. Looking at the most frequent predominant and secondary forage (red and orange bars, Figure 19), *Rubus* spp., *T. repens* and *Brassica* spp. were prevalent. *Acer* spp. were the fourth most abundant taxa, followed by the Maleae tribe, *Impatiens glandulifera, Vicia* spp. *Ulex* spp., and *Calluna vulgaris*.

The most abundantly found plants were found across all months (April to October). Time, measured as the calendar month the honey sample was taken, was found to be a good predictor of plant taxa composition (Figure 19; $LR_{428, 1} = 397.2$, $p = 0.001$). The relative

proportion of taxa within each month tracked the flowering phenology. The spring flowering *Acer* spp., Maleae tribe, *Prunus* spp., *Taraxacum officinale, and Ulex spp.* were all found more in May and June honey samples, while late summer species such as *Calluna vulgaris* and *Impatiens glandulifera* were found more in August and September (Figure 19). The frequency of both *Rubus* spp. and *T. repens* peaked in July.

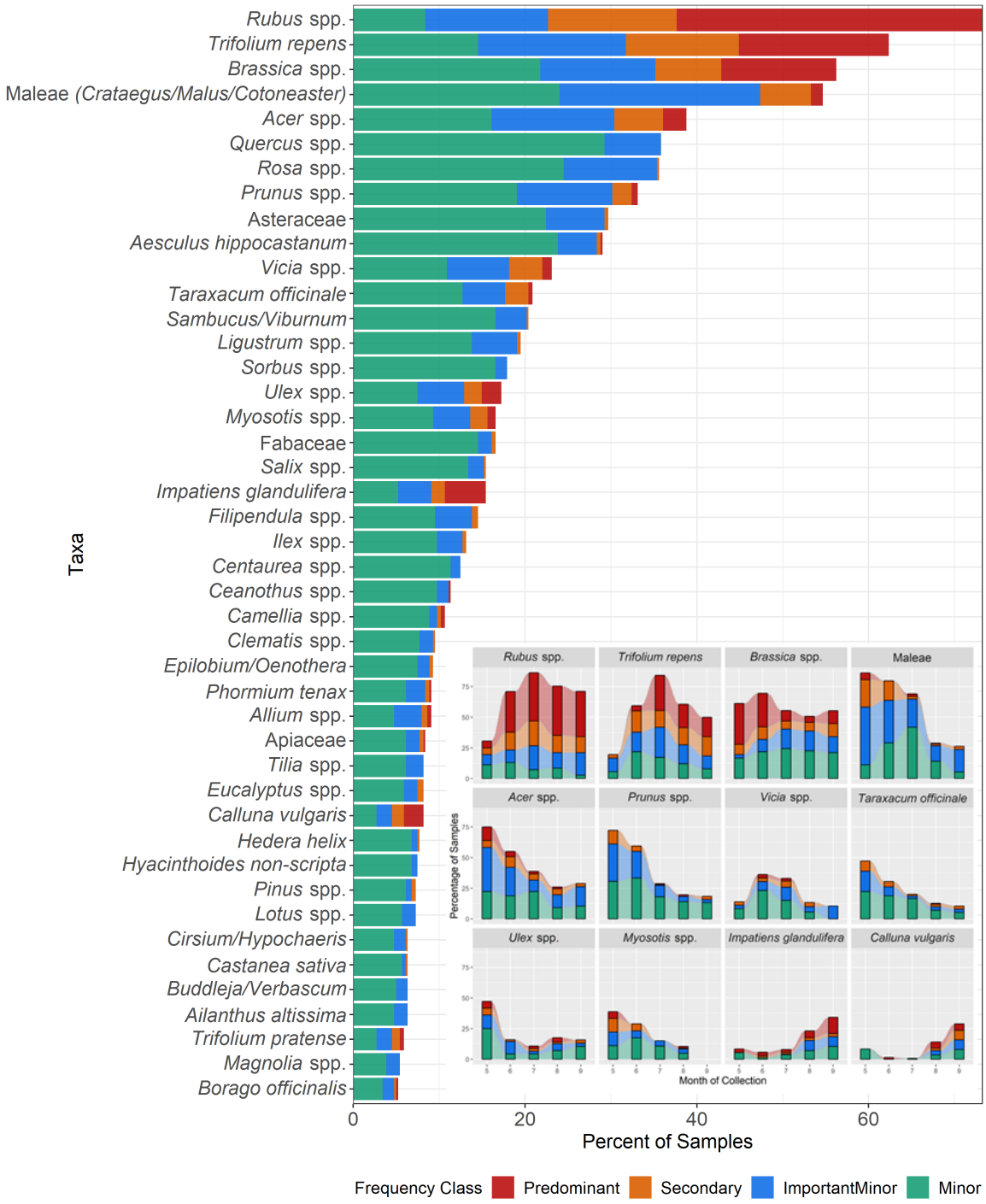


Figure 19: The 44 plant taxa found in over 5% of honey samples (n = 441). A further 113 taxa identified were present in less than 5% of samples. Colour indicates the proportion of samples classified as the different frequency classes. Predominant is over 45% of sequences returned in a sample, secondary is between 15-45%, important minor is between 1-15% and less than 1% of sequences is minor. **Inset:** the most frequent taxa found at a predominant and secondary level are summarised as a proportion of samples through the season. Samples collected in April (n = 3) and October (n = 7) were excluded. Sample sizes for the months were: May (n = 39), June (n = 71), July (n = 147), August (n = 155), September (n = 43).

4.4.4. Spatial patterns

No significant regional differences were found in the abundances of the plant taxa when examining the effect of latitude and longitude (Figure 20; Latitude $LR_{426, 1} = 220.1$, $P = 0.335$; Longitude $LR_{476, 1} = 341.3$, $P = 0.118$). However spatial patterns were detected in less frequently found taxa. *Ailanthus altissima*, a non-native tree species often planted along city streets, has a distribution associated with urban areas, with 21 of the samples located in or near Greater London and six from cities around the UK (Figure 21). This is reflected in the local land cover around the hives with *A. altissima* present, with 68% of the land within a 2 km radius being urban or suburban compared to 24% across all samples (Figure 21).

The presence of two plants, *Brassica* spp. and *Vicia* spp., was found to have a relationship with the presence of crop fields for oilseed rape (*Brassica napus*) and field beans (*Vicia faba)* respectively (Figure 22). The presence of *Brassica* spp. in the honey significantly differed between samples with oilseed rape present in a 2 km radius and those without ($x^2 = 50.71$, d.f. = 4, $p < 0.0001$), with an increase in *Brassica* spp. in the honey taken from hives with oilseed rape in the locality. Similarly, the present of *Vicia* spp. increased with the presence of field beans in a 2 km radius ($x^2 = 52.83$, d.f. = 4, $p < 0.0001$), with no honey samples returning *Vicia* spp. as predominant forage in areas without field beans present (Figure 22).

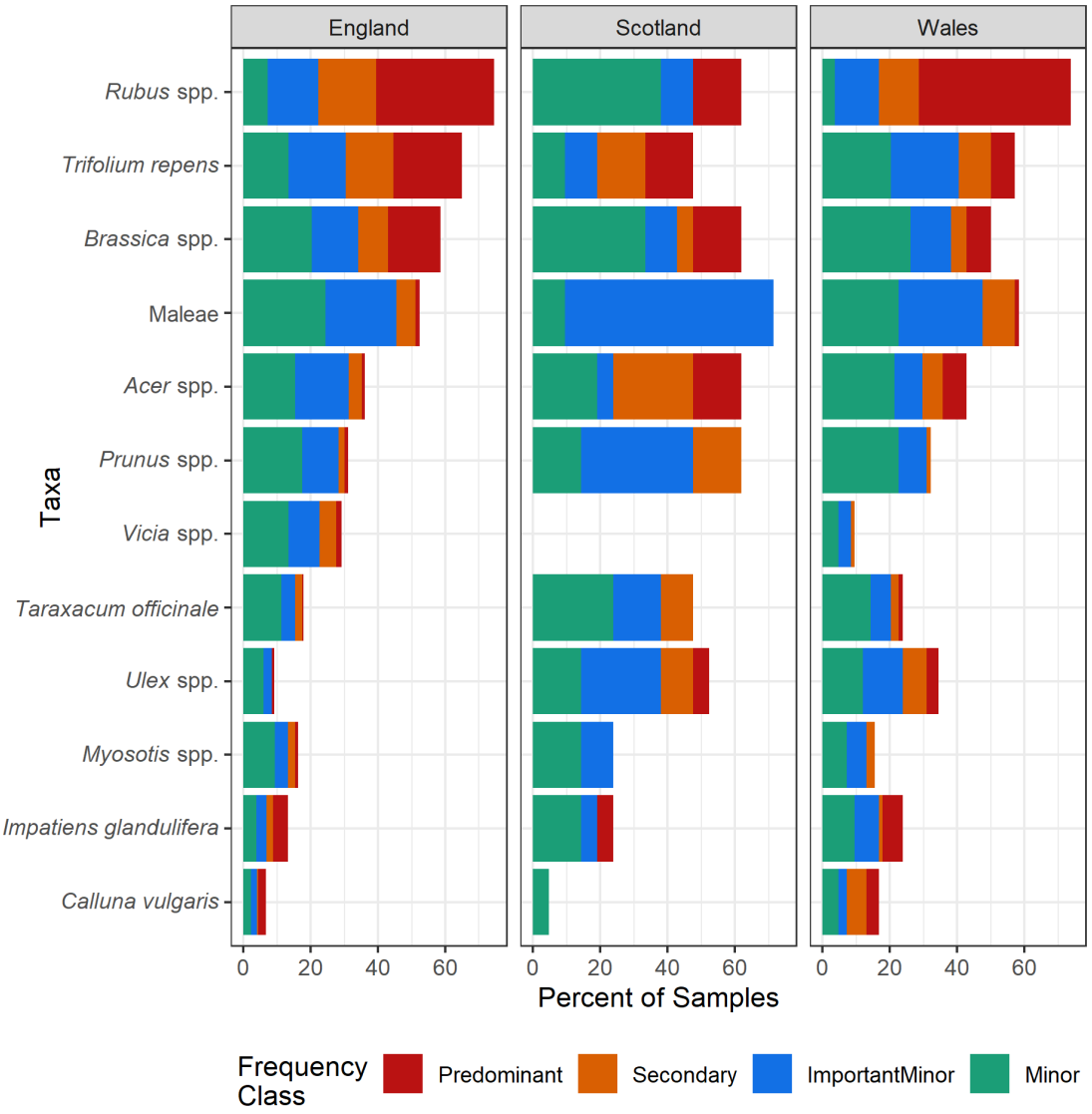


Figure 20: Top predominant and secondary taxa (>15% sequences) found in the DNA in the different regions, England (n = 319), Scotland (n = 21), and Wales (n = 84). Regions with smaller samples sizes were excluded: Ireland (n = 3), Northern Ireland (n = 4), Guernsey (n = 3), and the Isle of Man (n = 7).

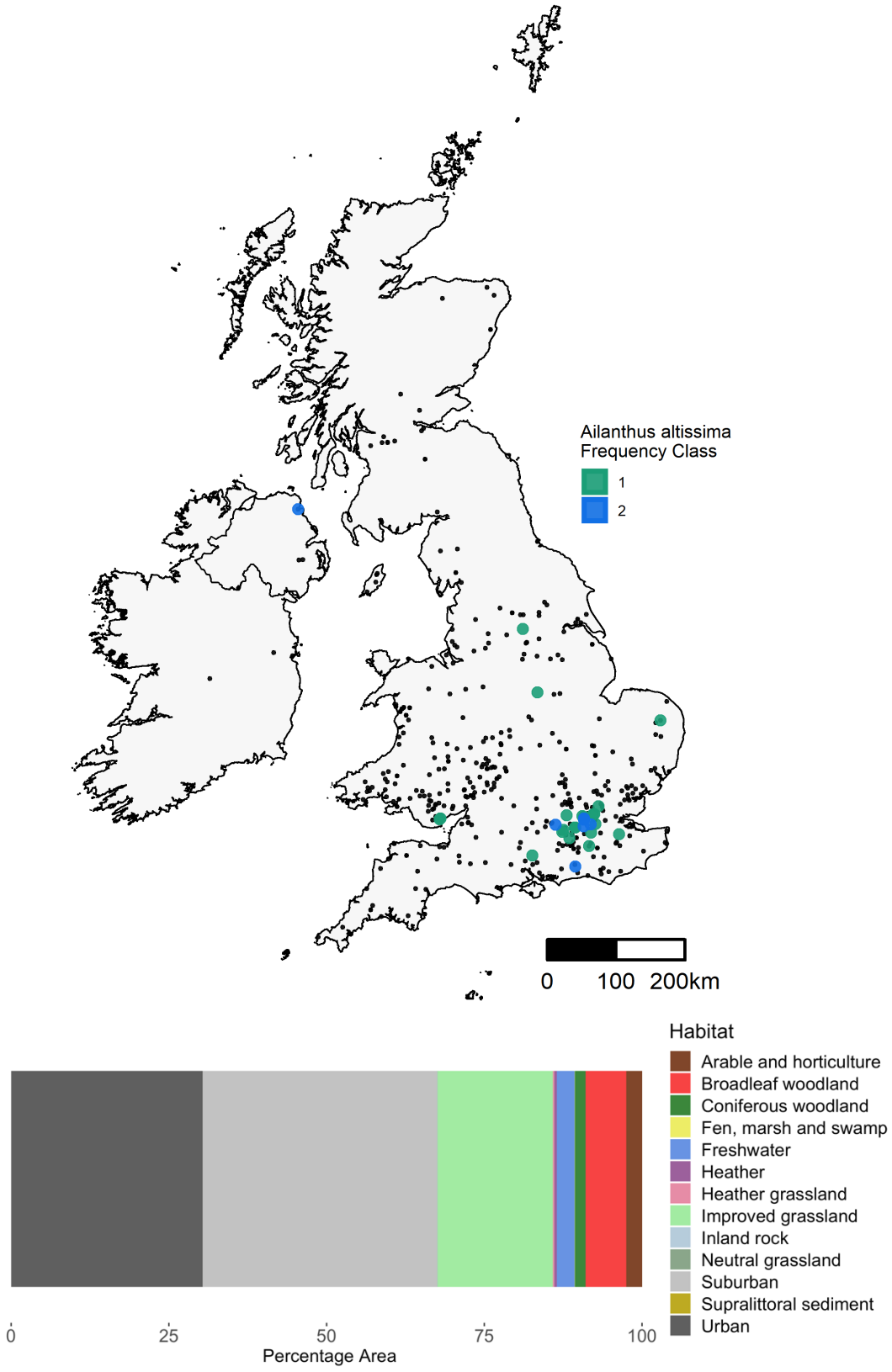


Figure 21: The distribution of honey samples containing *Ailanthus altissima*. Black dots indicate absence from a honey sample. The percentage habitat in a 2 km radius is summarised for honey samples with *A. altissima* present (n = 28) with 68% of the habitat around the hive classed as urban or suburban.

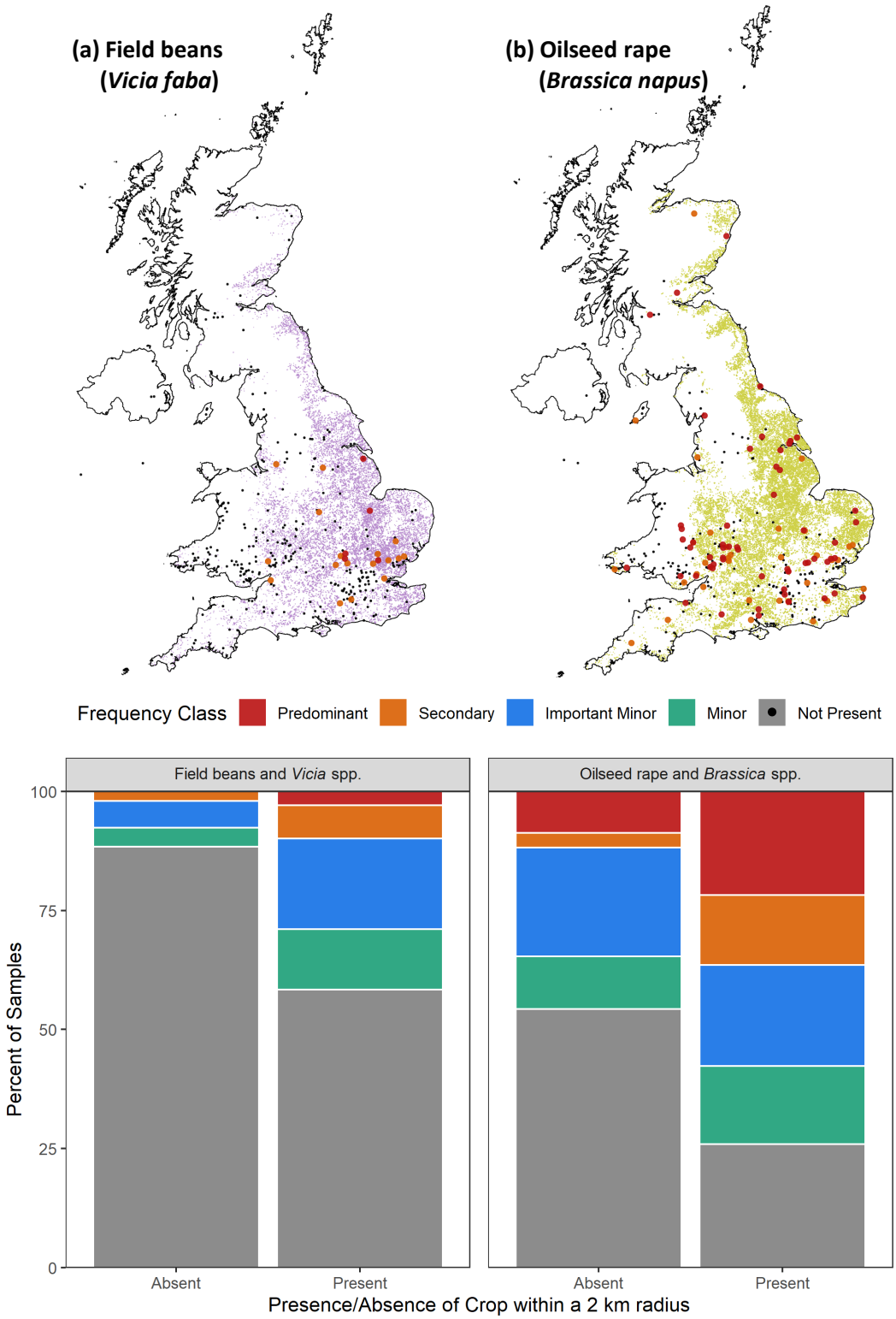


Figure 22: The locations of honey samples with *Vicia sp.* (a) and *Brassica* sp. (b) present at predominant (>45%) and secondary (>15%) level over the cover of oilseed rape and field beans in 2017 from CEH Land Cover plus: Crops map. The presence of *Brassica* in the honey differed between samples with oilseed rape present in a 2km radius and those without (x^2= 50.71, d.f. = 4, $p < 0.0001$). The same pattern was found for *Vicia* sp. and field beans (x^2 = 52.83, d.f. = 4, $p < 0.0001$).

4.4.5. Comparison with 1952 honey survey

From the 855 honey samples characterised in 1952 using melissopalynology, a total of 66 plant taxa were identified; 5% to species, 65% to genus, 12% to type (a broader classification used in melissopalynology to indicate a wider range than the genus named), and 18% to family (Supplementary Figure 3). After adjusting for differing levels of taxonomic identification between the two surveys (1952 and 2017), 47 of taxa were found in both the 1952 and 2017 surveys (Figure 24). The same four plant taxa were identified in over 50% of samples in 1952 as in 2017, with *Trifolium repens* dominating, followed by a Maleae grouping including *Crataegus* spp., *Prunus* spp., and *Pyrus* spp., then *Rubus* spp. and *Brassica* spp. There was a positive correlation between 1952 and 2017 for the frequency of occurrence for the 47 matched taxa (Kendall's τ correlation coefficient $\tau = 0.389$, $p = 0.0001$) (Figure 23).

Of the 19 taxa found in 1952 but not in 2017, nine taxa could not be directly matched to results in the DNA due to differences in taxonomic resolution. For example, in 1952, some of the Asteraceae genera identified (*Tussilago* spp., *Achillea* spp.) may be represented under the family level Asteraceae identification in the DNA. The remaining ten taxa were not represented in the DNA, e.g. *Saxifraga* spp. In 2017, there were an additional 99 taxa not found in 1952 with 15 of these taxa occurring in more than 5% of samples. The majority of these taxa were not found at a predominant and secondary abundance with the notable exception of *Ulex* spp. (predominant and secondary in 4% of samples).

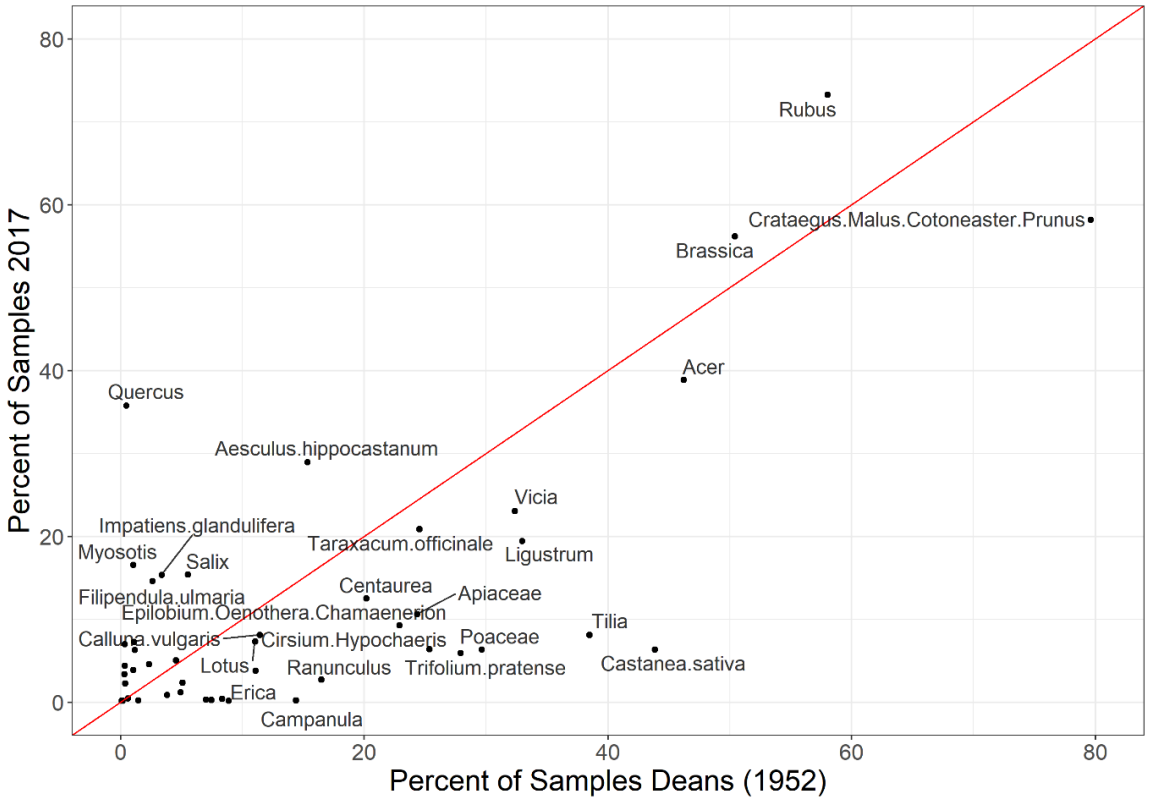


Figure 23: Comparing the total proportion of samples found in 1952 and 2017 for the plant taxa found in both surveys. There is a significant positive correlation (Kendall's τ correlation coefficient $\tau = 0.389$, $p = 0.0001$). Taxa are labelled when they appear in over 10% of samples for either the 1952 survey or the 2017 survey.

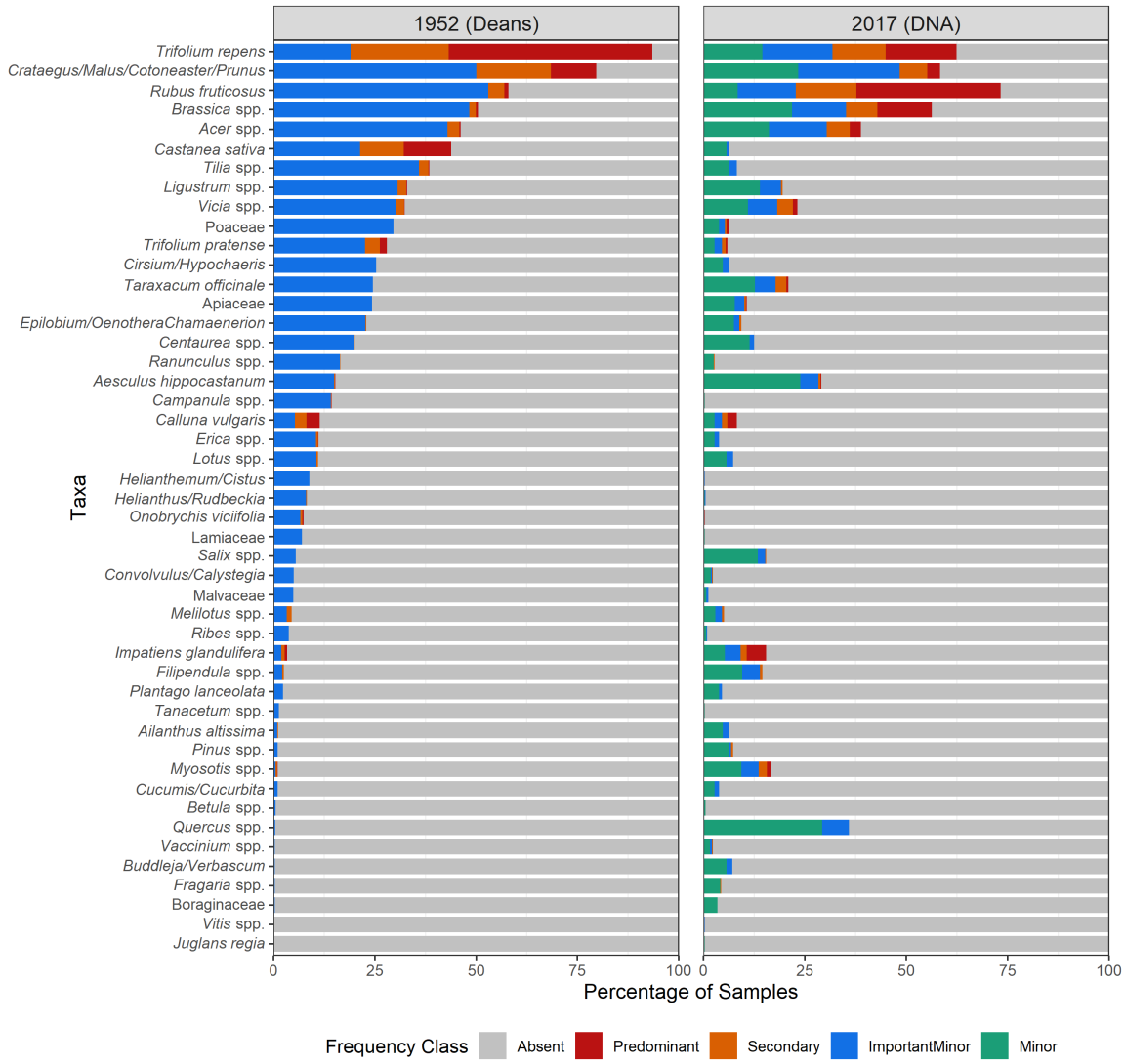


Figure 24: Comparing 47 taxa in honey samples analysed by Deans in 1952 (n = 855), using melissopalynology, with honey samples analysed using DNA metabarcoding (n = 441). Overall, there was a positive correlation between the two honey surveys for the total percentage of samples found for each taxa (r_τ = 0.389, p = 0.0001). In order to compare with the *Prunus, Pyrus, Crataegus* group identified by Deans, the DNA reads for the Maleae tribe and *Prunus* were grouped, to create the consensus group of *Crataegus, Malus, Cotoneaster,* and *Prunus*.

4.4.6. Comparison of dominant forage sources

There were significant changes in frequency in seven of the nine taxa found at a predominant and secondary level in over 1% of samples in both 1952 and 2017 (Table 6, Figure 25). These changes in frequency between the honey were compared with frequency changes in the Countryside Survey data between 1978 and 2007 (Table 6). The top forage found in 1952, *Trifolium repens,* was reported as predominant or secondary forage in 74% of samples, decreasing to 31% of honey samples in 2017 ($x^2 = 229.51$, d.f = 1, $p < 0.0001$). *Trifolium pratense* also decreased in use from 5% to 1% ($x^2 = 11.18$, d.f = 1, $p = 0.027$). Based on the Countryside Surveys, both *T. repens* and *T. pratense* showed a decline in frequency between 1978 and 2007, with a 13% and 28% decline respectively.

Brassica spp. were the predominant or secondary forage source in only 1% of samples in 1952 compared with 21% in 2017 ($x^2 = 131.46$, d.f = 1, $p < 0.0001$), almost certainly reflecting the increase of oilseed rape (*Brassica napus*) as a crop in the UK since the 1970s (Chamberlain and Fuller, 2000; Robinson and Sutherland, 2002; Sutcliffe and Kay, 2000). No significant difference was found between the honey surveys for another crop species, *Vicia* spp. ($x^2 = 7.15$, d.f = 1, $p = 0.255$), despite an increase in production of field beans (*Vicia faba*) since 1945 and in recent years (Breeze et al., 2011; DEFRA, 2017; Williams and Carreck, 1994).

Contrasting the decline in *T. repens*, there has been an increase in the forage usage of *Rubus* spp. compared to 1952, where it was the third most common taxa found in the honey; in 2017, it is the dominant forage plant. In 1952, *Rubus* spp. were the predominant or secondary forage in only 5% of samples, compared to 31% of samples in 2017 ($x^2 = 367.07$, d.f = 1, $p < 0.0001$), while in the Countryside Survey data, *Rubus fruticosus* agg. was recorded as increasing by 21% between 1978 and 2007.

A small increase in the predominant or secondary forage use of *Acer* spp. was also observed between 1952 and 2017 ($x^2 = 14.92$, d.f = 1, $p = 0.004$), with a 16% increase in the frequency found in 2007 compared to 1978 in the Countryside Survey. A significant decrease was seen for the predominant or secondary use of the *Crataegus* spp.*, Malus* spp.*, Cotoneaster* spp.*,* and *Prunus* spp. group ($x^2 = 62.25$, d.f = 1, $p < 0.0001$). This is less clearly resolved in the Countryside Survey data with local frequency increases in *C.*

monogyna and *Prunus spinosa* between 1978 and 2007. No significant difference was found between honey in 1952 and 2017 for *Calluna vulgaris* ($x^2 = 3.04$, d.f = 1, p = 1.00), despite a decline in the frequency of *C. vulgaris* from 1978 to 2007, and loss of suitable heathland habitat (Preston et al., 2002).

Impatiens glandulifera increased as a predominant and secondary forage from 1% of samples in 1952 to 6% in 2017 ($x^2 = 22.17$, d.f = 1, $p < 0.0001$), representing an invasive species which has spread widely in the UK during the 20th century (Rich and Woodruff, 1996; Usher et al., 1986).

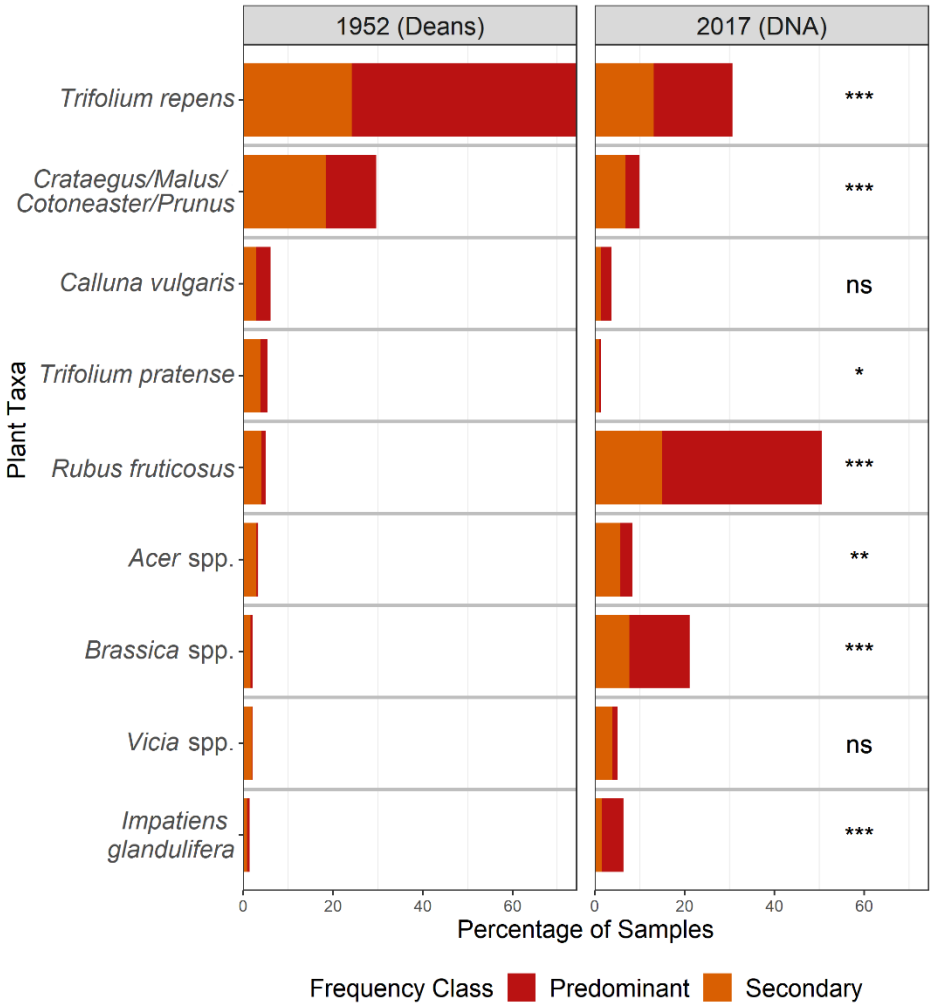


Figure 25: Change (%) in plant taxa used by honeybees from 1952 to 2017. The taxa included are those found as predominant and secondary within honey samples (>15% of pollen grains in melissopalynology or >15% of DNA sequences) for more than 1% of samples in both surveys.

Table 6: Summary of top taxa between 1952 and 2017, for plants representing the major forage (predominant and secondary: >15% of pollen grains or >15% of DNA sequences) for greater than 1% of samples in both surveys. The p-value for the chi-squared tests used Bonferroni correction for multiple testing is given, testing the change in major forage use between the two surveys.

Plant Taxa	1952 Present N	2017 Present N	1952 Major Forage	2017 Major Forage	x^2	p-value	New Atlas Change Index	CS Plot frequency 1978	CS Plot frequency 2007	CS % Change (1978-2007)
Trifolium repens	799 (93%)	275 (62%)	636 (74%)	135 (31%)	229.51	0.000	1.31	561	488	-13.01
Crataegus Malus							*-0.76 (Crataegus mongoyna),*	*198 (C. monogyna)*	*216 (C. monogyna)*	*+9.09 (C. mongoyna),*
Cotoneaster Prunus	681 (80%)	257 (58%)	253 (30%)	44 (10%)	62.25	0.000	*+0.57 (Malus sylvestris), +0.40 (Prunus spinosa)*	*71 (P. spinosa)*	*101 (P.spinosa)*	*+42.25 (P. spinosa)*
Calluna vulgaris	98 (11%)	36 (8%)	52 (6%)	16 (4%)	3.05	1.000	-0.64	370	312	-15.68
Trifolium pratense	239 (28%)	26 (6%)	46 (5%)	6 (1%)	11.18	0.027	-0.18	144	104	-27.77
Rubus	496 (58%)	323 (73%)	43 (5%)	223 (51%)	367.07	0.000	*-0.29 (R. fruticosus agg.)*	281	341	*+21.3 (R. fruticosus agg.)*
Acer	395 (46%)	171 (39%)	28 (3%)	37 (8%)	14.92	0.004	*+0.35 (A. campestre), -0.40 (A. pseudoplatanus)*	74	86	+16.22 (*Acer* spp.)
Brassica	431 (50%)	248 (56%)	18 (2%)	93 (21%)	131.46	0.000	*+2.88 (B. napus), +0.74 (B. rapa), +0.90 (B. oleraceae)*	18	20	*+11.11 (Brassica spp.)*
Vicia	277 (32%)	102 (23%)	18 (2%)	22 (5%)	7.15	0.247	*-0.37 (V. cracca), +0.05 (V. hirsuta), +0.19 (V. sativa)*	132	98	*-25.76 (Vicia sp.)*
Impatiens glandulifera	29 (3%)	68 (15%)	12 (1%)	28 (6%)	22.17	0.000	+1.85	2	4	+100

4.5. Discussion

4.5.1. Changes in forage plants reflecting agricultural intensification and changes in crop cultivation

In the UK, the post-war intensification of agriculture, with the introduction of inorganic fertilisers and herbicides, has led to a decline in flower rich habitat (Green, 1990; Robinson and Sutherland, 2002). Improved grasslands covering large areas of the UK have been estimated to provide the greatest contribution to nationwide nectar provision, with *Trifolium repens* as the dominant source of nectar (Baude et al., 2016). However, clover availability has reduced within grasslands, due to decreasing use of clover leys in rotation and the increased application of inorganic nitrogen fertilizers and herbicides, which reduce forb diversity and increase graminoids (Robinson and Sutherland, 2002; Wilson et al., 1999), with a 13% decline in local-scale frequency between 1978 and 2007 for *T. repens* and a 28% decline for *Trifolium pratense* seen in the Countryside Survey. Clover leys would have represented much of the temporary grassland available in the late 1940s but by the 1980s it was less than 1% of all temporary grassland (Chamberlain et al., 2000). This reflects the decline seen in the predominant or secondary use of *T. repens* from 74% to 31% of honey samples. Nevertheless, *T. repens* still represented the second most present taxa found in the 2017 honey suggesting that, despite declines in land cover, honeybees are still actively seeking out clover.

At the same time, the UK has seen an increase in area of the crop oilseed rape (*Brassica napus*), corresponding to an increase in *Brassica* spp. found in the honey. Oilseed rape was introduced to the UK as a crop in the 1960s, with 4884 ha grown in 1969 compared to 279,030 in 1988 (Chamberlain and Fuller, 2000; Robinson and Sutherland, 2002; Sutcliffe and Kay, 2000). Hives within 2 km of oilseed rape fields were more likely to contain *Brassica* spp. in the honey, a signal detected despite the likely presence of other native *Brassica* species. In *Plants and Beekeeping*, Howes, (1945) notes only *Brassica nigra* and *Brassica arvensis* (now *Sinapsis*) as honeybee plants, but does not include *Brassica napus*. The revised edition from 1979 included oilseed rape under major honey plants, noting the increase in production during the 1970s.

Vicia spp. includes another insect pollinated crop, field beans (*Vicia faba*), which has increased in production since 1945 and in recent years (DEFRA, 2017; Williams and Carreck, 1994), however no significant change in the predominant or secondary presence in the honey was detected between the two surveys. A greater occurrence of *Vicia* spp. in the honey from hives with field beans in a 2 km radius was detected. Despite the increase in field bean crop, other *Vicia* species in the landscape may have declined in availability, such as vetches used as fodder crops associated with leys, similar to the decline in clover (Bryant and Hughes, 2011; Wilson et al., 1999).

4.5.2. Changes in forage on native species

Comparatively to the decline in *Trifolium repens*, *Rubus* spp. has seen an increase in usage compared to 1952, where it was the third most common taxa found in the honey. In 2017, it is the most common plant found, with a large increase in occurrence as a predominant or secondary forage within the honey samples, from 5% to 51%. *Rubus* spp. includes the native *Rubus fruticosus* agg., an aggregate of over 320 microspecies, which is widespread and found in woods, hedges, scrub and waste ground, as well as *Rubus idaeus*, which is cultivated as raspberry. *T. repens* and *Rubus* spp. have similar flowering periods covering June to September and the increased predominance of *Rubus* spp. in the honey may also be reflecting the decreased availability of *T. repens* during the same flowering period. *Rubus* spp. and *T. repens* offer both pollen and nectar forage. However, the protein content and proportion of essential amino acids is lower in *R. fruticosus* compared to *T. repens,* meaning the honeybees may not be gaining the same nutritional benefits if substituting *T. repens* with *Rubus* spp. (Franco et al., 2008).

There was no significant difference found in foraging for the native plant *Calluna vulgaris* between 1952 and 2017, despite a decline in suitable heathland and a local frequency decline between 1978 and 2007 (Preston et al., 2002). Honeybees are known to forage greater than average distances (10 km) for heather (Beekman and Ratnieks, 2000), and the findings may also reflect beekeepers targeting *C. vulgaris* as a honey crop by locating their hives to maximise the output of heather honey.

Foraging on the *Crataegus* spp.*, Malus* spp.*, Cotoneaster* spp.*,* and *Prunus* spp. group decreased between 1952 and 2017, with predominant or secondary use declining from

30% to 10% of samples. This fruit tree group has been identified as a top source of forage (Synge, 1947) and is taxonomically difficult to distinguish, both using the DNA and by morphology of the pollen. A factor impacting the availability of this group of species includes the decline in area of orchards in the UK (*Malus* spp., *Prunus* spp. and *Pyrus* spp.), from 30,389 in 1985 to 24,449 ha in 2017 (DEFRA, 2017). Additionally, in the wider landscape, characteristic hedging species such as *Crataegus monogyna* have likely been affected by the increasing mechanisation of farmlands since the 1960s leading to the widespread removal of hedgerows (Barr and Gillespie, 2000; Robinson and Sutherland, 2002). Due to their high potential nectar productivity in a small area, hedgerows and linear features have been suggested as an efficient way to increase local nectar resource within agri-environment schemes (Baude et al., 2016). The nectar of plants with exposed nectaries such as *C. monogyna* and *Prunus spinosa* are influenced more by the local temperature and humidity, with the nectar of both *C. monogyna* and *P. spinosa* being described as sparse and concentrated (Corbet et al., 1979; Gyan and Woodell, 1987), which may relate to the reputation of *C. monogyna* with beekeepers as an inconsistent source of nectar flow for honey, within and between seasons (Howes, 1945; Kirk and Howes, 2012).

4.5.3. Changes in the availability of invasive species

Introduced to the UK in 1839, *Impatiens glandulifera* represents an invasive plant species that shows an increased predominance in the honey after increasing in distribution during the 20th century. In 1900, there were only eight 10 km square records, with 75 recorded by 1940 and 614 squares by 1960 (Rich and Woodruff, 1996; Usher et al., 1986). While *I. glandulifera* was twentieth overall in occurrence, it represented the sixth most frequent species when looking at predominant and secondary forage, indicating that when the honeybees were foraging on *I. glandulifera* they were using it as their main source of nectar. *I. glandulifera* likely occurs at a lower overall frequency across the honey samples due to the height of its flowering occurring after the peak in honey sampling. *I. glandulifera* is extremely attractive to pollinators, as its sugar production is higher than other plant species associated with the same habitat, with its presence negatively affecting seed set in plant species competing for pollinators (Chittka and Schürkens, 2001).

4.5.4. Comparison between DNA metabarcoding and melissopalynology

The read counts returned with DNA metabarcoding sequencing are commonly considered semi-quantitative, with biases through the process of sampling, DNA extraction, PCR and sequencing (Alberdi et al., 2019; Zinger et al., 2019). DNA metabarcoding is often able to detect more taxa compared to traditional techniques: firstly, through identifying rarer species in the sample, and secondly, through achieving higher taxonomic resolution. This was seen here with several plant families, including Poaceae and Apiaceae (Kraaijeveld et al., 2015; Prosser and Hebert, 2017), although certain groups are still taxonomically difficult to resolve even with DNA, such as the Maleae tribe.

Here we show a significant correlation in the overall presence and absence of plant taxa found by both the DNA metabarcoding in 2017 and by melissopalynology in 1952 (Deans, 1958, 1957), with significant differences when just predominant and secondary foraging is examined. Overall, previous studies have found a concordance between the taxa found using microscopy techniques and DNA, with both techniques returning the plants found abundantly within a sample (Hawkins et al., 2015; Richardson et al., 2015; Smart et al., 2017). Rarer species in a sample, however, are less likely to be found consistently by both methods, making it more appropriate to examine differences in frequently found, abundant taxa. Using frequency classes is a good approach for retaining abundance information without assuming the data to be fully quantitative. Here we put the proportion of DNA sequence reads into classes to compare against the melissopalynology.

The results of the honey analysis conducted by Deans (1958, 1957) are supported by other contemporary sources, with differing levels of taxonomic identification of the pollen. Synge, (1947) analysed pollen loads through the 1945 and 1946 season in England, with *Trifolium repens* and *Trifolium pratense* identified as the top species found overall, while Percival (1947) found *T. repens* and *Rubus fruticosus* top in pollen loads collected from hives in South Wales. There were 15 plant taxa found in the DNA in over 5% of samples that were not found by Deans, (1958, 1957), nine of which were identified in the pollen loads by Synge (1947) and Percival (1947). *Ulex* spp. was the only top taxa found predominantly in the DNA that Deans did not also identify. Contemporary sources identify *Ulex* spp. in honeybee pollen loads (Percival, 1947; Synge, 1947), so this genus was likely

missed by Deans (1958, 1957). Deans (1958, 1957) did not report the plant species classed as minor (less than 1% of pollen grains), which will have affected the overall species list found, and in addition 44% of samples contained pollen at important minor levels that could not be identified, making it likely that some plant taxa were missed.

In both surveys, beekeepers were asked to send honey from their normal extraction. The period of heaviest nectar production is known by beekeepers as honey flows, which can be affected by weather conditions both in terms of the plants' nectar production and the honeybees' ability to fly and forage during poor weather (Winston, 1987). The frequency of plant taxa which flower in the earlier and later extremes of honeybee foraging may be more affected by this; which includes the fruit trees, a key group in April and May, as well as *I. glandulifera*, which can flower from July to October. These patterns follow those found when sampling the honey on a monthly basis (Chapter 3). However, while the greatest proportion of samples with the Maleae tribe present occur in May and June, and the greatest proportion of *I. glandulifera* occurred in August and September, a smaller proportion of predominant and secondary honey samples were still found in other parts of the season, suggesting that sampling occurring at peak honey flows still captures information about taxa from the start and end of the season.

4.5.5. Implications for foraging behaviour and habitat management

Honeybees have been referred to as super-generalists within plant-pollinator networks, as they visit a wide range of plant taxa, however we know them to be selective with the plants they forage on (Chapter 3). Even as generalist foragers honeybees are making choices in their foraging, and the decreased availability of *Trifolium repens* may have resulted in the increased use of *Rubus* spp. While abundance within the landscape is one factor affecting foraging choice, it may not be the single main driver for selection (Chapter 3). *Rubus* spp. is a genus which has been found to be a relatively high-quality mono-floral pollen source (Di Pasquale et al., 2013), potentially mitigating the benefit of a poly-floral diet. A small number of key taxa across the UK were frequently and abundantly foraged, representing the species which provide the greatest abundance of nectar nationally within the UK, (Baude et al., 2016). Although honeybees represent good surveyors of nectar within the landscape, with their ability to communicate sources of forage, they are

also likely buffered against the interacting stressors affecting wild bees and other pollinators. Consequently, the loss of potential forage may impact more negatively on wild pollinators. On a landscape scale, two main management recommendations could mitigate declines in available forage. Improved grasslands represent the greatest area of forage availability. Increasing the amount of flowering clover, *Trifolium repens*, is likely to have the most potential beneficial impact on honeybee forage provision. In addition, the presence of top forage plants such as *Rubus* spp. or *Crataegus monogyna* could be increased in hedgerows and linear features, efficiently increasing their nectar productivity.

4.5.6. Conclusions

Here, we present the first UK wide floral analysis of honey samples since the melissopalynology survey by Deans in 1952 (Deans, 1958, 1957). The hive location was not found to have a significant effect on the floral composition of the honey, with the key forage species occurring across regions. However spatial patterns were associated with the presence of geographically restricted species, including urban planted tree species *Ailanthus altissima*, and the crop species; oilseed rape (*Brassica napus*) and field beans (*Vicia faba*). The changes found between 1952 and 2017 in the most abundant forage in the honey are reflecting widespread changes in the available forage in the landscape since the 1940s. Evidence of agricultural intensification on agricultural fodder and crop species, changes in native plants, and the spread of invasive species are all recorded in environmental DNA traces in the honey. Charting these changes has implications for pollinators in managing habitat for the landscape availability of nectar and pollen forage.

4.6. References

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M., Gilbert, M.T.P., 2019. Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources* **19**, 327–348.

Barr, C.J., Gillespie, M.K., 2000. Estimating hedgerow length and pattern characteristics in Great Britain using Countryside Survey data. *Journal of Environmental Management* **60**, 23–32.

Batáry, P., Dicks, L. V., Kleijn, D., Sutherland, W.J., 2015. The role of agri-environment schemes in conservation and environmental management. *Conservation Biology* **29**, 1006–1016.

Baude, M., Kunin, W.E., Boatman, N.D., Conyers, S., Davies, N., Gillespie, M.A.K., Morton, R.D., Smart, S.M., Memmott, J., 2016. Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature* **530**, 85–88.

Beekman, M., Ratnieks, F.L.W., 2000. Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology* **14**, 490–496.

Benton, T.G., Vickery, J.A., Wilson, J.D., 2003. Farmland biodiversity: Is habitat heterogeneity the key? *Trends in Ecology and Evolution* **18**, 182–188.

Breeze, T.D., Bailey, A.P., Balcombe, K.G., Potts, S.G., 2011. Pollination services in the UK: How important are honeybees? *Agriculture, Ecosystems & Environment* **142**, 137–143.

Brodschneider, R., Crailsheim, K., 2010. Nutrition and health in honey bees. *Apidologie* **41**, 278–294.

Bryant, J.A., Hughes, S.G., 2011. Vicia. In: Wild Crop Relatives: Genomic and Breeding Resources: Legume Crops and Forages. Springer Berlin Heidelberg, pp. 273–289.

Carvell, C., Roy, D.B., Smart, S.M., Pywell, R.F., Preston, C.D., Goulson, D., 2006. Declines in forage availability for bumblebees at a national scale. *Biological Conservation* **132**,

481–489.

Chamberlain, D.E., Fuller, R.J., 2000. Local extinctions and changes in species richness of lowland farmland birds in England and Wales in relation to recent changes in agricultural land-use. *Agriculture, Ecosystems & Environment* **78**, 1–17.

Chamberlain, D.E., Fuller, R.J., Bunce, R.G.H., Duckworth, J.C., Shrubb, M., 2000. Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *Journal of Applied Ecology* **37**, 771–788.

Chittka, L., Schürkens, S., 2001. Successful invasion of a floral market. *Nature* **411**, 653–653.

Corbet, S.A., Unwin, D.M., Prŷs-Jones, O.E., 1979. Humidity, nectar and insect visits to flowers, with special reference to *Crataegus*, *Tilia* and *Echium*. *Ecological Entomology* **4**, 9–22.

de Vere, N., Jones, L.E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S., Ford, C.R., 2017. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports* **7**, 42838.

de Vere, N., Rich, T.C.G., Ford, C.R., Trinder, S.A., Long, C., Moore, C.W., Satterthwaite, D., Davies, H., Allainguillaume, J., Ronca, S., Tatarinova, T., Garbett, H., Walker, K., Wilkinson, M.J., 2012. DNA Barcoding the Native Flowering Plants and Conifers of Wales. *PLoS ONE* **7**, e37945.

Deans, A.S.C., 1957. *Survey of British Honey Sources*. Bee Research Association.

Deans, A.S.C., 1958. The Pollen Analysis of some British Honeys. Thesis for the National Bee Diploma.

DEFRA, 2017. Agriculture in the United Kingdom.

Di Pasquale, G., Salignon, M., Le Conte, Y., Belzunces, L.P., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L., Alaux, C., 2013. Influence of Pollen Nutrition on Honey Bee Health: Do Pollen Quality and Diversity Matter? *PloS ONE* **8**.

Eyer, M., Neumann, P., Dietemann, V., 2016. A look into the cell: Honey storage in honey bees, *Apis mellifera*. *PLoS ONE* **11**, 1–20.

Franco, M., Darvill, B., Goulson, D., Hanley, M.E., Pichon, S., Franco, M., Pichon, S., Darvill, B., Goulson, D., 2008. Breeding system, pollinator choice and variation in pollen quality in British herbaceous plants. *Functional Ecology* **22**, 592–598.

French, K.E., 2017. Species composition determines forage quality and medicinal value of high diversity grasslands in lowland England. *Agriculture, Ecosystems and Environment* **241**, 193–204.

Fuller, R.M., 1987. The changing extent and conservation interest of lowland grasslands in England and Wales: A review of grassland surveys 1930-1984. *Biological Conservation* **40**, 281–300.

Green, B.H., 1990. Agricultural intensification and the loss of habitat, species and amenity in British grasslands: a review of historical change and assessment of future prospects. *Grass and Forage Science* **45**, 365–372.

Gyan, K.Y., Woodell, S.R.J., 1987. Nectar Production, Sugar Content, Amino Acids and Potassium in *Prunus spinosa* L., *Crataegus monogyna* Jacq. and *Rubus fruticosus* L. at Wytham, Oxfordshire. *Functional Ecology* **1**, 251.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., Baillie, L., Adams-Groom, B., 2015. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. *PLOS ONE* **10**.

Howes, F.N., 1945. *Plants and Beekeeping*. Faber & Faber, London.

IPBES, 2016. The Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on Pollinators, Pollination and Food Production.

Jones, B., Wilford, J., Cierniak, D., 2011. Managing Bees to Maximise Honey Production.

Kirk, W.D.J., Howes, F.N., 2012. Plants for Bees. IBRA.

Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., den Dunnen, J.T., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources* **15**, 8–16.

Louveaux, J., Maurizio, A., Vorwohl, G., 1978. Methods of melissopalynology. *Bee World* **5**, 139–153.

Manley, R.O.B., 1936. *Honey Production in the British Isles*.

Maurizio, A., Hodges, F.E.D., 1951. Pollen Analysis of Honey. *Bee World* **32**, 1–5.

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology* **10**.

Ollerton, J., Erenler, H., Edwards, M., Crockett, R., 2014. Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. *Science* **346**, 1360–1362.

Percival, M., 1947. Pollen collection by *Apis mellifera*. *New Phytologist* **46**, 142–165.

Petit, S., Stuart, R.C., Gillespie, M.K., Barr, C.J., 2003. Field boundaries in Great Britain: Stock and change between 1984, 1990 and 1998. *Journal of Environmental Management* **67**, 229–238.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010a. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25**, 345–353.

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., Vanbergen, A.J., 2016. Safeguarding pollinators and their values to human well-being. *Nature* **540**, 220–229.

Potts, S.G., Roberts, S., Dean, R., Marris, G., Brown, M., Jones, R., Neumann, P., Settele, J., 2010b. Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research* **49**, 15.

Preston, C.D., Pearman, D.A., Dines, T.D., 2002. *New Atlas of the British and Irish Flora: An*

Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands. Oxford University Press, Oxford.

Prosser, S.W.J., Hebert, P.D.N., 2017. Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry* **214**, 183–191.

R Development Core Team, R., 2011. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

Requier, F., Odoux, J.F., Henry, M., Bretagnolle, V., 2017. The carry-over effects of pollen shortage decrease the survival of honeybee colonies in farmlands. *Journal of Applied Ecology* **54**, 1161–1170.

Rich, T.C.G., Woodruff, E.R., 1996. Changes in the vascular plant floras of England and Scotland between 1930-1960 and 1987-1988: The BSBI monitoring scheme. *Biological Conservation* **75**, 217–229.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015. Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. *Applications in Plant Sciences* **3**.

Robinson, R.A., Sutherland, W.J., 2002. Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology* **39**, 157–176.

Rowland, C.S., Morton, R.D., Carrasco, L., McShane, G., O'Neil, A.W., Wood, C.M., 2017. Land Cover Map 2015 (vector, GB). *NERC Environmental Information Data Centre*.

Schürch, R., Couvillon, M.J., Ratnieks, F.L.W., 2015. Determining the foraging potential of oilseed rape to honey bees using aerial surveys and simulations. *Journal of Apicultural Research* **54**, 238–245.

Seeley, T.D., 1995. *The Wisdom of the Hive*. Harvard University Press.

Smart, M.D., Cornman, R.S., Iwanowicz, D.D., McDermott-Kubeczko, M., Pettis, J.S., Spivak, M.S., Otto, C.R.V., 2017. A comparison of honey bee-collected pollen from

working agricultural lands using light microscopy and its metabarcoding. *Environmental Entomology* **46**, 38–49.

Stace, C., 2010. New Flora of the British Isles, Third. ed. Cambridge University Press.

Stevens, Hopkins, Mackintosh, Rimes, Robertson, Blackstock, Jefferson, 2003. The extent of semi-natural grassland communities in lowland England and Wales: a review of conservation surveys 1978-96. *Grass and Forage Science* **54**, 1–18.

Sutcliffe, O.L., Kay, Q.O.N., 2000. Changes in the arable flora of central southern England since the 1960s. *Biological Conservation* **93**, 1–8.

Synge, A.D., 1947. Pollen collection by honeybees (*Apis mellifera*). *Journal of Animal Ecology* **16**, 122–138.

Usher, M.B., Kornberg, H., Horwood, J.W., Southwood, R., Moore, P.D., 1986. Invasibility and Wildlife Conservation: Invasive Species on Nature Reserves. *Philosophical Transactions of the Royal Society B: Biological Sciences* **314**, 695–710.

Vanbergen, A.J., 2013. Threats to an ecosystem service: pressures on pollinators. *Frontiers in Ecology and the Environment* **11**, 251–259.

Visscher, P.K., Seeley, T.D., 1982. Foraging Strategy of Honeybee Colonies in a Temperate Deciduous Forest. *Ecology* **63**, 1790.

Von Der Ohe, W., Persano Oddo, L., Piana, M.L., Morlot, M., Martin, P., 2004. Harmonized methods of melissopalynology. *Apidologie* **35**, S18–S25.

Williams, I.H., Carreck, N.L., 1994. Land use changes and honey bee forage plants. In: Forage for Bees in an Agricultural Landscape. pp. 7–20.

Wilson, J.D., Morris, A.J., Arroyo, B.E., Clark, S.C., Bradbury, R.B., 1999. A review of the abundance and diversity of invertebrate and plant foods of granivorous birds in northern Europe in relation to agricultural change. *Agriculture, Ecosystems and Environment* **75**, 13–30.

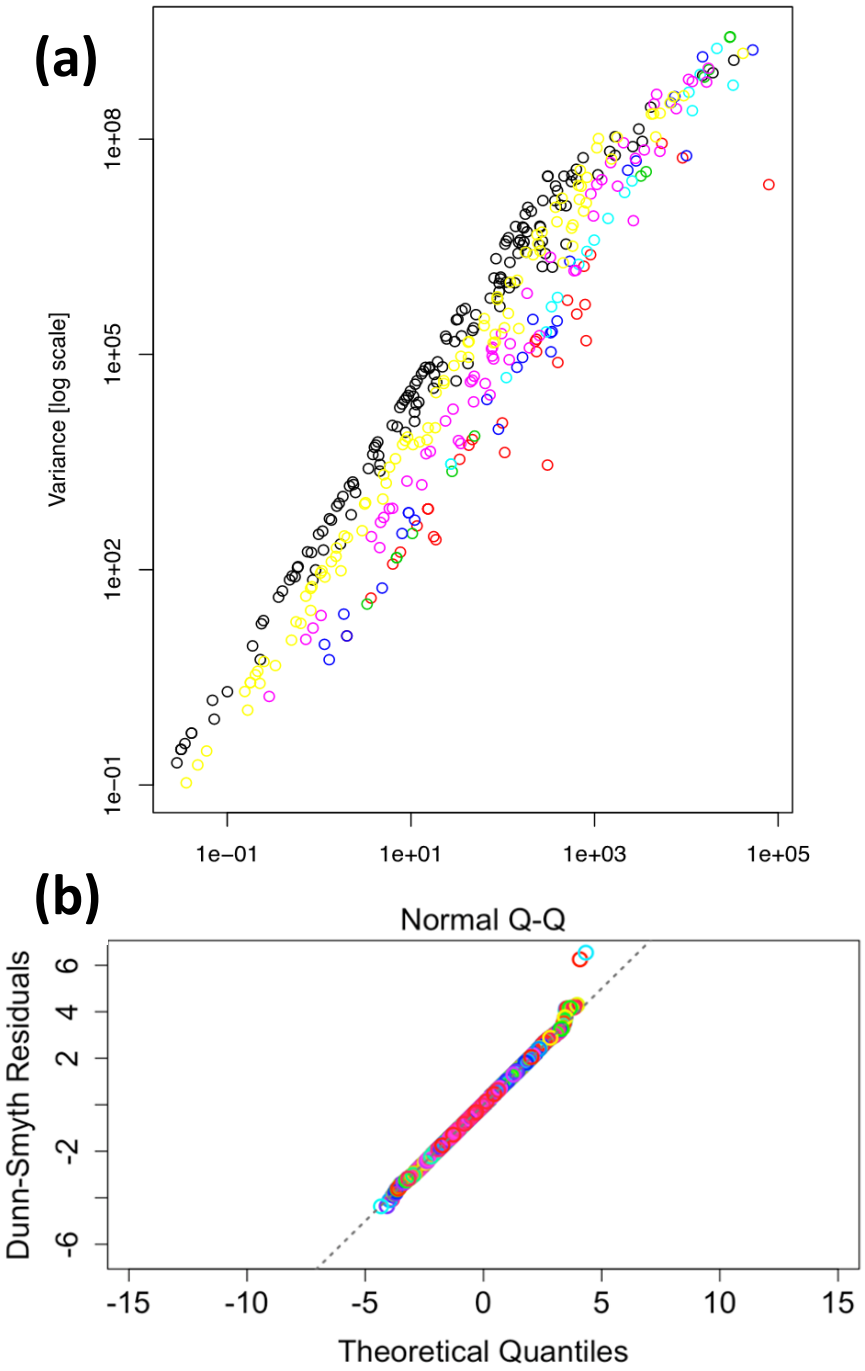
Winston, M.L., 1987. *The Biology of the Honey Bee*. Harvard University Press.

Wood, C.M., Smart, S.M., Bunce, R.G.H., Norton, L.R., Maskell, L.C., Howard, D.C., Scott, W.A., Henrys, P.A., 2017. Long-term vegetation monitoring in Great Britain - The Countryside Survey 1978-2007 and beyond. *Earth System Science Data* **9**, 445–459.

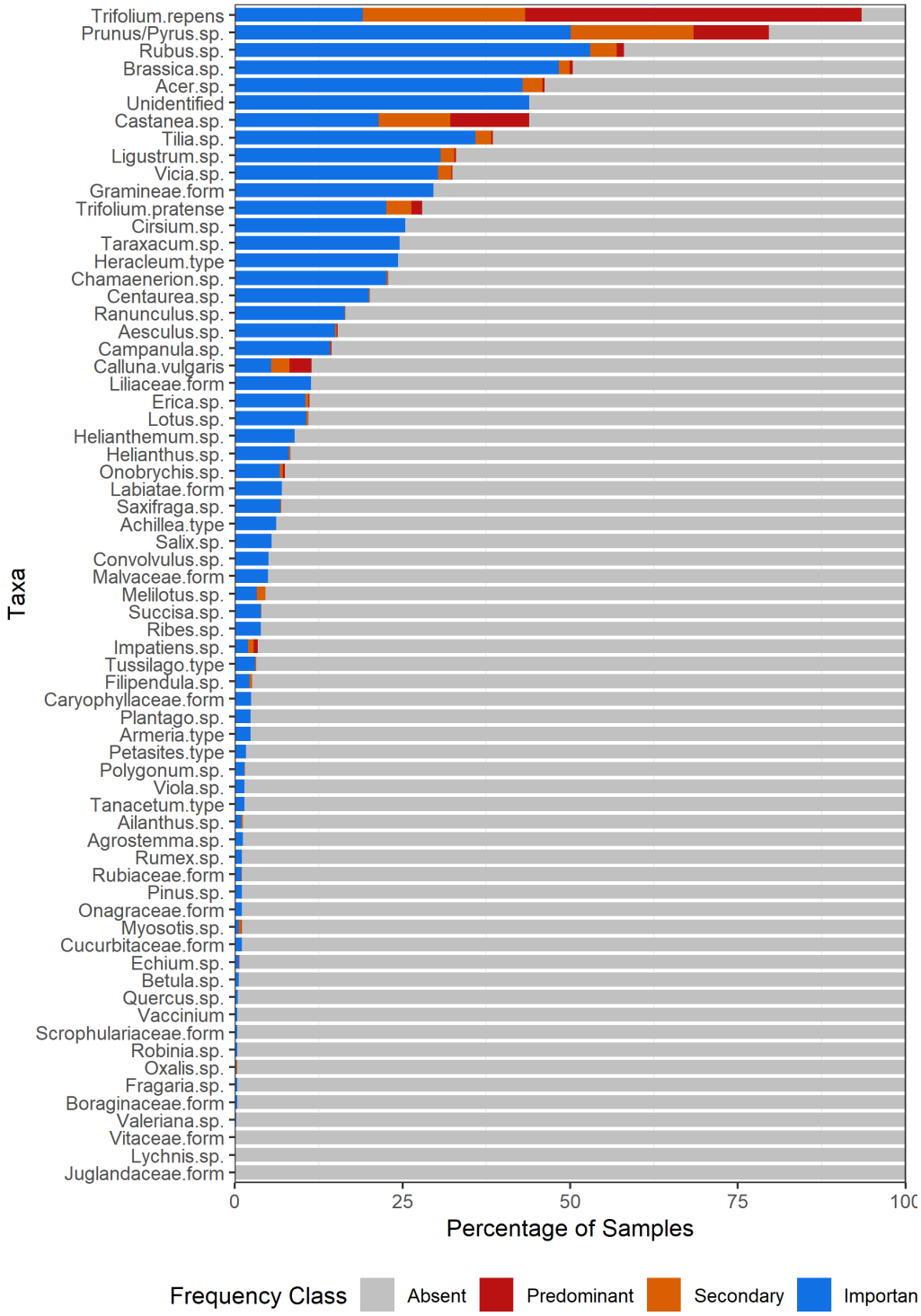
Wright, G.A., Nicolson, S.W., Shafir, S., 2018. Nutritional Physiology and Ecology of Honey Bees. *Annual Review of Entomology* **63**, 327–344.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., De Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., Kauserud, H., Orlando, L., Pansu, J., Pawlowski, J., Tedersoo, L., Thomsen, P.F., Willerslev, E., Taberlet, P., 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* **28**, 1857–1862.

4.7. Supplementary Figures



Supplementary Figure 2: (a) There is a strong relationship between the mean proportion of sequences and the variance of the proportion of sequences from each sampling site. Coloured circles indicate sampling region. The plots were produced using the meanvar.plot function in the mvabund package in R (Wang et al., 2012). (b) Scatter plot of theoretical quantile values and the residuals output from the model used to analyse the abundance data produced by metabarcoding (*rbcL* and ITS2 markers). Deviations from the straight line are minimal indicating a normal distribution and suggests that the model selected is plausible and the mean-variance assumption of the negative binomial regression is correct. Coloured circles denote different genera in the abundance data. The plot was produced using the plot.manyglm function in the mvabund package.



Supplementary Figure 3: Full list of taxa identified in 855 honey samples by Deans (1952) labelled with the taxa as in 1952.

CHAPTER 5

Discussion

5.1. Discussion

5.1.1. Aims and objectives of project

- Create a DNA barcode reference library for the UK native flowering plant and conifer species, using three DNA barcode loci, *rbcL*, *matK*, and ITS2.
    - How do *rbcL*, *matK*, and ITS2 vary in their ability to successfully recover a sequence, within the UK flora?
    - How do *rbcL*, *matK*, and ITS2 vary in their species level discrimination?
- Create a reference library including UK non-native plant species from all available DNA sequences on GenBank to support the applications of DNA barcoding.
    - How representative are the currently available sequences for the UK non-native and native plant species?
    - What is the species level discrimination ability within the reference library for the selected native and non-native UK plant species?

A high-quality DNA barcode reference library was created for the UK native flowering plant species, providing a resource for this and future DNA barcoding and metabarcoding projects. This was augmented with sequences from GenBank to create a curated resource for wider applications such as investigating pollinator foraging. The current representation and taxonomic discrimination ability of the UK plant species was assessed, giving key context to any applications of the reference library.

- Investigate the foraging preferences of honeybees using DNA metabarcoding within a diverse floral landscape.
    - Which plants are honeybees choosing throughout the season (April to September) when offered a diverse floral resource?
    - Are there differences in foraging between hives placed near horticultural plants, compared with hives situated further away but within foraging distance?
    - How does the abundance of a plant within the landscape relate to the abundance found within the honey?

    - Is there any pattern to the characteristics of the forage plants found throughout the season, in terms of native status, growth form and associated habitat?

Honeybee foraging was examined within a diverse landscape and characterised over the foraging season and across two years. The hives proximity to a diverse floral resource was found to predict some of the differences in floral composition of honey but the major foraging plants were found to be consistent across hives. Honeybees were found to use a small proportion of plants to supply the majority of their nutritional needs throughout the season, with many of the taxa found being characterised as native plants which occur in hedgerow and linear features and grassland habitats.

- Characterise the plants honeybees are using across the UK using DNA metabarcoding
    - Are there spatial and temporal patterns to honeybee foraging within the UK?
    - How does the presence of crop species within foraging distance of the hive relate to the floral composition of the honey?
    - Have landscape scale changes in forage availability since the early 20th century affected the use of key honeybee forage plants in the UK?

Honeybee foraging was examined at a national scale using honey from beekeepers across the UK for the first time since 1952. The top forage plants of the UK were characterised and temporal patterns to honeybee foraging were revealed, matching the detailed survey of Chapter 3. The hive location was not found to have a significant effect on the floral composition of the honey, however spatial patterns were associated with the presence of geographically restricted species, including urban planted tree species *Ailanthus altissima*, and crop species oilseed rape and field beans. By comparing with the results of the last survey which occurred in 1952 (Deans, 1957, 1958), we evidence the declines in the landscape availability of nectar resource, with changes in the use of native species relating to agricultural intensification since the mid 20th century.

5.1.2. The future of DNA barcoding reference libraries

Plants are one of the more well understood groups of species, with around 374,000 described and accepted species and 2,000 new species described each year (Christenhusz and Byng, 2016). However, an estimated 70,000 flowering species are yet to be described, leaving the continuing characterisation of diversity an ongoing challenge, often limited by the availability of taxonomic expertise and financial support (Bebber et al., 2010). The ability to accurately identify species is vital to understanding ecosystems and DNA barcoding offers a strategy to provide progress on this challenge by describing both existing and unknown biodiversity.

In this PhD, I have created a comprehensive national DNA barcode reference library with associated voucher specimens and metadata for the UK native flowering plants and conifers, representing 1,482 species for three DNA barcode markers. Overall, 97% of the UK plants are represented with at least one marker. For non-native plant species, the UK natives reference library was supplemented with sequences from GenBank, creating a curated restricted reference library suitable for wider applications within the UK. I have used this high-quality, well curated resource to investigate the foraging preferences of honeybees, and it can now also be used to support many other applications requiring the identification of plants. For example, in Fahner et al., (2016), where biodiversity monitoring was completed by using soil samples to investigate plant diversity and therefore the quality of sites, or in the regulation of food products as in De Boer et al., (2017), where illegal plant species were detected in commercial tea products.

When comparing between the DNA barcode markers, *rbcL* and ITS2 were found to show similar patterns of performance with both the UK native flora and with naturalised or horticultural species. Taxonomic universality was greatest with *rbcL*, paired with relatively lower species discrimination. ITS2, in contrast, showed improved species discrimination but this success was moderated by its more limited ability in initially recovering a sequence. Meanwhile *matK*, in the UK natives, showed a higher level of species level discrimination than *rbcL*, but lower than ITS2. In contrast, sequence recovery with *matK* was better than ITS2, but worse compared to *rbcL*, with *matK* requiring several family specific primer pairs to achieve a comparable number of successful sequences. While

matK is not suitable for DNA metabarcoding applications, it is useful in phylogenetic studies as it has been shown to evolve faster in comparison to *rbcL* (Hilu et al., 2003; Lim et al., 2014). The *matK* region has also been investigated for potential mini-barcodes, where a shorter region is used for application in DNA metabarcoding, however success was limited (Garbett, 2016).

There are relative benefits to the use of different markers for different projects and the choice of marker should be undertaken with an understanding of what is required in the study, in terms of taxonomic range, target species and the source of DNA (Creer et al., 2016; Deiner et al., 2017). For studies which do not require a wide taxonomic range, the regions which discriminate well within the targeted species would be best regardless of their wider ability. For environmental DNA studies with highly degraded plant DNA, such as with ancient DNA, the non-coding plastid *trnL* (UAA) intron P6 loop has been used, as its short length (10-143 bp) can make it particularly suited (Taberlet et al., 2007). However, with this short length, species discrimination can again be limited unless curated reference databases are used (Fahner et al., 2016). This was demonstrated in a DNA metabarcoding study using *trnL* with soil samples from the tundra of Svalbard, an area with a small and intensively studied flora, where species discrimination was improved by knowing the availability of congeners within the system (Edwards et al., 2018).

For the data presented in Chapter 3, all honey samples were sequenced with both *rbcL* and ITS2, but only the *rbcL* results were presented in detail. During the initial analysis of the sequence results, *rbcL* was able to identify more taxa compared to ITS2 (*rbcL*: n = 137, ITS2: n = 37), and 34 of taxa identified by ITS2 were also found by *rbcL*. In this case, the ITS2 region was useful in providing additional support for using relative read abundance as a measure for the *rbcL* results but did not add species information to the overall picture of plants the honeybees were foraging on. In Chapter 4, both the *rbcL* and ITS2 results were used in combination. While again *rbcL* returned more plant taxa than ITS2 (*rbcL*: n = 121, ITS2: n = 84), ITS2 did return a more comparable number and provided additional plant taxa information (n = 36) which was not found by *rbcL.*

There are two main issues that face the ongoing development and use of DNA barcoding resources. Firstly, the need to increase the representation of species with high-quality DNA barcodes that are available in databases such as GenBank and BOLD (Barcode of Life Database). Secondly, the need to improve the taxonomic resolution that is achievable by using current and emerging technology and techniques, allowing finer scale questions to be answered.

To increase the amount of species with DNA barcodes and understand more about the species which are not described, BIOSCAN, a project by IBOL (International Barcode of Life Consortium) was announced in June 2019. It is a seven-year, $180 million project with the aim of sequencing 15 million single specimens, 90% from undescribed species. In addition, the project plans on metabarcoding 100,000 bulk sample collections from 2,000 sites across the world to study species interactions (IBOL, 2019). This takes advantage of the increasing output ability of sequencing for decreasing cost. The results here show that biologically meaningful results can be generated with the current level of species discrimination with plant DNA barcode regions. Future options for improving the species level discrimination achieved with DNA barcoding have been suggested, including additional amplicon sequencing from the nuclear genome and the use of entire plastid genomes (Hollingsworth et al., 2016).

One suggested method of achieving both organelle genome coverage and additional nucleotide coverage is genome skimming; the low-coverage, shotgun sequencing of genomic DNA (Coissac et al., 2016). The benefits of using genome skimming include that it is back compatible with DNA barcoding, as the standardised barcode regions can be recovered, in addition to being achievable with degraded DNA from herbarium specimens (Nieuwenhuis et al., 2015). The amount of data retrieved by genome skimming can allow for the near-complete assembly of the plastids and is being used with projects such as PhyloNorway and PhyloAlps to construct plastid genome reference libraries (Coissac et al., 2016; Parducci et al., 2019). Using the whole plastid can avoid problems arising from the use of different supplementary DNA regions from different research groups. In addition, it may allow for an increase in species discrimination. However, a remaining concern of targeting plastid genomes for improved species discrimination is that even with a greater sequence length, they still do not always effectively track the differences between species

(Hollingsworth et al., 2016, 2011). Chloroplasts are inherited maternally and chloroplast capture can occur unexpectedly through hybridisation and introgression with comparatively limited nuclear gene flow (Rieseberg and Soltis, 1991). In addition to this organelle sequence data, genome skims provide nuclear sequences. The low level of coverage from single copy nuclear DNA means that recovery of homologous sequences between samples is limited. This presents significant challenges for improved species discrimination as it requires the ability to compare between samples with a highly variable set of fragment DNA with highly variable overlap (Coissac et al., 2016; Hollingsworth et al., 2016). One approach may be to use all of the genomic information present without any assembly and compute the genomic distances between query genome skims, and reference genome skims (Sarmashghi et al., 2019). Before genome skims can be used for applications, high-quality reference libraries need to be constructed to improve the potential and accuracy of species identification (Parducci et al., 2019).

The continuing trajectory of DNA based identification may reach the level of whole genomes, especially paired with the ongoing decrease in costs of sequencing. For model organisms sequence reads can be mapped to high quality reference genomes (Berardini et al., 2015; The Arabidopsis Genome Initiative et al., 2000), while for non-model species reference genomes have been assembled *de novo* (Daccord et al., 2017). However, the limitations to the *de novo* assembly of whole genomes for plant species include the size of the genomes, which can vary widely (ca. 2,400 fold) and be very large, as well as the presence of repetitive regions of sequences (Pellicer et al., 2018). In addition, for the purposes of species identification and effective discrimination, the whole genome may supply a surplus of data beyond what is required to answer research questions and processing the results of whole genome sequencing requires additional bioinformatic support and resources compared to DNA metabarcoding.

The ability to identify species rapidly when out in the field is another ambition in the concept of DNA barcoding. Emerging technologies may be realising this goal, with the Bento Lab, which includes a thermocycler, centrifuge and gel electrophoresis equipment contained in a portable pack and the MinION sequencer, a portable long-read DNA sequencer which uses a USB port of a computer as the power supply (Runtuwene et al., 2019). Krehenwinkel et al., (2019) used the MinION platform in a field trial in rainforest in

Peru, performing the DNA extractions, PCR and library preparation in the field using portable equipment. Using ribosomal DNA they were able to identify eukaryote species, although there were taxonomic biases associated with sequencing from bulk community samples (Krehenwinkel et al., 2019).

5.1.3. The quantitative abilities of DNA metabarcoding

In this project, I have used DNA metabarcoding to characterise the key forage plants of honeybees, throughout the season and on a national scale. While DNA metabarcoding can confidently produce a qualitative list of taxa within a sample, it also provides a count of the DNA reads of each taxon and thus has the potential to provide quantitative results. However, throughout the process of DNA metabarcoding, from sampling, DNA extraction, PCR, and sequencing, biases are introduced (Alberdi et al., 2019; Zinger et al., 2019). How best to treat and interpret read count data is a source of debate between research groups (Bell et al., 2016; Lamb et al., 2019). There are two main approaches to presenting quantitative data. Often, the read count data is used to calculate the relative proportion of the taxa present in a sample, providing a semi-quantitative measure of the relative biomass contribution of the taxa from the original sample (Brennan et al., 2019; Erickson et al., 2017; Kartzinel et al., 2015; Richardson et al., 2018). The other approach, thought of as being more conservative, is to take the read count information and generate occurrence data for taxa, showing whether they are present or absent within a sample. However, using occurrence only data can overemphasise rare taxa in and across samples (Deagle et al., 2019).

Studies have tried to assess how quantitative the read abundance measure is with mock communities, where DNA mixes of known proportions are sequenced. Ideally, these would be made from starting tissues which are the same as the source of the sampled DNA, but due to the difficulty in homogenising tissue sources, mock communities are sometimes made from pre-existing DNA extracts. The results of comparisons using mock communities have ranged from finding no relationship between the composition of the mock community and the sequence results, to good correlations (Deagle et al., 2019). The variety of success in these comparisons are likely reflecting both biological differences between systems, but also technical differences, for example in the abilities of the

primers or the sequencing platform used. Overall, using a meta-analysis of papers, a weak quantitative relationship was found between the number of sequences produced and starting biomass in the system, albeit with a large degree of variation (Lamb et al., 2019). Similarly, Deagle et al., (2019) found, using simulations of sequencing data, that overall relative read abundance gave a more accurate picture of the starting composition compared with the proportion of samples where a species occurred, but with increased variance compared to the more consistent estimates from just occurrence measures.

Bell et al., (2019) tested the quantitative abilities of DNA metabarcoding using pollen mock communities with *rbcL* and ITS2. A significant but weak correlation was found between the relative abundance of pollen within a mixture and the proportion of reads for each species, for both markers, but they conclude that the proportion of starting pollen grains did not explain a satisfactory amount of the variance in the returned sequence number to use read abundance as a quantitative measure. Other studies have detected a relationship, such as Pornon et al., (2016) which found a positive relationship using mixtures from pollen DNA, with higher variability occurring with low DNA amounts.

In addition to mock communities, comparison with other methods of identifying the species present in a sample can be used to validate results. With pollen this is achieved by comparing DNA metabarcoding with morphological identification under a light microscope. When looking at the pollen loads from honeybees, studies have found a positive relationship between abundance estimates for the two methods (Keller et al., 2015; Richardson et al., 2018; Smart et al., 2017). In Richardson et al., (2018), the correlation in abundances between the two techniques was improved by taking the median abundance found across the four markers used in the study, although this restricted analysis to taxa which were found by all markers. Given that the universality of certain markers can be limited (Chapter 2), by using this technique information on biological relevant taxa will be lost. For honey samples, comparisons between DNA metabarcoding and melissopalynology found that the most abundantly found plant taxa were the most consistently detected between the two survey methods, whereas rare taxa were less likely to be detected by both surveys (Hawkins et al., 2015; Richardson et al., 2015).

In Chapter 3, the relative abundance of reads from the DNA metabarcoding was used to give a semi-quantitative measure of the contribution of each plant found within the honey. Using percentage of reads, we found a strong phenological signal for the plant species in the honey which would be more difficult to detect using only occurrence data, due to the low level carry over of plant species foraged from earlier in the season. Each sample was sequenced with *rbcL* and ITS2 and a correlation was found between the two markers for the relative abundance. Restricting the data to only species found by both *rbcL* and ITS2 would have created a less complete picture of honeybee foraging (22 taxa compared to 136).

In Chapter 4, using both *rbcL* and ITS2, the relative abundance of reads within a sample were placed into frequency classes, to allow comparison with the melissopalynology data from 1952 (Deans, 1958, 1957). The plant taxa returned at over 45% of reads in a sample were classed as predominant, between 15% and 45% were secondary, 1-15% were important minor taxa, and less than 1% of reads were classed as minor taxa. In order to allow the best comparison between the two surveys, the overall relationship was assessed based on the proportion of samples with a taxon present or absent. When comparing between the most frequently found forage sources, only the proportion of samples which were classified into predominant and secondary (>15%) were examined to reduce the impact of potentially increased variance related to rarity, as the most abundantly found taxa are the most consistently discovered between the two survey methods.

Future improved quantitative information from sequencing results may be achieved by using approaches that avoid a PCR amplification stage, such as genome skimming (Peel et al., 2019). However, the coverage of DNA reference libraries needs to be increased to improve plant identification with these methods (Parducci et al., 2019). Another method to potentially gain quantitative results from mixed DNA samples is to use species-specific real-time qPCR (quantitative PCR) assays, but this would be limited in studies where characterisation of the entire community is desired. It may be more appropriate when trying to detect specific rare species from a mixture, as in Harper et al., (2018), where detection of the target species, the great crested newt, was improved when using qPCR compared with DNA metabarcoding.

5.1.4. Honeybee foraging preferences throughout the season

In Chapter 3 even within a landscape rich with horticultural plant species, we show that honeybees are using a small number of plants throughout the season to supply most of their nutritional needs. This was also seen when examining spring forage (de Vere et al., 2017). The seasonality of nectar and pollen supply should be taken into consideration when providing planting recommendations, as well as contextualising with the requirements of the hive throughout the year. Earlier in the season (April and May), the plants found in honey tended to be tree species, reflecting similar patterns seen in the overall phenology of insect-pollinated plant species of the British flora, with 82% of tree species in flower in April and May (Balfour et al., 2018). During this time, the hive requires large quantities of pollen in order to support the growth and development of the larvae, with wild colonies estimated as containing 20,000 immature bees during spring, and each bee that is reared requiring around 130 mg of pollen (Seeley, 1995). The spring flowering tree species found in both honey surveys reflect this, with pollen rich, wind pollinated species such as, *Salix* spp., *Acer* spp., and *Quercus* spp. found in spring samples. Nectar requirements during this time are focused on providing the bees with energy, rather than for creating honey. Spring tree species that supply both pollen and nectar include, *Prunus* spp., *Sorbus* spp., and *Aesculus hippocastanum*. Top herbaceous plants early in the season include *Taraxacum officinale* and *Brassica* spp.

During early summer, from June to August, the honeybees are foraging for nectar to build up honey stores, while beekeepers are looking to benefit from honey production. The majority of DNA identified in the honey sampled from the Botanic Garden hives during this time was *Rubus* spp. and *Trifolium repens.* These taxa were also the top two most frequently found plants in the honey characterised across the UK. For both the UK sampling and the Botanic Garden, July represented the peak occurrence for *Rubus* spp. and *Trifolium repens*. Baude et al., (2016), using UK land cover data and estimates of nectar production for 260 plant species, estimated the seasonal peak in UK nectar production as July and August, with 60% of nectar being provided during these months. This coincides with the most common time for beekeepers to take honey off the hives. Due to the large area of improved grassland in the UK, allowing *T. repens* to flower within these areas offers the most potential for increasing nectar provision on a landscape scale

(Baude et al., 2016). In the case of *Rubus* spp., *Rubus fruticosus* agg. is a widespread plant found in habitats such as scrub, woodland, and hedge bank and in the context of land management, field margins and hedgerows are potential areas for gains in the nectar resource of this plant.

Late season is an important time for the colony, where they continue to build food reserves which will be needed for colony survival through the winter. In September, the top plant found was *Hedera helix,* a key autumn flowering nectar and pollen species (Garbuzov and Ratnieks, 2014)*. Calluna vulgaris* represented an autumn flowering species not found in the Botanic Garden survey but was found as a predominant forage in the UK survey. The plant *Impatiens glandulifera* was also a top species found in honey samples from August onwards. However, *I. glandulifera* is also a Schedule 9 species under the Wildlife and Countryside Act 1981 in England and Wales making it illegal to plant and grow in the wild. Planting lists therefore need to take into consideration the potential ecological impact of any plant species recommended.

Compared to the number of flowering plants available in the landscape for the honeybees, only a small percentage of these were being used as forage and a smaller proportion of these were present as major forage. Despite the most frequently found species tending to be widespread, native plants available in abundant floral events, such as *Rubus* spp., *Salix* spp., *Crataegus monogyna, Prunus* spp. or *Taraxacum officinale*, no relationship was found between the percentage cover of each plant found in the study and the abundance within the honey. This indicates that the honeybees are choosing plants based on factors in addition to their abundance within the landscape.

One of the key continuing research questions is why honeybees use these plants and how it relates to both the availability in the landscape and how honeybees may be seeking out the plants that supply them with specific nutritional needs. The plants identified here by DNA metabarcoding can be used as a starting point for relating to further nectar and pollen analysis. While the sugar concentration, composition and rate of production within flowers has often been a focus of nutritional analysis for pollinators, the relative quality of the sources of pollen available may be having a larger impact on long term health of a hive, given the importance of protein to healthy brood growth and development and its

longer term impact on the colony (Alaux et al., 2010; Brodschneider and Crailsheim, 2010; Filipiak, 2019). While chemical analysis of freshly collected pollen is possible, the quantities required are often difficult to achieve from what is available from flowers. Studies often focus on pollen loads collected by bees from pollen traps, which may lead to an underestimation of the protein composition due to the bees packing the pollen with nectar or saliva (Wright et al., 2018). Future work could look at incorporating measures of floral reward into the abundance data recorded here, including a finer measurer of floral abundance on the plant, and measurements of nutritional reward for pollen and nectar.

Honeybees can be considered as a potential indicator species for landscape health, due to both their large foraging range and their generalist foraging compared to other bees (Couvillon and Ratnieks, 2015). In Chapter 4, I showed how honeybees can be used to assess foraging on a national scale, benefitting from the wide distribution of honeybees within the landscape, as well as data from the honey representing the forage effort of multiple honeybees over a longer time period, compared with sampling individual insects. However, honeybees do differ in comparison to other wild pollinators such as solitary bees, bumblebees, and hoverflies. While their large foraging range allows an insight into the forage of the wider landscape, other pollinators will not be able to reach resources at the same scale. Their larger forage range comes from their large social structure coupled with their ability to communicate forage using waggle dances. In addition to behavioural differences, they will have physiological differences to other pollinators, including aspects such as tongue length, which can restrict access to certain known important sources of forage. For example, both *Trifolium repens* and *Trifolium pratense* were detected in this study, with *T. repens* being much more frequently found than *T. pratense*. *T. pratense* is less used by honeybees due to its longer corolla and is more associated with long-tongued bumblebees (Comba et al., 1999).

The DNA metabarcoding techniques developed here have excellent scope to be expanded to other pollinators within the same study system. This study based in a botanic garden showed that despite a wide variety of plant species on offer, honeybees foraged on a comparatively small number of plants. By looking at wild pollinators within the same system, a larger picture of pollination networks within a diverse habitat can be established. Using the DNA metabarcoding techniques developed here, the pollen from

the insect's body or their pollen baskets can be identified, as with hoverflies in species rich grasslands (Lucas et al., 2018), and bees in urban planted wildflower mixes (Potter et al., 2019).

5.1.5. Planting for pollinators

There is an appetite for pollinator friendly gardening recommendations from both beekeepers and the general public, with 77% of garden centre customers wanting to help pollinators by growing pollinator-friendly plants (Wignall et al., 2019). Chapter 3 investigated the foraging of honeybees within a diverse landscape, focusing on a landscape spatial scale, and looked at how foraging changed throughout the season. In Chapter 4, honeybee foraging was assessed on a national scale, throughout the UK, and there are recommendations to be made from both these studies together.

Many of the top forage plants represent native species that people would not necessarily choose to plant within their garden environments, such as *Rubus* spp., *Trifolium repens*, or *Taraxacum officinale. S*ome represent species that can only be planted with great caution such as a *Cotoneaster* species like *C. horizontalis* while some, like *Impatiens glandulifera,* should not be planted at all. These results provide a scientific backing for encouragement from pollinator and plant conservation schemes to let mown green spaces grow and flower such as the charity Plantlife's "No Mow May" campaign, launched in 2019. This management benefits species such as *T. repens* and *Taraxacum officinale* (Lerman et al., 2018).

The tree species found in the honey represent an opportunity for planting in a variety of contexts: for key hedgerow species, such as *Crataegus monogyna;* for gardeners with often ornamental species such as *Prunus* spp.; and for urban planting with trees such as *Salix* spp., *Acer* spp. and *Tilia* spp.

Figure 26: Illustrating some of the top forage plants found by this study. Clockwise from the top left: *Trifolium repens*, *Rubus* sp., *Brassica napus*, *Impatiens glandulifera*, and *Prunus spinosa*.

5.1.6. Top plants used throughout the season

Table 7: The top forage plants found throughout the season identified using DNA metabarcoding of honey throughout the season from hives set within a botanic garden and from 441 honey samples from across the UK.

April	May	June	July	August	September
Salix	Maleae e.g. *Crataegus monogyna*	*Rubus*	*Rubus*	*Rubus*	*Hedera helix*
Prunus	*Taraxacum officinale*	*Trifolium repens*	*Trifolium repens*	*Trifolium repens*	*Impatiens glandulifera*
Taraxacum officinale	*Acer*	*Cirsium*	*Cirsium*	*Impatiens glandulifera*	*Calluna vulgaris*
Brassica	*Brassica*	*Vicia*	*Filipendula ulmaria*	*Calluna vulgaris*	
Ulex	*Sorbus*	*Rosa*	*Vicia*	*Filipendula ulmaria*	

5.1.7. Overall conclusions

This work aimed to provide a comprehensive picture of honeybee foraging in the UK, assessing the seasonal forage plants and placing this into a national context, providing an important evidence base for the plant species most important to honeybees. In doing so the major forage plants for honeybees throughout the season were revealed showing the importance of native plants and semi-natural habitats to honeybees. This information can aid beekeepers in understanding how best to support their hives, as well as providing gardeners and landowners with recommendations for planting and management. To support this and future DNA barcoding applications, a reference library for the UK native flora has been created, representing a vital resource for a wide range of studies that require identification of plants from unknown material.

5.2. References

Alaux, C., Ducloz, F., Crauser, D., Le Conte, Y., 2010. Diet effects on honeybee immunocompetence. *Biology Letters* **6**, 562–565.

Alberdi, A., Aizpurua, O., Bohmann, K., Gopalakrishnan, S., Lynggaard, C., Nielsen, M., Gilbert, M.T.P., 2019. Promises and pitfalls of using high-throughput sequencing for diet analysis. *Molecular Ecology Resources* **19**, 327–348.

Balfour, N.J., Ollerton, J., Castellanos, M.C., Ratnieks, F.L.W., 2018. British phenological records indicate high diversity and extinction rates among late-summer-flying pollinators. *Biological Conservation* **222**, 278–283.

Baude, M., Kunin, W.E., Boatman, N.D., Conyers, S., Davies, N., Gillespie, M.A.K., Morton, R.D., Smart, S.M., Memmott, J., 2016. Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature* **530**, 85–88.

Bebber, D.P., Carine, M.A., Wood, J.R.I., Wortley, A.H., Harris, D.J., Prance, G.T., Davidse, G., Paige, J., Pennington, T.D., Robson, N.K.B., Scotland, R.W., 2010. Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences* **107**, 22169–22171.

Bell, K.L., Burgess, K.S., Botsch, J.C., Dobbs, E.K., Read, T.D., Brosi, B.J., 2019. Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology* **28**, 431–455.

Bell, K.L., de Vere, N., Keller, A., Richardson, R.T., Gous, A., Burgess, K.S., Brosi, B.J., 2016. Pollen DNA barcoding: current applications and future prospects. *Genome* **59**, 629–640.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E., 2015. The *Arabidopsis* information resource: making and mining the “gold standard” annotated reference plant genome. *genesis* **53**, 474–485.

Brennan, G.L., Potter, C., de Vere, N., Griffith, G.W., Skjøth, C.A., Osborne, N.J., Wheeler,

B.W., McInnes, R.N., Clewlow, Y., Barber, A., Hanlon, H.M., Hegarty, M., Jones, L., Kurganskiy, A., Rowney, F.M., Armitage, C., Adams-Groom, B., Ford, C.R., Petch, G.M., Creer, S., 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution* **3**.

Brodschneider, R., Crailsheim, K., 2010. Nutrition and health in honey bees. *Apidologie* **41**, 278–294.

Christenhusz, M.J.M., Byng, J.W., 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201.

Coissac, E., Hollingsworth, P.M., Lavergne, S., Taberlet, P., 2016. From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology* **25**, 1423–1428.

Comba, L., Corbet, S.A., Barron, A., Bird, A., Collinge, S., Miyazaki, N., Powell, M., 1999. Garden Flowers: Insect Visits and the Floral Reward of Horticulturally-modifed Variants. *Annals of Botany* **83**, 73–86.

Couvillon, M.J., Ratnieks, F.L.W., 2015. Environmental consultancy: dancing bee bioindicators to evaluate landscape "health." *Frontiers in Ecology and Evolution* **3**.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C., Bik, H.M., 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution* **7**, 1008–1018.

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E.A., Gouzy, J., Rees, D.J.G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., Aubourg, S., Quesneville, H., Weigel, D., van de Weg, E., Troggio, M., Bucher, E., 2017. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49**, 1099–1106.

De Boer, H.J., Ghorbani, A., Manzanilla, V., Raclariu, A.C., Kreziou, A., Ounjai, S., Osathanunkul, M., Gravendeel, B., 2017. DNA metabarcoding of orchid-derived

products reveals widespread illegal orchid trade. *Proceedings of the Royal Society B: Biological Sciences* **284**.

de Vere, N., Jones, L.E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M.J., Creer, S., Ford, C.R., 2017. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports* **7**, 42838.

Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L., Kartzinel, T.R., Eveson, J.P., 2019. Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology* **28**, 391–406.

Deans, A.S.C., 1957. *Survey of British Honey Sources*. Bee Research Association.

Deans, A.S.C., 1958. The Pollen Analysis of some British Honeys. Thesis for the National Bee Diploma.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* **26**, 5872–5895.

Edwards, M.E., Alsos, I.G., Yoccoz, N., Coissac, E., Goslar, T., Gielly, L., Haile, J., Langdon, C.T., Tribsch, A., Binney, H.A., Von Stedingk, H., Taberlet, P., 2018. Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. *Holocene* **28**, 2006–2016.

Erickson, D.L., Reed, E., Ramachandran, P., Bourg, N.A., McShea, W.J., Ottesen, A., 2017. Reconstructing a herbivore's diet using a novel *rbcL* DNA mini-barcode for plants. *AoB PLANTS* **9**.

Fahner, N.A., Shokralla, S., Baird, D.J., Hajibabaei, M., 2016. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four DNA markers. *PLoS ONE* **11**, 1–16.

Filipiak, M., 2019. Key pollen host plants provide balanced diets for wild bee larvae: A lesson for planting flower strips and hedgerows. *Journal of Applied Ecology* **56**, 1410–

1418.

Garbett, H.L., 2016. Development of Bioinformatic Tools and Procedures to Computationally Support DNA barcoding. University of South Wales.

Garbuzov, M., Ratnieks, F.L.W., 2014. Ivy: An underappreciated key resource to flower-visiting insects in autumn. *Insect Conservation and Diversity* **7**, 91–102.

Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S., Hänfling, B., 2018. Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution* **8**, 6330–6341.

Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., Baillie, L., Adams-Groom, B., 2015. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. *PLOS ONE* **10**.

Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T.A.B., Rohwer, J.G., Campbell, C.S., Chatrou, L.W., 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* **90**, 1758–1776.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. *PloS ONE* **6**.

Hollingsworth, P.M., Li, D., van der Bank, M., Twyford, A.D., 2016. Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150338.

IBOL, 2019. Scientific abstracts from the 8th International Barcode of Life Conference. *Genome* **62**, 349–453.

Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W., Pringle, R.M., 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National*

Academy of Sciences **112**, 8019–8024.

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., Ohe, K. Von Der, Ohe, W. Von Der, Rost, S., Hartel, S., Steffan-Dewenter, I., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology* **17**, 558–566.

Krehenwinkel, H., Pomerantz, A., Henderson, J.B., Kennedy, S.R., Lim, J.Y., Swamy, V., Shoobridge, J.D., Graham, N., Patel, N.H., Gillespie, R.G., Prost, S., 2019. Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience* **8**, 1–16.

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G., Taylor, M.I., 2019. How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* **28**, 420–430.

Lerman, S.B., Contosta, A.R., Milam, J., Bang, C., 2018. To mow or to mow less: Lawn mowing frequency affects bee abundance and diversity in suburban yards. *Biological Conservation* **221**, 160–174.

Lim, J., Crawley, M.J., de Vere, N., Rich, T., Savolainen, V., 2014. A phylogenetic analysis of the British flora sheds light on the evolutionary and ecological factors driving plant invasions. *Ecology and Evolution* **4**.

Lucas, A., Bodger, O., Brosi, B.J., Ford, C.R., Forman, D.W., Greig, C., Hegarty, M., Jones, L., Neyland, P.J., de Vere, N., 2018. Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding. *Scientific Reports* **8**.

Nieuwenhuis, M., Mohammadin, S., Alquezar-Planas, D.E., Lei, D., Yu, J., Gravendeel, B., van de Kerke, S., Holmer, R., Wei, Z., Bakker, F.T., Staats, M., 2015. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. *Biological Journal of the Linnean Society* **117**, 33–43.

Parducci, L., Alsos Greve, I., Unneberg, P., Pedersen, M.W., Han, L., Lammers, Y., Salonen, S.J., Väliranta, M.M., Slotte, T., Wohlfarth, B., 2019. Shotgun ancient DNA, pollen and macrofossil analysis of Lateglacial lake sediments from southern Sweden. *Frontiers in Ecology and Evolution* **7**, 189.

Peel, N., Dicks, L. V., Clark, M.D., Heavens, D., Percival-Alwyn, L., Cooper, C., Davies, R.G., Leggett, R.M., Yu, D.W., 2019. Semi-quantitative characterisation of mixed pollen samples using MinION sequencing and Reverse Metagenomics (RevMet). *Methods in Ecology and Evolution* 2041–210X.13265.

Pellicer, J., Hidalgo, O., Dodsworth, S., Leitch, I.J., 2018. Genome size diversity and its impact on the evolution of land plants. *Genes* **9**.

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Pellizzari, C., Iribar, A., Etienne, R., Taberlet, P., Vidal, M., Winterton, P., Zinger, L., Andalo, C., 2016. Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports* **6**, 1–12.

Potter, C., de Vere, N., Jones, L., Ford, C.R., Hegarty, M.J., Hodder, K.H., Diaz, A., Franklin, E.L., 2019. Pollen metabarcoding reveals broad and species-specific resource use by urban bees. *PeerJ* **7**.

Richardson, R.T., Curtis, H.R., Matcham, E.G., Hua Lin, C., Suresh, S., Sponsler, D.B., Hearon, L.E., Johnson, R.M., 2018. Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. *Molecular Ecology* **28**, 686–697.

Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015. Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. *Applications in Plant Sciences* **3**.

Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**, 65–84.

Runtuwene, L.R., Tuda, J.S.B., Mongan, A.E., Suzuki, Y., 2019. On-Site MinION Sequencing.

In: Advances in Experimental Medicine and Biology. pp. 143–150.

Sarmashghi, S., Bohmann, K., Gilbert, P.M.T., Bafna, V., Mirarab, S., 2019. Skmer: Assembly-free and alignment-free sample identification using genome skims. *Genome Biology* **20**, 1–20.

Seeley, T.D., 1995. *The Wisdom of the Hive*. Harvard University Press.

Smart, M.D., Cornman, R.S., Iwanowicz, D.D., McDermott-Kubeczko, M., Pettis, J.S., Spivak, M.S., Otto, C.R.V., 2017. A comparison of honey bee-collected pollen from working agricultural lands using light microscopy and its metabarcoding. *Environmental Entomology* **46**, 38–49.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**, e14.

The Arabidopsis Genome Initiative, Kaul, S., Koo, H.L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L.J., Feldblyum, T., Nierman, W., Benito, M.I., Lin, X.Y., 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.

Wignall, V.R., Alton, K., Ratnieks, F.L.W., 2019. Garden centre customer attitudes to pollinators and pollinator-friendly planting. *PeerJ* **7**, e7088.

Wright, G.A., Nicolson, S.W., Shafir, S., 2018. Nutritional Physiology and Ecology of Honey Bees. *Annual Review of Entomology* **63**, 327–344.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., De Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., Kauserud, H., Orlando, L., Pansu, J., Pawlowski, J., Tedersoo, L., Thomsen, P.F., Willerslev, E., Taberlet, P., 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* **28**, 1857–1862.

APPENDIX A

BeeCraft Article: Honeybee foraging, new techniques to barcode the natural world

NEW TECHNIQUES TO BARCODE THE NATURAL WORLD

Honey Bee Foraging

Laura Jones

Laura Jones

Artwork created by Stitching Botanicals, each plant found at the garden was stitched by the group who are based at the Botanic Garden.

Honey bees vitally contribute to humans both as a critically important pollinator of wild and crop plants, and by providing honey and wax products. The ease with which honey bees can be managed, compared with other pollinators, makes them crucial to crop pollination.

There is therefore, considerable concern worldwide over the vulnerability of honey bee colonies to the interacting effects of pests and diseases, apicultural mismanagement, climate change and agricultural intensification which results in the loss and fragmentation of habitat.

Leading the World

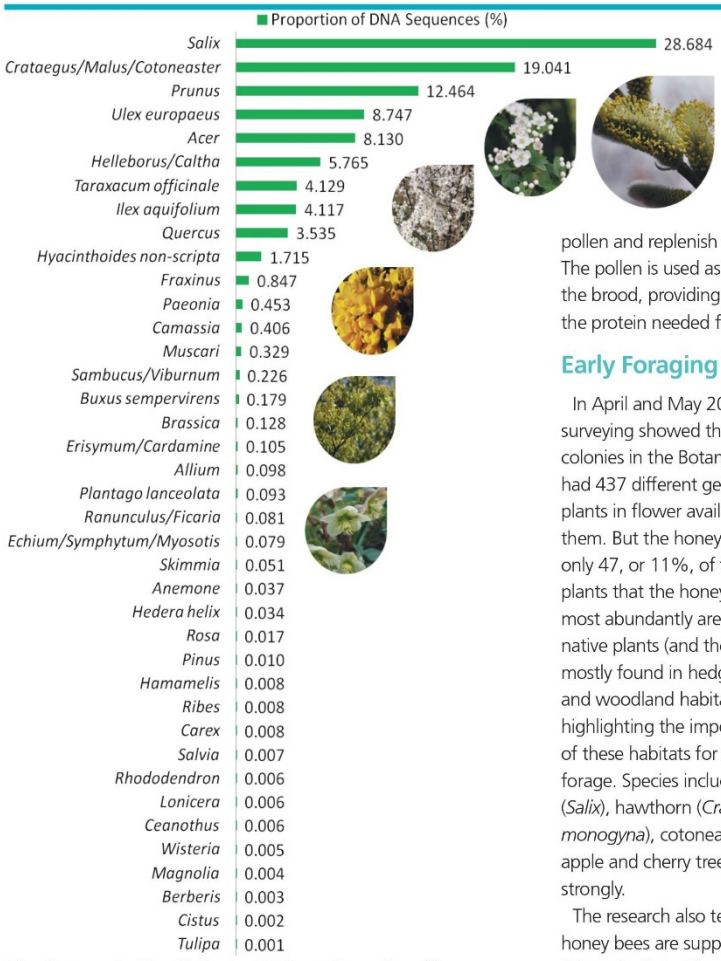
The National Botanic Garden of Wales is interested in understanding which flowers honey bees forage to help indicate suitable plants for healthy honey bee colonies. Our research paper forms the beginnings of my PhD research with the Botanic Garden and Bangor University, and is funded by Knowledge Economy Skills Scholarships (KESS).

We have been using the Botanic Garden and its own beehives as a study site, collecting honey samples for analysis on a monthly basis throughout the season. The Botanic Garden provides a rich foraging resource for pollinators, with over 8,000 plant taxa in our horticultural plantings, in addition to being surrounded by native habitat in our National Nature Reserve. The Botanic Garden also made Wales the first nation in the world to DNA barcode every native flowering plant, a total of 1,143 species, and it's this DNA barcoding expertise that is used in my work.

Honey and DNA Barcoding

How do we find out the plants that make up honey? By extracting the pollen from honey, we can use DNA barcoding to identify the plants it came from. This uses short regions (small segments) of DNA that are unique to a species, called DNA barcodes. DNA is made up from just four bases, represented by the letters A, T, C and G. It is the differences between the order of these bases that allows us to identify their source and therefore which plants bees were visiting to make the honey in the hives. Using these DNA techniques allows us to identify plants where traditional identification by appearance would not be possible.

Each DNA barcode is stored in the Barcode of Life Database (BOLD) where it is available to view and use by everyone. Each record contains the DNA



The plant species identified using DNA metabarcoding of honey, ordered by proportion of DNA (%) for three colonies sampled in April and May 2015.

sequence as well as information on the plant individual the DNA came from, including a scan of the sampled herbarium specimen.

Surveying the Botanic Garden

DNA barcoding our honey gives us the plants the bees are visiting, but how do we know the plants that were in flower and available to the bees? That is where our survey comes in. Once a month, to match our honey sampling, we survey all of the plants in flower in the Botanic Garden. We have the support of a fantastic group of conservation volunteers who contribute to our dataset and research by helping us to record the plants in flower. Each month between April and September we sample honey from each of our hives and record everything in flower. Presented above are our first results for April and May 2016.

Spring is a crucial time for honey bees. Over winter, they will use up their stores of honey. As winter progresses, they will start to produce brood and, as soon as the sun comes out in early spring, they need to gather pollen and replenish their stores. The pollen is used as food for the brood, providing them with the protein needed for growth.

Early Foraging Results

In April and May 2016 our surveying showed that the colonies in the Botanic Garden had 437 different genera of plants in flower available to them. But the honey bees used only 47, or 11%, of these. The plants that the honey bees used most abundantly are nearly all native plants (and their relatives) mostly found in hedgerow and woodland habitat, highlighting the importance of these habitats for spring forage. Species including willow (*Salix*), hawthorn (*Crataegus monogyna*), cotoneaster, and apple and cherry trees featured strongly.

The research also tells us that honey bees are supplementing this main diet with smaller amounts from garden plants, showing that what we do in our own backyard is crucial. The plants that the honey bees used less tended to be horticultural species, including peonies (*Paeonia*), camassia, and grape hyacinth (*Muscari*).

Continuing Work

Now that we have the early foraging results, my work will track which plants honey bees use for the continuing months and how that varies from year to year.

We are also inviting beekeepers to send us samples of honey from the 2017 season to find out the plants used by honey bees throughout different habitats in Wales and the UK, and how this relates to hive health.

Science and Society

The National Botanic Garden of Wales is always working to engage, educate and collaborate with as wide a range of people as possible. As part of this, we often collaborate with artists to create artworks that interpret our scientific research.

The early foraging results presented here were illustrated by a group called the Stitching Botanicals, who are based at the Garden. Each plant found in the honey was meticulously stitched onto a hexagon, creating a moveable and adjustable illustration of our work that we use to engage people with our research and honey bees.

If you are interested in contributing honey to the project please contact laura.jones@gardenofwales.org.uk.

The full research paper and list of plants used by the honey bees is published in the open access journal Scientific Reports available at www.nature.com/articles/srep42838.

Knowledge Economy Skills Scholarships is a pan-Wales higher-level skills initiative led by Bangor University on behalf of the higher education sector in Wales. It is part-funded by the Welsh Government's European Social Fund convergence programme.

Barcode of Life Database: www.boldsystems.org

APPENDIX B

Using DNA metabarcoding to investigate honeybee foraging reveals limited flower use despite high floral availability

doi: 10.1038/srep42838

[TEXT REMOVED]

APPENDIX C

Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding

doi: 10.1038/s41598-018-23103-0

[TEXT REMOVED]

APPENDIX D

Pollen metabarcoding reveals broad and species-specific resource use by urban bees

doi: 10.7717/peerj.5999

[TEXT REMOVED]

APPENDIX E

Temperate airborne grass pollen defined by spatio-temporal shifts in community composition

doi: 10.1038/s41559-019-0849-7

[TEXT REMOVED]

APPENDIX F

Conference presentations, art exhibitions and media

5.3. Conferences

June 2019: Oral presentation at the 8th International Barcode of Life Conference, Trondheim, Norway. *Using pollen DNA metabarcoding to investigate floral visitation by honeybees and wild pollinators*.

May 2018: Oral presentation at the 5th European Congress of Conservation Biology, Jyväskylä, Finland. *Using DNA metabarcoding to investigate the foraging of honeybees*. Helped lead a technical workshop on environmental DNA in conservation management.

November 2017: Oral presentation at the 7th International Barcode of Life Conference, Kruger, South Africa. *Using pollen DNA metabarcoding to investigate the foraging preferences of honeybees*.

May 2016: Oral presentation at the Congenomics Conference, Porto, Portugal. *Investigating honeybee foraging using DNA metabarcoding*.

5.4. Art exhibitions

2019. *Stitching Botanical 'Brood Frame'*. The Stitching Botanical group took the results from Chapter 3 and created a textile installation showing a honeybee brood frame with the different plants illustrated.

2017. *Cross-Pollination. Revaluing Pollinators through Arts and Science Collaboration.* Led by Prof Andrea Liggins and Sarah Tombs. Funded by the Arts and Humanities Research Council (AHRC). Scientists and artists worked together to explore the decline of pollinating insects and what can be done about it. The artists created responses to this in a variety of forms, including photography, sculpture and textiles. The installations were exhibited at the National Botanic Garden of Wales, Sarah Beynon's Bug Farm and at Cornell University.

2016. *Stitching Botanical. Illustrating the early season foraging of honeybees based on results from DNA metabarcoding.* The Stitching Botanical group at the National Botanic Garden of Wales took the results of the paper in Appendix B and illustrated each of the plants (Figure 27). The installation has been exhibited at shows and events throughout the UK.

2015. *Barcode Sculpture*. Blockley, K., Howes, J., Jackson, D., Kincaid, P., Lane, M., Mably, L., Morris, G., Spowers, A., ISBN 978-0-956583-4-1. The group of sculptors spent time in the labs at the National Botanic Garden of Wales finding out about DNA barcoding. A series of sculptures were created in response to this and exhibited at the National Botanic Garden of Wales.

Figure 27: The Stitching Botanical's work illustrating the early season honeybee forage plants. It is designed to be portable, allowing it to be displayed at events, as seen here at the Royal Welsh Show horticultural tent.

5.5. Media

2017: BBC Gardeners' World. Feature on my honeybee research at the Botanic Garden. Beekeepers were asked to send honey samples for DNA metabarcoding to support my PhD study (Figure 28).

2017: BBC News. Bees' favourite plants revealed by Botanic Garden study. Available at: https://www.bbc.co.uk/news/uk-wales-south-west-wales-39003201

2017: BBC Countryfile. Feature discussing our ongoing honeybee research and the early foraging results

Figure 28. Filming during Gardeners' World to publicise the UK honey survey.