Using a nanopore sequencing approach to investigate the pollen diet of bumblebees in an agricultural landscape

Eleanor Sophia Kent

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia
School of Biological Sciences

April 2023

Abstract

Pollinators play an important role in ecosystem functioning and pollination services, which is especially true of agricultural landscapes that are suffering biodiversity losses due to land intensification, chemical use and reduction of natural habitats. The mechanisms by which foraging occurs can be better understood by identifying the floral resources most commonly used by pollinators using DNA sequencing. A whole genome sequencing approach, 'Reverse Metagenomics', has previously shown evidence of semi-quantitative characterisation of mock DNA mixes. Here, RevMet is applied to pollen collected from foraging commercial bumblebees (*Bombus terrestris*) on farms growing highbush blueberry (*Vaccinium corymbosum*) and compared to microscopic proportions. The RevMet results provide evidence of qualitative and quantitative abilities, which are further tested comparatively using nanopore-sequenced ITS2 amplicons derived from mock pollen mixtures. The quantitative RevMet results from this study show a potential bias from plant genome size that can be improved using a correction factor. To apply RevMet in a wider ecological context, pollen loads are sequenced from foraging commercial bumblebees located on four farms growing highbush blueberry in southern England over the crop flowering season. Here, the most utilised plant taxa are revealed, providing recommendations for landowners of attractive foraging resources that can be used to inform landscape level decisions. Overall, this work demonstrates an improved understanding of the RevMet approach when applied to pollen loads, and provides evidence of the floral community most commonly used by commercial bumblebees on UK farms growing highbush blueberry crops.

Acknowledgements

I would like to start by thanking my primary supervisor, Lynn, for her support and guidance throughout the PhD. I would also like to thank my secondary supervisors Doug and Richard for their help and advice over the years. Richard, you welcomed me into the Technology Algorithms Group at the Earlham Institute and provided me with the opportunity to work in your lab, for which I am extremely grateful.

The molecular work was carried out at the Earlham Institute, so I'd like to thank all the staff and researchers who helped me along the way. I couldn't have done it without Darren, who was a limitless source of patience and experience, and taught me everything I know about working with a lot of pollen and tiny amounts of DNA. You were, as always, right in saying we stand on the shoulders of giants. Another huge thank you to Ned, the creator of RevMet, for answering my many questions and allowing me to pick his brains and ask for advice on the bioinformatics, I'm very grateful for all your help.

Additional thanks to Emma, Derik, BerryWorld and the farm managers for providing access to field sites and accommodation, particularly when COVID-19 made movement more challenging.

To the wider Agro-Ecology group at UEA and Cambridge, it's been a pleasure getting to know everyone. A special thanks to Midge and Nat, who were a huge source of support, both academically and personally, I'm so glad we got to share this experience. Another thank you to Katie, who gave me some excellent pieces of advice during the last months of the PhD.

I have met so many lovely people during my time at UEA, but a special thank you to the members of the Gherkin office and the folk on Floor 01. Particular shout outs to Midge, Chris, Enya, Leigh and Tom, who have made this time a lot of fun, even with a global pandemic halfway through.

I couldn't have done any of this without an immense amount of support from friends and family. First, to my friends who, over the years, have helped me more than words can describe - you never know what life's going to throw at you. Fee, you especially have been an amazing source of positivity and laughter.

My family has continuously inspired my love of nature and the outdoors. I would like to say a particular thank you to Mum and Dad for supporting me in following a path that has taken me across the world. You've made me believe all of this was exciting and possible.

Finally, I would like to dedicate this thesis to Anna Crossman, who I wish more than anything was here to celebrate this achievement with me.

Table of Contents

List of Figures

List of Tables

List of Supplementary Figures

List of Supplementary Tables

Chapter One

Introduction

1.1 Pollinators in agriculture

Animal pollinators are of vital importance to ecosystem services, as well as ecological biodiversity and agriculture. Pollinators, which include insects, birds, bats and reptiles, are responsible for the pollination of 90% of flowering plants (Kearns et al., 1998). An estimated 70% of the 124 main crops grown for human consumption, including vegetables, fruits and nuts, are animal pollinated to some degree (Klein et al., 2007). Bees are believed to be the most important pollinators, with a global contribution to the production of crops for human consumption worth an estimated US$150 billion (Gallai et al., 2009). Of the 20,000 species of bee worldwide, an estimated 12% contribute to crop pollination (Kleijn et al., 2015; Lautenbach et al., 2012).

Agricultural production that is dependent on animal pollination has increased by 300% over the last 60 years (Aizen & Harder, 2009). The increase in global managed honey bee (*Apis mellifera*) colonies has not been at the same rate as the increase in crop production, so it is likely that honey bees cannot meet this demand. This, coupled with a steadily increasing global population, estimated to reach almost 10 billion by 2050 (UN DESA, 2022), could pose a serious threat to future food security.

However, there may not need to be such a reliance on honey bees, as wild pollinators have proved to be more efficient at pollinating certain crops (Garibaldi et al., 2013). There is evidence to suggest that enhancing wild pollinator diversity and abundance can increase seed production (Greenleaf & Kremen 2006), fruit weight (MacInnis & Forrest 2019), fruit set (Garibaldi et al., 2013), pollen transfer (Woodcock *et al.* 2013) and pollination effectiveness (Javorek et al., 2002). Furthermore, the presence of commercial bees can have a negative impact on local populations of wild bees (Goulson 2003). Commercial bees, such as honey bees and some species of bumblebee, compete with wild bees for resources such as pollen, nectar and nesting sites. An added pressure is the prevalence of disease that occurs in densely populated areas of managed bees that can spread to wild populations (Furst et al., 2014). A systematic review of 146 studies on the effects of managed bees on wild pollinators

concluded variable results (Mallinger et al., 2017). In terms of resource competition, 55% of the studies reported a negative impact of managed honey bees, depending on resource availability and landscape heterogeneity. Competitive effects between managed and wild bees were found to be relatively local (<800m to the nest), suggesting that at a landscape level in a heterogeneous environment, it is possible that wild and managed populations of bees could coexist without negative impacts.

In terms of ecological importance, it has been suggested that a diverse population of pollinators is less critical for crop production (Kleijn et al., 2015; Senapathi et al., 2021). It is argued that a few species provide the majority of the pollination services, so should not be used in arguments in support of biodiversity conservation. Kleijn et al. (2015) used data from 90 studies and concluded that 80% of crop pollination was provided by 2% of bee species.

This study contradicts a multitude of research that suggests species-rich communities of bees are most important in crop production (Carvalheiro et al., 2011; Garibaldi et al., 2013, 2016; Winfree et al., 2018). Winfree et al. (2018) suggest the value of a diverse community of pollinators is underestimated at the landscape scale and the effects of species dominance and turnover cannot be recognised in field level studies. Highly diverse communities significantly increase seed and fruit set in a range of crops (Klein et al., 2003; Mallinger & Gratton, 2015; Rogers et al., 2014; Senapathi et al., 2021). A diverse flower-visitor density was found to be the most important predictor of crop yield across 344 fields of 33 crop systems in Africa, Asia and Latin America (Garibaldi *et al.* 2016). Species-rich populations of wild bee also increase the resilience of pollination services, that is the ability of a pollinator community to maintain pollination function under environmental disturbances. Functionally similar species do not respond in the same way to environmental changes, so a diversity of species provides a buffer against fluctuations (Brittain et al., 2013; Senapathi et al., 2021; Vicens & Bosch, 2000).

1.2 Drivers of pollinator loss

Although the number of commercial honey bee colonies is on the rise, wild pollinators are experiencing a severe decline (Potts et al., 2016). Wild honey bees have decreased by 62% in the same period (60 years) in the US and wild non-honey bees have experienced a loss in numbers of 23% over just a five-year period, from 2008 to 2013 (Fox et al., 2021; Powney et al., 2019; van Engelsdorp & Meixner, 2010). In the United Kingdom, 23 pollinating bee and wasp species have been declared extinct since 1850 (Ollerton et al., 2014).

The main drivers of loss include habitat destruction and fragmentation, pesticides, climate change and disease (Goulson et al., 2015; Potts et al., 2010). Pollinators require diverse habitats, including flowering plants, for nesting sites and shelter. Human activities including deforestation, urbanisation, and intensive agriculture, have removed natural habitats and those remaining become increasingly fragmented, limiting the resource availability for pollinators (Winfree et al., 2009)

The application of chemicals in agricultural ecosystems is a significant driver of pollinator decline. Chemicals used for herbicides, pesticides and insecticides are sprayed onto crops, but also harm pollinators directly or indirectly. Neonicotinoids, a popular class of insecticides, were found to increase mortality, and reproduce reproductive output in bumblebees and solitary bees (Rundlöf et al., 2015; Whitehorn et al., 2012). What's more, chemicals can accumulate in the environment for many years, posing long term risks to pollinators and other non-target species (Goulson, 2013).

Global climate change is causing increasing fluctuations in temperature and extreme weather patterns, with heatwaves becoming more common events (Ummenhofer & Meehl, 2017). Temperatures over 40°C can kill insects (Martinet et al., 2015; Rasmont & Iserbyt, 2012), limit bee reproduction (Martinet et al., 2021) and negatively affect foraging activity (Hemberger et al., 2023). Increasing temperatures are also altering the timing of flowering plants. This phenological shift has been found to cause a mismatch in resource availability

for pollinators, which are emerging later than the flowers and disrupting the timing of pollination events (Memmott et al., 2007).

Pollinators are susceptible to disease and parasites, which have been found to reduce their lifespan and reproductive output (Brown et al., 2003). The introduction of managed bees into ecosystems has been found to increase pathogen spill over to wild conspecifics and solitary bee communities (Fürst et al., 2014; Murray et al., 2013; Ravoet et al., 2014). Therefore care should be taken not to overpopulate natural areas with high densities of managed bees.

1.3 Mitigating against pollinator decline

The Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services (IPBES) published an 'Assessment Report on Pollinators, Pollination and Food Production' (IPBES, 2016). The report outlines the threats to pollinators and pollination, and practical solutions to mitigating these threats. One of the major threats the report discusses is agricultural intensification, as isolation from florally diverse habitats has been shown to decrease pollination events (Garibaldi et al., 2011). In order to support pollinators a system of 'ecological intensification' should be adopted. Ecological intensification is defined as methods by which farmland is actively managed to increase the intensity of ecological processes that support production, whilst minimizing environmental damage (Dicks et al., 2016; IPBES, 2016).

To combat losses of pollinator biodiversity, drivers of change include restoring areas of natural and semi-natural habitat to provide nesting and foraging resources. Greater landscape diversity results in more diverse pollinator communities and can be achieved with wildflower strips in field margins, hedgerows and weed flowers (Carvalheiro et al., 2011; Dicks et al., 2015; M'Gonigle et al., 2015). An increase in floral resources enhances pollen and nectar sources for managed and wild pollinators, leading to higher visitation and seed set (Blaauw & Isaacs, 2014; Carvell et al., 2017; Timberlake et al., 2021). Agri-environment

schemes (AES) advocate the creation of flower-rich habitats to conserve wild pollinators (Carvell et al., 2007). In the UK, AES flower mixes were first designed to support bumblebee communities, which has been a success in farms where the scheme was implemented (Wood, Holland, Hughes, et al., 2015). These mixes have been found to primarily attract certain insects such as bumblebees over other taxa, so there need to be further research into recommendations for a wider pollinator group (Wood et al., 2017; Wood, Holland, & Goulson, 2015). However, bumblebee species are an important focal group of pollinators because they are social, generalist foragers, and have been found to be the dominant visitors to a range of crops (Hutchinson et al., 2021).

1.4 Blueberry pollination

Pollinators contribute to the yield and quality of the highbush blueberry crop, *Vaccinium corymbosum,* an economically valuable crop worldwide. Blueberries are nutritious, containing high levels of antioxidants and vitamins and are widely recognised as a "superfood". In a comparative study with 40 other fresh fruit and vegetables, blueberries ranked number one in antioxidant levels (Prior et al., 1998). The fruit contains anthocyanins and phenolics, both of which have been found to possess antioxidant properties (McGhie et al., 2003; Sellappan et al., 2002).

Their perceived health benefits have led to a steady increase in production over the last decade. Native to eastern North America, there has been a global expansion in blueberry production, increasing by 53% between 2013 and 2018 (FAO, 2020). *V. corymbosum* flowers are small, with poricidal dehiscent anthers that require buzz-pollination, the behaviour of vibrating the anthers to release the pollen, an ability that bumblebees (but not honey bees) possess (Buchmann, 1983; De Luca & Vallejo-Marín, 2013). Although bumblebees are better adapted to pollinate blueberries through their ability to "buzz" and their large bodies that can carry a high number of pollen grains (Javorek et al., 2002), honey bees are often the most abundant visitors to the crop (Benjamin & Winfree, 2014; Button & Elle, 2014). High levels of agricultural intensity have led to a lower abundance of bumblebees that are able to

fulfil pollination services, although their hibernation cycles over the winter months make it unlikely that wild bumblebees could dominate in early-spring flowering crops such as *V. corymbosum* (Hutchinson et al., 2021). This produces a reliance on commercial bees to carry out this service. Commercial colonies contain approximately 150-300 workers and are often implemented in high densities in soft fruit crops used to supplement the pollination services carried out by wild pollinators (Isaacs & Kirk, 2010; Ricketts et al., 2008).

Although *Vaccinium coyrmbosum* is self-fertile to a degree, cross-pollinated flowers increase the fruit yield by producing more seeds, heavier fruit and greater fruit set (Harrison et al., 1993; Nicholson & Ricketts, 2018). In field studies, pollinators were found to have a significant impact and the presence of bees increased the yield of fruit by 50-80% (Button & Elle, 2014). Pollination by honey bees and wild bees have been found improve the size and uniformity of berries, creating a more attractive and economically valuable product (Benjamin & Winfree, 2014; Blaauw & Isaacs, 2014; Isaacs & Kirk, 2010). The dependence on pollinators for fruit yield and quality coupled with high market value make *V. corymbosum* crops an important study system in which to investigate the pollen resource use by visiting insects. By identifying the most attractive floral resources for the most effective pollinators of *V. corymbosum*, the pollination services and therefore economic output for growers can be increased (Nicholson & Ricketts, 2018).

1.5 Foraging resources

Foraging resources are thought to be key influencers in shaping pollinator communities. The quantity, quality, phenology and spatial distributions of resources have been found to affect insect communities and populations. It is not well understood how the different factors interact, but the key findings are summarised below.

1.5.1 Resource quantity

Resource quantity can be described as the number of flowering units available in the landscape. A high abundance of flowering resources can affect pollinators foraging, which has been observed in mass flowering crops (MFCs). MFCs have been found to attract wild pollinators at the landscape scale, which produces a temporal dilution effect on the pollinator community (Holzschuh et al., 2011). The impact of MFCs can be positive, as they act as a huge burst of pollen or nectar resource, increasing wild bee populations (Holzschuh et al., 2013; Riedinger et al., 2014; Westphal et al., 2006). However, MFCs have also been found to draw pollinators away from other resources, such as wildflower species. The presence of oilseed rape (*Brassica napus*) was found to decrease the pollination of cowslip (*Primula veris*), a co-flowering wildflower species that was growing in the field margins (Holzschuh et al., 2011). There is also evidence of co-flowering crops drawing pollinators away from the crop of interest, affecting the yield of the fruit (Grab et al., 2017). It is likely that floral constancy, the mechanism by which bees visit flowers of the same species to increase foraging efficiency, makes it more likely that bees will forage from the most abundant flower taxon close to the colony (Grüter & Ratnieks, 2011; Osborne et al., 1999).

At a smaller scale, plants contain varying numbers of flowers per plant, which can then produce different quantities of nectar and pollen (Hicks et al., 2016). This information could be valuable in predicting the relative rewards of flower taxa and has been recorded for some species, but there is not currently a complete dataset that is suitable for landscape-scale studies (Baude et al., 2016).

1.5.2 Landscape composition

Landscape features can affect populations of wild pollinators in agricultural ecosystems (Kennedy et al., 2013). Natural habitats surrounding farmland provide nesting and food resources for insects, which improve diversity and density of visiting pollinators (Kremen et al., 2002; Williams & Kremen, 2007). As previously mentioned, wild pollinators improve the

quality and yield of fruit, so it is important to manage farms in a way that support insect communities. Increasing areas of natural habitat has been found to increase the pollinator abundance in fruit orchards, with associated benefits to the pollination services for the crops (Carvalheiro et al., 2012). Local-level improvements can be made by increasing the proportion of flower habitats on the farm, which has been found to support a greater diversity and abundance of bumblebees (Carvell et al., 2007; Pywell et al., 2011). This can be achieved by planting wildflower margins and allocating areas of the farm that are wild or semi-natural habitats. Flower strips are often encouraged by agricultural environment schemes (AES) as a method of mitigating against the negative effects of agricultural intensification (Albrecht et al., 2020; Batáry et al., 2015; Potts et al., 2016; Requier & Leonhardt, 2020). The creation of flowering hedgerows was found to increase the population of insects in surrounding fields, demonstrating the potential of within-farm management practices (Morandin & Kremen, 2013).

Bee species differ in their foraging ranges and habitats, and therefore show different responses to landscape configuration (Greenleaf et al., 2007). Solitary bees fly shorter distances compared to bumblebees and honey bees, which is predicted by their smaller body size, and must be considered when designing pollinator-friendly landscapes (Gathmann & Tscharntke, 2002; Greenleaf et al., 2007; Osborne et al., 2008).

1.5.3 Pollen resource quality

Pollen is a crucial component of bee diet and an important source of protein, lipids, sterols, amino acids, carbohydrates, vitamins and minerals (Vaudo et al., 2020; Vaudo, Patch, et al., 2016; Vaudo, Stabler, et al., 2016). The nutritional profile of pollen contributes to the health and development of the colony, since pollen is brought back to the colony as food provisioning for larvae and growing workers (Vaudo et al., 2018). The protein to lipid ratio is thought to be an important factor in discerning whether a pollen source is good quality or not, with diets high in protein producing the most beneficial impacts to the colony, such as higher mass and reproductive output (Baloglu & Gurel, 2015; Moerman et al., 2017;

Moerman, Roger, et al., 2016). When foraging in the landscape, bumblebees have been found to favour protein-rich pollens suggesting they are able to discern the pollen quality in different floral resources (Leonhardt & Blüthgen, 2012; Ruedenauer et al., 2016).

The profile and quantity of amino acids have also been found to impact colony fitness, as bumblebee larvae development was slower when fed a low amino acid diet (Moerman, Vanderplanck, et al., 2016). Like protein, bumblebees have also been found to preferentially gather pollen with higher levels of amino acid content (Kriesell et al., 2017; Leonhardt & Blüthgen, 2012; Somme et al., 2016). Less is known about bumblebee diets and how a mixed diet compares to a monofloral diet in terms of health. In honey bees a polyfloral diet increased immunocompetence and tolerance of parasites (Alaux et al., 2010; di Pasquale et al., 2013). It is thought a diverse pollen diet is less important than the quality of the pollen collected, with Moerman et al. (2017) finding a high quality monofloral pollen diet produces larger larvae than bees fed difloral low quality pollen diet.

1.5.4 Phenological progression

The timing of floral resources, known as phenology, affects the availability of plant taxa in the foraging landscape. There are several studies that have examined the temporal effects of foraging throughout the year in different insect groups by recording the pollen taxa collected (Bertrand et al., 2019; De Vere et al., 2017; Lowe et al., 2022). Woody taxa such as *Salix* and *Prunus* have been found to be important early season foraging resources for bumblebees (Bertrand et al., 2019; Kämper et al., 2016), whilst early-flowering tree species such as *Acer* and *Quercus* are more prevalent in solitary bee diets (Allen & Davies, 2023; Bertrand et al., 2019; Persson et al., 2018). Later in the year shifts are observed from woody plants to herbaceous flowers, and taxa such as *Crataegus, Ranunculus, Rubus* and *Trifolium* become more common in the diet (Bertrand et al., 2019; Kleijn & Raemakers, 2008; Lowe et al., 2022).

Landscape availability of resources can be used to study pollinator preferences, whereby the pollen taxa is compared to the floral taxa in the landscape. This data can also be used in assessing whether there is a “phenological gap” in resources at certain times of the year, which indicates the months there should be improved efforts to provide floral resources (Timberlake et al., 2019). Such gaps have been identified in farmland in southwest England in early spring (until late March), June and August, which could be filled by mass-flowering crops (Timberlake et al., 2019).

1.6 Monitoring floral visitation

The flowers that pollinators visit for resources can be recorded and used to create plant-pollinator networks. Networks allow us to understand which resources are most frequently used by insect taxa and to estimate levels of specialisation and resilience to environmental change (Lonsdorf et al., 2009; Memmott, 1999).

One method of investigating interactions is through visitation surveys, where any insect visiting a flower is recorded, for example, on transect walks, timed searches, point counts or in experimental conditions. The advantages to visitation surveys are being able to record multiple insect taxa visiting a habitat and being able to discriminate between pollen visits and nectar visits. However, surveys may miss important interactions as they are biased towards easily accessible habitats such as flower margins and may miss difficult to survey taxa such as tree canopies, or smaller insect and plant taxa groups (Allen & Davies, 2023)

The movement of bees in the landscape can also be recorded using coloured paint to mark them or small radio trackers. Harmonic radar tracking, where small radar emitting devices are attached to bees, has been used to record flight paths in the landscape (Osborne et al., 1999, 2013; Riley et al., 1993). Tracking approaches provide information about how social bees explore and exploit heterogenous landscapes around their colonies and provide data on the distance and direction bees take. However, the devices have a cost on foraging activity even with weights as low as 200 mg, with lower distances and longer rest times recorded for

bees with trackers attached (Hagen et al., 2011). Tracking technology using passive radio frequency identification is being developed that have longer ranges and lighter weight, which could provide a low-cost alternative to current options (Barlow et al., 2019).

Palynology, the study of pollen, can be used to provide information about plant-pollinator interactions. Pollen can be washed from the body of the insect, from where there has been contact between the pollen producing anther and the insect's head or abdomen (Wood et al., 2016). Social insects such as honey bees and bumblebees collect pollen from flowers by gathering it in corbicular loads, structures on their hind legs that act as pollen baskets, to transport it back to their colony. Pollen can be removed from the bee by capturing it whilst foraging in the landscape or at the entrance to their colony on their return. Once removed the pollen can be stained and identified morphologically using a microscope (Rahl, 2008). This process requires a high level of training and expertise and if dealing with large sample numbers, it is a time-consuming process (Khansari et al., 2012). Identifying pollen to species level has an estimated <78% accuracy and is mostly only possible to genus or family level (Mander et al., 2014). A lack of taxonomic resolution is limiting in studies where the plant species of interest cannot be distinguished from hedgerow flowers. Foulis and Goulson (2014) analysed bee pollen loads to identify the floral resources used by commercial bees on raspberry crops. Raspberry and blackberry (genus *Rubus*) frequently flower at the same time and are morphologically indistinguishable, so it was difficult for the researchers to estimate the importance of hedgerow pollen in comparison to crop pollen.

1.7 DNA Metabarcoding

Over the past decade molecular techniques have become cheaper and more accessible, making them attractive alternatives to field studies and identification using microscopy. Next Generation Sequencing (NGS) can be used as a pollen identification method to process mixed samples quickly (Valentini et al., 2010). DNA barcoding and metabarcoding currently dominate the field of pollen identification and quantification (Bell et al., 2022). DNA barcoding uses a specific DNA region, or barcode, that is amplified via polymerase chain

reaction (PCR) using universal primers, sequenced using high-throughput sequencing technologies and matched with known sequences in a reference database (Hebert et al., 2003). Barcoding is used for single species samples, whereas metabarcoding allows for the simultaneous identification and quantification of multiple species in a mixed sample (Ji et al., 2013).

The barcode, a short nucleotide sequence from a standard genetic locus (300-800 base pairs), is selected for small intraspecific and large interspecific differences within a group of organisms. For example, the mitochondrial cytochrome oxidase subunit I (COI) gene is most commonly used in animals, while different barcodes are used for plants (e.g. *rbcL,* ITS2), fungi (ITS) and microorganisms (16s rRNA) (Hollingsworth, 2011). Projects such as the Barcode of Life aim to create a collection of references sequences for all species of life (Adamowicz, 2015). Barcode sequences are uploaded to the Barcode of Life Database (BOLD), a library that is freely available to the public and a widely used tool.

DNA metabarcoding has been applied to pollinator research to allow large-scale identification of pollen taxa. Pollen from honey (Galimberti et al., 2014), airborne samples (Kraaijeveld et al., 2015) and bee-collected pollen loads (Cornman et al., 2015; Hawkins, De Vere, et al., 2015; Keller et al., 2015; Richardson, Lin, Sponsler, et al., 2015) have been used to develop metabarcoding approaches in this field. There are several different barcode markers that are used in pollen metabarcoding studies; including chloroplast derived markers such as *rbcL* and *matK*, and nuclear markers such as internal transcribed spacer 2 (ITS2).

Pollen metabarcoding has been found to provide a higher level of species richness in mixed pollen samples than traditional techniques (Arstingstall et al., 2021; Pornon et al., 2017). Rare plant taxa can be detected using metabarcoding, with pollen abundance as low as five pollen grains (Pornon et al., 2017). Problems can arise with closely related plant species due to similarities in barcode sequences, which produce false identifications (Hollingsworth et al., 2016). Usually only genus- or family-level taxonomic identification can be achieved, because

there is not a high enough level of variation between barcode sequences of closely related taxa to capture species-level information (Bell et al., 2018).

In addition to measuring the species richness and diversity of pollen samples, metabarcoding can be used to produce quantitative results, which provide a more informative analysis than presence / absence studies. There have been positive correlations observed between the proportion of pollen grains and amplicon sequences in studies of honey bee-collected pollen (Richardson et al., 2019, 2021) and airborne pollen samples (Kraaijeveld et al., 2015).

Whilst the factors affecting qualitative success are well understood, the same cannot be said for quantitative DNA metabarcoding. When mixed samples of known biomass are analysed using metabarcoding pipelines, the resulting sequence data is compared to the original sample biomass to examine whether there is a correlation. There has been some disagreement over the accuracy of metabarcoding's ability to predict pollen proportions. Some studies have suggested metabarcoding can provide a semi-quantitative result (Pornon et al. 2016), but several have found the results to be unreliable (Richardson et al. 2015; Sickel et al. 2015; Bell et al. 2018; Kamo et al. 2018). Certain taxa exhibit a strong relationship between marker proportion and pollen proportion, but this can vary within and between (Baksay et al., 2020). Contamination of samples can also lead to taxa being given an unrealistically high weighting in plant-pollinator networks (Pornon et al., 2017) and sequence results do not always accurately quantify the frequency of flowers visited, as the number of sequence reads do not represent the number of pollen grains (Bell et al., 2017).

There are several factors thought to be involved in the mixed quantitative abilities of pollen metabarcoding. One factor is variation in pollen morphology and characteristics. Pollen grain size, exine (pollen wall) structure and genome size, have been shown to have an effect on the DNA extraction and amplification processes in different plant taxa (Pornon et al., 2016; Swenson & Gemeinholzer, 2021). These differences can lead to species-specific biases in quantitative results.

There are also biases associated with the polymerase chain reaction (PCR) step in metabarcoding. Sequence divergence in priming sites between plant taxa have been found to affect priming efficiency and the resulting sequence abundance. Nucleotide mismatches occur between the primer and the sample's target site that differ between samples and can account for 75% of variation observed when testing the relationship between primer mismatch and amplification efficiency (Piñol et al. 2015). Other causes of bias are regions of high or low GC contents that amplify to a lesser extent than highly variable regions, and sequence polymorphisms in the primer site that mean some plant species are amplified less efficiently than others (Kraaijeveld et al. 2015).

It is not only pollen studies that have had mixed results over the effectiveness of metabarcoding as a means of species quantification. In a meta-analysis of metabarcoding studies there was a weak quantitative relationship between sample biomass and the number of sequences in a range of organisms (Lamb et al., 2019). It is thought that by increasing the number of samples the variance could be reduced, potentially improving quantitative results.

Further recommendations to improve metabarcoding's quantitative ability include the use of multiple barcodes and combining the sequence results, which has been found to be more effective than using a single locus (Richardson et al., 2019). Using correction factors could also be used to improve taxon-specific biases, which would require laboratory tests of mock mixes of known mass to calibrate sequencing results (Garrido Sanz et al., 2022). Using pollen weight as an indicator is one such method, which has been used in honey bee samples to sum the pollen weight of each taxa (grouped by colour) and calibrate the DNA sequence proportions accordingly (Kamo et al. 2018).

Comprehensive barcode databases are essential in answering ecological questions using metabarcoding. There is a near-complete reference database for native and non-native flowering species in the UK, targeting three commonly used barcodes (Jones et al., 2021) but this is not true for other countries. It is estimated that only 25% of species worldwide have barcodes that are publicly available (Bell et al., 2021). Databases are being continuously

added to, but using a range of different barcodes that might differ across countries and regions (Kress 2017).

1.8 PCR-free approaches

Creating reference genomes for eukaryotes is difficult and costly due to their complexity and length. High-throughput sequencing technologies have made this process much simpler. Genomic data can be taken from any plant species and used in phylogenetic, population genetic and barcoding studies. Genome skimming, also known as whole-genome shotgun sequencing, is a way of "navigating the tip of the genomic iceberg" (Straub et al*.,* 2012). Straub's study was the first to demonstrate this approach to create a low cost, low coverage shotgun sequencing method to obtain deep sequencing of the high-copy fraction of the genome (Dodsworth 2015). Coverage refers to the proportion of fragments sequenced. At a high coverage it is easier to piece the genome back together, but it is also a more expensive option (Berger et al*.,* 2017). At a low sequence coverage, approximately 0.1-10x, a fragmented nuclear assembly can be generated. Rather than specific amplified regions, such as the genes targeted in metabarcoding, random collections of regions where sequences overlap are stitched back together and assembled into a library (Srivathsan et al*.,* 2015; Hollingsworth et al., 2016).

As a suggested alternative to PCR-based methods, genome skimming is still relatively underused in ecology and conservation biology. Unlike techniques that use PCR, the copy number of the genetic material does not change during sample processing, so in theory, results can be studied quantitatively without PCR-biasing problems (Paula et al., 2015) Genome skimming has enabled the quantification of mixed samples in freshwater macroinvertebrates (Bista et al., 2018), terrestrial arthropods (Zhou et al. 2013), bulk bee samples (Tang et al., 2015) and pollen (Lang et al., 2019). Lang et al. (2019) sequenced pollen genome skims on an Illumina platform to identify mock mixtures, correctly predicting the ratios of pollen taxa with a high level of repeatability and accuracy.

1.9 Nanopore technology

Oxford Nanopore Technologies' (ONT, Oxford, UK) MinION device is a portable, handheld device that can sequence RNA or DNA data almost immediately. It is low-cost (especially if reagents are bulk ordered) in comparison to other technologies that need a lab set up, making it an accessible option for research groups globally (Oxford Nanopore Technologies, 2019). Due to its accessibility and portability it has been used in several remote locations, including the International Space station, the Antarctic Dry Valleys and a montane rainforest in Tanzania (Castro-Wallace et al*.,* 2016; Johnson et al*.,* 2017; Menegon et al. 2017). A new feature of the MinION is the "read until" function, which allows user to pre-select a sequence of interest and only analyse those strands, which could have an important role in emerging pathogen detection in plants as well as human health (Quick et al., 2016).

The main difference between nanopore and other sequencing technologies, such as Illumina and Ion Torrent, is the method by which the DNA is sequenced. In nanopore sequencing, DNA or RNA strands are passed through protein nanopores embedded in a membrane. When a voltage is applied to the membrane, ions flow through the pore. Sensors detect the change in the voltage when a polynucleotide passes through, and as each nucleotide base produces a different electrical signal (a "squiggle") they are recorded and computationally translated to produce basecalled sequences (Rang et al., 2018).

When the MinION was first launched, the only basecalling option was to upload raw signal files to a cloud-based server which were then downloaded back onto the same laptop. This proved to be a problem when using MinION in the EBOLA epidemic when internet connection was slow and unreliable, or even non-existent (Quick *et al.* 2016). With growing use of local basecallers, connection to the internet is no longer a requirement. Local basecalling can be done on the same laptop without uploading to a server and has opened up third-party programmes that remove the need for the internet, which has advantages when working in remote field locations (Leggett & Clark 2017). Currently, there are a range of ONT basecaller options that can improve the read quality of the sequences produced; Fast, High Accuracy, and Super Accurate. The Fast basecaller is best suited to keep up with the

"read-until" function, whilst the Super Accurate version is for the highest quality reads, but at the expense of read length and yield (Ferguson et al., 2022).

An additional advantage of using the nanopore platform over other sequencing technologies is the use of long read sequences. Read lengths tens of kb long are possible, with one study looking at bacterial genome sequence assemblies achieving a 98kb sequence (Laver *et al.* 2015). This was, however, not achieved without problems; high error rates and problems sequencing GC-rich regions produce lower quality reads (Wang et al., 2021). Choice of DNA extraction method can have an effect on the length of DNA fragments that can be sequenced. Using the phenol-chloroform extraction method rather than kits such as Macherey Nagel produces a higher yield of DNA and with minimal pipetting can generate much longer reads (Jain *et al.* 2018).

Since its arrival in 2014, ONT and its software have been receiving continuous updates to the device, flow cell design, sequencing kits, and software. More recent studies have demonstrated a lower error rate and higher yields (Jain *et al.* 2017), and accuracy and yield are only expected to improve over the coming years. It is currently considered that accuracy is ~95% and yields of 3-5 Gb are standard (Delahaye & Nicolas, 2021; Leggett & Clark, 2017). Although the ONT platform still lags behind other sequencing platforms in terms of read quality, this might not present a significant disadvantage when the long read lengths, low cost and portability are taken into account. Leidenfrost et al. (2020) found comparable compositional results when sequencing the ITS2 region in pollen loads on the nanopore and Illumina platforms.

1.10 RevMet

A recent study in pollen characterisation used whole genome sequencing without the use of reference genomes, producing semi-quantitative results (Peel et al., 2019). Whole genome skims targeting a 1x coverage were sequenced using the Illumina platform. These short reads are mapped to long read query sequences generated by ONT's MinION to identify mixed

pollen samples, which is the opposite to typical metagenomic studies, hence the name 'Reverse Metagenomics' (Figure 1.1). The method demonstrated accuracy in qualitative and semi-quantitative analysis of the species compositions, with low false negative and false positive rates. One potential improvement would be to generate plant genome reference skims with a higher coverage. Peel et al. (2019) used skims with an average coverage of ~0.3x, which may have contributed to the misidentification of closely related species. To improve accuracy, higher coverage skims are recommended to give higher sequencing depth, and increase the number of positive matches. RevMet has the potential to be expanded to other organisms to answer ecological questions, especially with regular updates from ONT making DNA analysis more cost-effective and portable.

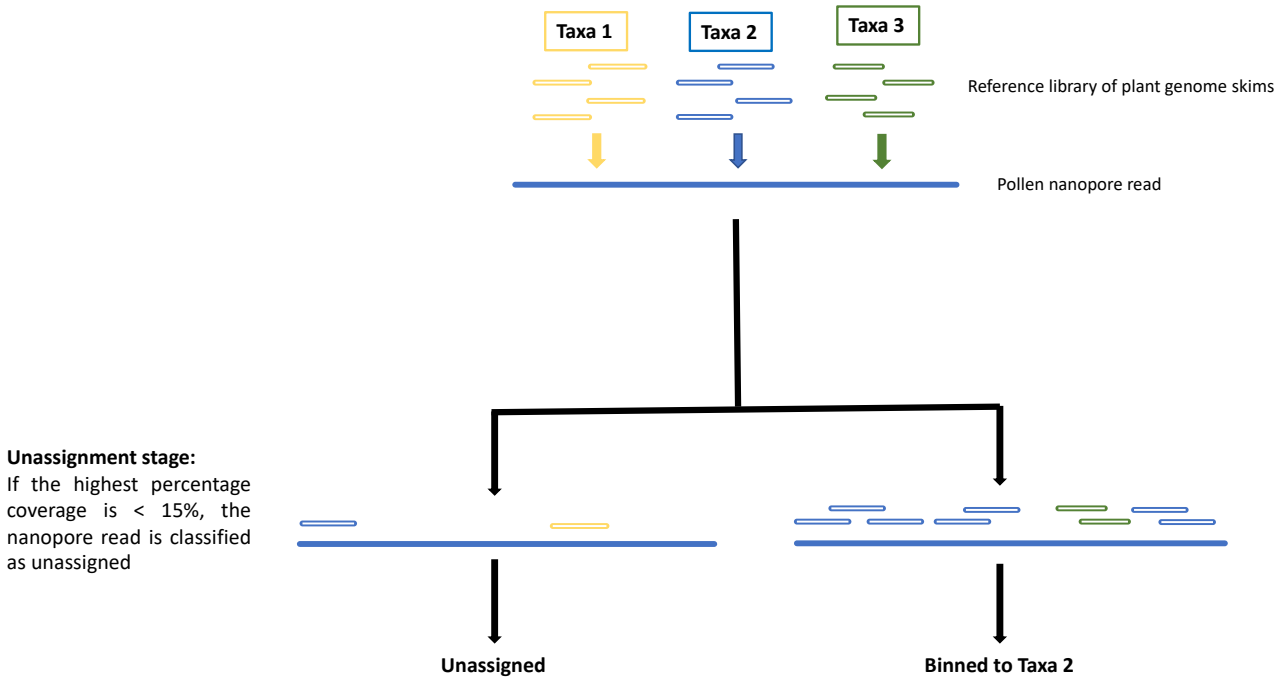


Figure 1.1 The classification process of reads in the RevMet pipeline. When the plant reference skims are aligned to the pollen nanopore long reads a percentage coverage is calculated, which is the proportion of the long read that is aligned to skims belonging to a plant taxon. The plant reference skim with the highest coverage will be assigned to that nanopore read, unless the highest coverage is < 15%, in which case the nanopore read will be classified as unassigned.

The current set- back in using genome-free methods such as RevMet lies in the lack of publicly available assembled reference genomes. Projects such as the Darwin Tree of Life (DToL) are creating reference genomes for all known eukaryotes, which will be continuously updated over the coming years (Lewin et al., 2018). Until then, genome skims can provide a low-cost alternative, with a lane of Novaseq 6000 costing approximately £50 per sample for a 1x coverage of a 3Gb genome.

1.11 Knowledge gaps

The study of bee pollen diet has been well-researched, however there is no study (to our knowledge) of bee foraging on *Vaccinium corymsbosum* farms in the UK. In this thesis RevMet, a recently developed molecular approach, is tested and explored (Chapters Two and Three) and applied to a landscape-scale ecological study (Chapter Four).

1.11.1 Testing the RevMet approach

There are few studies that have used genome skimming or whole genome sequencing approaches to characterise and quantify mixed species pollen loads (Leidenfrost et al., 2020; Bell et al., 2021). The RevMet pipeline offers potential as a promising alternative to metabarcoding in characterising mixed pollen samples (Peel et al., 2019). Its quantitative abilities have been tested using mock mixes of extracted DNA (Peel et al., 2019), but have not yet been compared to microscopy or mock mixes of pollen grains. The RevMet approach should be further tested to evaluate its suitability for studies on bee pollen diets.

1.11.2 Bee foraging in blueberry crops

There is growing evidence of the unsuitability of agriculture land in supporting pollinators, primarily due to agricultural intensification, chemical output and lack of available resources (Carvell et al., 2017; Samuelson et al., 2018). This highlights the need for further research into pollinator foraging and the identification of frequently used plant taxa in farmland where there could be resource gaps (Timberlake et al., 2019). The study of bumblebee pollen diets has been well documented in many ecological and agricultural landscapes, but none have studied highbush blueberry (*V. corymbosum*) as a focal crop in the UK. *V. corymbosum* is a high value crop and has become increasingly popular to growers in the UK over the last decade. The commercial buff-tailed bumblebees (*Bombus terrestris*) used to pollinate the crop are active earlier in the year than their wild conspecifics, which provides an insight into the foraging of wild queens, the first individuals to emerge in early spring. The phenological progression of flowers in spring provides a varied and changing pollen diet for foraging bees, which we might expect to see in the pollen they return to the colony. The mechanisms by which foraging bees choose to visit certain plant taxa over others are not fully understood, because it is a complex and multi-factored process. By documenting the flowering plants available to bees the most frequently visited plants in the landscape can be identified, and foraging in the context of a mass-flowering crop can be better understood. In identifying the attractive plant taxa in the agricultural landscape the planting of certain taxa can be recommended to sustain early emerging pollinators.

1.12 Study regions and landscapes

The studies were conducted at two farms in southern England (Heathlands and Tuesley) in March to May 2019, which were extended to an additional two farms (Colworth and Winterwood) in March to June 2021 (Figure 1.2). There was no fieldwork conducted in 2020 due to the Covid-19 pandemic, which caused a nationwide lockdown from March 2020 to the summer of that year. The experiments were established in fields growing *Vaccinium corymbosum* crops, which are situated within Spanish polytunnels. Spanish polytunnels are

field-scale structures that have a metal frame covered by plastic sheets and are used in early spring due to low night-time temperatures and occasional frost, which can damage the fruit. Fields were surrounded by a matrix of semi-natural habitat, woodland, arable land, pasture and residential settlements, including gardens.

For the 2019 data collection (Chapter Two), plant samples were collected from a 500m radius around the farm, which were used to create the genome skim reference database for the RevMet analysis. In 2021 the radius was expanded to 1km to include the approximate foraging range of the bumblebees (Westphal et al., 2006), and collect additional samples for the genome skim reference database while conducting floral transect surveys (Chapter Four). The combined plant lists from 2019 and 2021, in addition to the skims created by Peel et al. (2019), were used for Chapters Two, Three and Four in the RevMet analyses.

Landscape maps were created for Chapter Four in order to compare bumblebee pollen load composition to resource availability in the landscape. Land cover types were classified in a 1km radius around each of the farms using the open source QGIS software and satellite images (GoogleEarth), which were later ground-truthed.

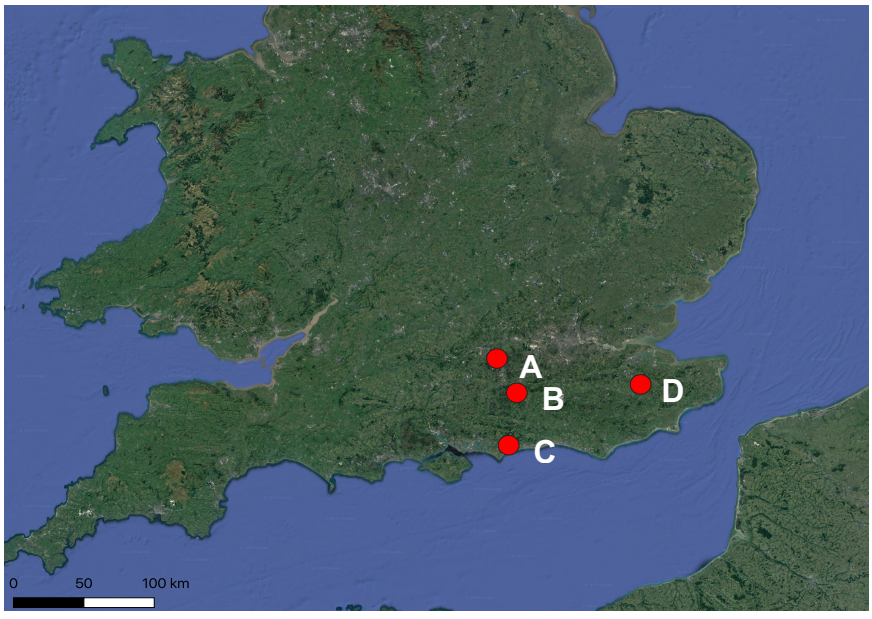


Figure 1.2 Map to show the location of the four farms used in this study: A) Heathlands, B) Tuesley, C) Colworth and D) Winterwood. QGIS v3.28.5

1.13 Thesis outline and aims

This introduction has highlighted the importance of pollinator research and the emerging role of DNA sequencing, in ecological research. In this thesis RevMet, a recently developed molecular approach, is tested and explored (Chapters Two and Three), and applied to a landscape-scale ecological study (Chapter Four). The specific aims of each chapter are outlined here.

In Chapter Two, numbers of RevMet reads are compared to counts of pollen grains using a light microscope from the same bee-collected samples, to investigate the quantitative relationship between the two different measures. Nanopore sequencing is an expanding method of sequencing that has been applied to few pollen studies (Leidenfrost et al., 2020; Peel et al., 2019).

Chapter Three uses the nanopore platform to sequence pollen mock mixtures in order to compare the qualitative and quantitative abilities of two molecular approaches. Using bee-collected pollen, six plant taxa were selected representing a range of genome sizes and used to create mock mixtures that varied in composition. The samples were analysed using the ITS2 marker and the long-read RevMet approach, and compared to the mock mixture pollen grain relative abundances estimated quantified using microscopy.

Chapter Four investigates the foraging of commercial bumblebees (*B. terrestris)* in four farms growing *V. corymbosum* over the crop flowering period. Using a combination of landscape floral surveys and pollen diet analysis using the RevMet approach, foraging preferences can be measured. The effect of crop flower cover and landscape floral richness are also used to investigate the factors impacting *V. corymbosum* pollen foraging.

Finally, in Chapter Five, the results from Chapters Two, Three and Four are synthesised and future directions for this research are discussed. The ability of RevMet in characterising and quantifying mixed species pollen loads is explored, including a wider assessment of the

nanopore platform and whole genome sequencing. Lastly, the results from this thesis are looked at in the context of pollinator foraging in agricultural landscapes, and key considerations for the farmland and floral resource management are outlined.

Author contributions

The three data chapters have been written as self-contained accounts of the research and are in preparation for publishing. This means there is a degree of overlap and repetition between the chapters, although certain methods have referred to the appropriate chapter for further detail. The work here is predominantly my own, but I have credited guidance and input from my supervisory team and others below.

Chapter Two (in prep)

The experiments were conceived and designed by Eleanor S. Kent and Lynn V. Dicks. E.S.K. carried out the data collection and conducted the lab work under the supervision of Darren Heavens. E.S.K conducted the bioinformatic analysis with input from Ned Peel. E.S.K conducted the data analysis and wrote the manuscript. Richard M. Leggett, Doug W Yu, and L.V.D. provided guidance on analyses and all authors contributed to manuscript editing.

Chapter Three (in prep)

The experiments were conceived and designed by Eleanor S. Kent and Lynn V. Dicks. E.S.K. carried out the data collection and conducted the lab work under the supervision of Darren Heavens. E.S.K conducted the bioinformatic analysis with advice from Ned Peel. E.S.K conducted the data analysis and wrote the manuscript. Richard M. Leggett, Doug W Yu, and L.V.D. provided guidance on analyses and R.M.L. and L.V.D. contributed to manuscript editing.

Chapter Four (in prep)

The experiments were conceived and designed by Eleanor S. Kent and Lynn V. Dicks. E.S.K. carried out the data collection and conducted the lab work under the supervision of Darren Heavens. E.S.K. conducted the bioinformatic analysis, the data analysis and wrote the manuscript. Katherine A. Berthon, Richard M. Leggett, Doug W. Yu, and L.V.D. provided advice and guidance on analyses, and K.A.B. and L.V.D. contributed to manuscript editing.

Chapter Two

A comparison between microscopy and RevMet for identification and quantification of plant taxa in bee-collected pollen

2.1 Abstract

Determining the plants most frequently visited by pollinators is important in understanding their ecology. Molecular methods such as metabarcoding have been widely adopted to characterise plant taxa in bee-collected pollen loads. However, there is evidence of quantitative bias when using metabarcoding, which has encouraged the development of PCR-free approaches that aim to improve accuracy in predicting pollen relative abundances. Here, we test the ability of the recently developed 'Reverse Metagenomics' (RevMet) technique, a PCR-free shotgun metagenomics approach, to characterise and quantify bee-collected pollen. We collected pollen loads from 96 bees on highbush blueberry (*Vaccinium corymbosum)* farms at the beginning and peak periods of crop flowering and compared the results generated by RevMet to those obtained using light microscopy. Our results show a positive quantitative correlation between the two methods, with RevMet identifying significantly more plant taxa than microscopy. At the beginning of crop flowering, the majority of pollen originated from non-crop plants such as *Salix caprea*, and during the peak blueberry flowering period the bees collect pollen largely from the crop itself but supplemented with a higher number of flowering taxa. A significant number of unassigned reads were identified with RevMet, which were explored and found to be an artifact of the reverse metagenomic approach with no significant effect on the method's ability to identify and quantify plant taxa in the pollen loads. This study provides evidence of commercial bumblebees visiting pollinator dependent crop flowers and the potential for RevMet to be used as an accurate approach in quantifying pollen taxa.

2.1 Introduction

Insect pollinators play a crucial part in maintaining ecosystem health, crop production and floral biodiversity (Potts et al., 2016). It is estimated that 75% of major food crops, including many fruits, seeds and nuts, rely to some degree on animal pollination (Klein et al., 2007). Concerningly, the diversity and abundance of insect pollinators has decreased, a result of agricultural intensification, loss of natural habitat and increased usage of pesticides (IPBES, 2016). To supplement wild pollinator services, commercial bees such as the western honey bee (*Apis mellifera)* and buff-tailed bumblebee (*Bombus terrestris*) are increasingly used to provide pollination services to crops (Potts et al., 2016).

Commercial bumblebee colonies are used to supplement pollination in a variety of soft fruit crops in the UK, including blueberry, raspberry and strawberry (Velthius & Doorn, 2006). An estimated 15,000 colonies are used every year for soft fruit pollination in open-ended polytunnels and open fields, allowing the bees to forage in the surrounding landscape and not only on the crop itself (Goulson, 2010). Bees collect pollen as their primary source of protein and lipids, essential for the healthy development and growth of the colony. Pollen sources differ in their nutritional content and these differences have been found to affect bees' reproduction, susceptibility to disease and larval development (di Pasquale et al., 2013; Leidenfrost et al., 2020; Roger et al., 2017). Social bee species are able to balance their colony's nutritional intake by gathering pollen from a diverse range of flowering plants, and by selecting high-quality pollen sources (Vaudo et al., 2018; Vaudo, Patch, et al., 2016). Understanding the foraging patterns of commercial bumblebee colonies is important to quantify the extent to which they can gain sufficient nutrition from the desired crop and can also be used to infer the pollen sources that wild colonies might use in the same landscape.

To identify the foraging patterns of pollinators, observational approaches focus on quantifying plant visitation frequencies (Baldock et al., 2019; Campbell et al., 2017). However, this approach can miss important interactions as they are limited both spatially and temporally (Fijen & Kleijn, 2017), so the analysis of bee-collected pollen from colony or

nesting sites may help to capture infrequent or difficult to observe interactions (Bell et al., 2017; Pornon et al., 2016).

Traditionally, pollen collected by bees has been identified morphologically using light microscopy, a method that is time-consuming, costly, and requires a high level of expertise (Rahl, 2008). Identification of pollen to species level is sometimes not possible where members of the same family or genus have identical morphologies (Khansari et al., 2012; Kremen et al., 2007; Mander et al., 2014). Moreover, microscopic identification can fail to record rare interactions if a pollen morphotype is present in low quantities, because only a small subsample is analysed.

For these reasons, pollen studies have adopted high-throughput sequencing (HTS) approaches as an alternative to microscopic analysis (Bell et al., 2022). DNA metabarcoding is a method that amplifies and sequences short regions of the genome with low intraspecific variation and high interspecific variation. These regions, or "barcodes", are matched against a reference database of plant barcodes to determine the species composition of the pollen load (Ji et al., 2013; Yu et al., 2012). For animals, the CO1 mitochondrial gene is most commonly used, but for plants, there is no universal barcode (Hebert et al., 2003; Hollingsworth et al., 2011). Instead, using combinations of multiple barcodes has proved to be most effective at characterising and quantifying pollen loads (Hollingsworth et al., 2011; Richardson et al., 2019).

In qualitative pollen metabarcoding studies a list of plant taxa present in mixed-species samples is produced, which has been applied to reveal plant-pollinator interaction networks (Bell et al., 2017; Pornon et al., 2017a), honey floral composition (De Vere et al., 2017; Hawkins, de Vere, et al., 2015), and pollen diversity in air-collected samples (Brennan et al., 2019; Leontidou et al., 2018). However, there are several steps in the metabarcoding pipeline that can produce erroneous results. False positives are possible from contaminant sources which, once amplified in the PCR step, could form a significant proportion of the species composition (Pornon et al., 2016). False negatives can also occur when species are not

included in the reference database, and therefore cannot be identified, which necessitates the use of large and regularly updated barcode reference databases (Bell et al., 2019).

Quantification in metabarcoding examines the relationship between the proportion of high throughput sequencing reads and the proportion of pollen grains present in the sample, i.e., the relative abundance. Studies that applied metabarcoding to the quantification of pollen concluded varying degrees of correlation. Significant positive correlations between abundance estimates from pollen grain counts (the number of grains identified per plant taxa using light microscopy) and amplicon counts (the number of reads assigned to a plant taxa) have been reported by some (Keller et al., 2015; Kraaijeveld et al., 2015; Richardson et al., 2019), but others have reported weak or inconsistent relationships (Bell et al., 2019; Richardson, Lin, Sponsler, et al., 2015; Smart et al., 2017). Variations in the metabarcoding pipeline – starting material, barcodes, primers, PCR cycles – affect the ability to accurately quantify relative abundance in mixed species pollen samples (Baksay et al., 2020a; Lamb et al., 2018; Sickel et al., 2015). The PCR step can introduce bias as primer-template mismatches occur, giving some taxonomic lineages higher weightings than others when compared to microscopy results (Bell et al., 2016; Krehenwinkel et al., 2017; Richardson, Lin, Sponsler, et al., 2015). Quantitative biases do not seem to be consistent, rather they are dependent on the other species present within the mixture (Bell et al., 2018; Pornon et al., 2016). The DNA extraction step includes rupturing the pollen exine, which varies in resiliency between plant species, introducing another source of bias (Pornon et al., 2016a; Swenson & Gemeinholzer, 2021). A higher level of accuracy may be achieved by creating mock mixes of extracted pollen DNA because species-specific biases resulting from the DNA extraction step are removed, but this is not representative of what happens in ecological studies, where pollen grains must be used in the first step.

More recently, PCR-free approaches such as whole-genome shotgun (WGS) sequencing have been adopted, where bulk samples of DNA are randomly fragmented and sequenced to produce a low-coverage set of short reads (Lang et al., 2019; Peel et al., 2019; Straub et al., 2012; Tang et al., 2015; Zhou et al., 2013). For example, a genome-skimming pipeline

detected all plant species in mock mixes created from bee-collected pollen and flower pollen, and pollen frequencies estimated from sequencing results were highly correlated to microscopic pollen grain counts (Lang et al., 2019). In comparison to the quantitative ability of metabarcoding, WGS sequencing was found to have a stronger correlation between pollen proportion and sequence proportion in mock pollen communities (Bell et al., 2021). One limitation of WGS sequencing is the lack of publicly available reference genomes, against which pollen sequences are aligned, although there are efforts to create a global database (Lewin et al., 2018).

The 'Reverse Metagenomics' (RevMet) pipeline incorporates low-cost, low-coverage plant genome skims in place of reference genomes (Peel et al., 2019). Sequencing and assembling eukaryote reference genomes is a costly process (Gilbert & Dupont, 2011), making genome skims an attractive alternative. Peel et al. (2019) generated long-read pollen sequences using Oxford Nanopore Technology's (ONT) MinION device, to which the plant genome skims were mapped. In addition to being used to characterise bee-collected pollen, the quantitative ability of the RevMet pipeline was tested using mock mixes created from extracted pollen DNA. The results provided evidence that RevMet could distinguish between high and low levels of DNA, therefore producing semi-quantitative results. In using extracted DNA to create mock mixes, the inherent challenges in pollen DNA extraction are diminished, leaving unanswered questions about its true suitability for quantifying the relative abundance of pollen samples in ecological studies.

In this chapter, we characterise the pollen loads collected by commercial *Bombus terrestris* workers from colonies located in highbush blueberry (Ericales: Ericaceae: *Vaccinium corymbosum*) crops. The aim is to identify and quantify the plant species from which the bumblebees collected pollen, both within the farm and from the surrounding landscape. Here, we use the RevMet pipeline, which has not yet been compared to results obtained using microscopy, in estimating the relative abundance of plant species.

Specifically, we use the RevMet methodology and microscopy to determine: 1) whether RevMet or light microscopy reveals a higher number of taxa, 2) whether there is correlation between the relative abundances of taxa derived from the two methods, 3) the variation in pollen taxa collected at the beginning of the *Vaccinium corymbosum* flowering season in comparison to peak flowering and 4) whether unassigned reads can be re-assigned or alternatively they are not represented in the database.

2.2 Materials and Methods

2.2.1 Pollen and leaf tissue sampling

Sampling took place across two farms growing *Vaccinium corymbosum* crops in southern England between 26th March and 6th May 2019. Heathlands farm (51°23 34"N 0°49' 06"W), and Tuesley farm (51° 10' 02"N, 0° 37' 16"W) are conventional, medium-sized (80-120 ha) farms growing *Vaccinium corymbosum* and strawberry (*Fragaria x ananassa*) at the time this study took place. Both farms are surrounded by a landscape mosaic of managed woodland, semi-natural habitat, grass pasture, arable land, and urban areas. The semi-natural habitat can be further classified into hedgerows, field margins, road verges and small areas of unmanaged woodland within the farm.

In 2019 leaf samples were collected from 38 species of flowering plants that were encountered during the survey period, within a 500 m radius of the colony locations. In March - May 2021 a further 37 plant species were sampled in a 1 km radius of the farms and used to create a wider reference database. Bumblebees commonly make foraging trips within 500 m of their colony but have been found to forage over 1 km, so the radius was extended to include the possibility of longer foraging trips (Osborne et al., 1999). A section of leaf tissue approximately 2 cm^2 was cut from a young leaf and wiped with Kimwipe tissue to remove any foreign DNA from the leaf surface. Leaf tissue was stored in a labelled bag and

placed in a -20°C freezer on site. The leaf samples were transported to the Earlham Institute on ice where they were stored at -80°C.

Pollen was collected from 96 bees originating from four commercial *Bombus terrestris* colonies (Tripol, Koppert Biological Systems, the Netherlands) situated within the highbush blueberry crop (*Vaccinium corymbosum*), Duke cultivar. The crops were grown within polytunnels that contained between four and six Tripol colonies per tunnel. Pollen sampling took place on dry days, between 9 am and 4 pm, when the temperature was above 13°C and the polytunnel entrances were open. On each farm, two fields were selected (<500 m apart) and a Tripol colony within that field was picked at random. Each 'Tripol' hive contained three colonies, one of which was closed for a maximum of 1 hour so as not to significantly impact the colony's pollen supply. Returning foragers carrying corbicular loads were captured using a queen marking tube and both pollen loads were removed using a mounted needle, after which the bee was released. Both pollen loads from individual bees were stored in 2 ml Eppendorf tubes at -20°C in an onsite freezer, followed by -80°C at the Earlham Institute. When 12 foraging bees had been captured and released, or the 1-hour time period was up, the colony was reopened. Colonies were visited at the two farms in two time periods, between 26th March - 11th April, and 24th April – 6th May in order to compare pollen foraging during early and peak periods of *Vaccinium corymbosum* flowering.

2.2.2 Microscopic identification and quantification of pollen

1 ml of 99% ethanol was added to the 2 ml tube containing the pollen loads from both legs and vortexed to mix the loads and suspend the pollen grains. 100 µl of the ethanol/pollen mix was pipetted into a new 2 ml microcentrifuge tube for microscope analysis.

For each pollen sample, 75 µl of glycerine and 10 µl of Calberla's stain was added to the tube containing the pollen (Smart et al., 2017). After pipetting up and down to suspend the

solution, 20 µl was pipetted onto a microscope slide and topped with a coverslip. The pollen was left to absorb the dye for a minimum of 20 minutes before visualisation. A Zeiss Axioplan 2ie Motorized Microscope was used to identify pollen grains at 400 – 1000x magnification in order to count 100 grains and determine the taxonomic identities. This step was repeated on three slides of pollen for a total of 300 grains per corbicular sample. For identification purposes, pollen reference slides were created from collections at the Natural History Museum, London and by using online databases (globalpollenproject.org), and textbook (Moore et al., 1991; Sawyer et al., 2018).

2.2.3 Leaf tissue DNA extraction, library preparation, and Illumina sequencing

DNA was extracted following the PowerSoil Pro Kit (Qiagen) protocol according to the manufacturer's instructions with the leaf samples disrupted using a SPEX SamplePrep 2010 Geno/Grinder set at 15,000 rpm for 5 min. Concentration of the extracted DNA samples was assessed using the dsDNA BS assay on a Qubit 3.0 fluorometer (Thermo Fisher) and the DNA size distribution was checked with a Genomic DNA Analysis ScreenTape on the TapeStation 2200 (Agilent).

For the library preparation, whole genome libraries were created from 75 plant samples using the NEBNext Ultra II FS DNA Library Prep Kit (New England Bio Labs). DNA fragmentation targeted an average insert size of 600 bp. Adapter sequences were added to the ends of fragmented DNA to generate paired-ends libraries.

In order to achieve a ~1x genome coverage the libraries were pooled based on estimated genome size, obtained from the Royal Botanic Gardens Kew Plant DNA C-values database. Pooled libraries were sequenced on one lane of the NovaSeq 6000 SP flow cell with 2 x 250bp paired end reads at the Earlham Institute in Norwich, UK.

2.2.4 Pollen DNA extraction, library preparation, and MinION sequencing

The tubes containing ethanol and pollen were spun down on a centrifuge and the ethanol was removed by pipette, being careful not to disturb the pollen. Beads from the Powersoil Pro kit (Qiagen) were added to 2 ml microcentrifuge tubes containing both the pollen samples and disrupted with a SPEX SamplePrep 2010 Geno/Grinder at 15,000 rpm for 5 min. DNA was extracted following the PowerSoil Pro Kit protocol according to the manufacturer's instructions. Concentration of the extracted DNA samples was assessed using the dsDNA BR assay on a Qubit 3.0 fluorometer (Thermo Fisher) and DNA size distribution was checked with a Genomic DNA Analysis ScreenTape on the TapeStation 2200 (Agilent).

The samples were adjusted to 400ng of input DNA. The ONT library was created using the SQK-NBD110-24 barcoding kit, following the protocol, and included negative controls that were created using ultrapure water in place of DNA. The runs were sequenced on ONT's GridION for 72 hours using four FLO-MIN106 flow cells and MinKNOW software (v3.6.14). Basecalling and demultiplexing were performed using Guppy (v3.2.8, ONT).

2.2.5 Bioinformatics Pipeline

We used the bioinformatics pipeline of Peel et al. (2019) which involved processing the Illumina and MinION reads and assigning the pollen taxonomy. First, duplicate reads were removed from the 75 Illumina plant reference skims using *NextClip 1.3.2* (Leggett et al., 2014) and *cutadapt 1.10* (Martin, 2011) was used to trim Illumina adapters and remove reads shorter than 100 bp. The MinION reads originating from organelles were removed because they have been found to be highly conserved across species, leading to a higher number of incorrect matches. Organelle reads were identified by aligning the long-read datasets to organellar genomes downloaded from NCBI Entrez using *minimap2 2.7* (Li, 2018) and then removed from the datasets. The genome skim reference database containing 49 species from Peel et al. (2019) was used in addition to the skims library created here in order to make a more complete database of plant taxa.

Using Python scripts adapted from Peel et al. (2019), we used *minimap2* to map the Illumina reads from each plant reference to every long-read MinION sequence. Briefly, *SAMtools 1.7* (Li et al., 2009) removed unassigned reads and calculated the "depth" of the mapping coverage, i.e., the proportion of the MinION read that was covered by reads from one or more Illumina species skims. A MinION read was assigned the plant taxa with the highest percentage coverage, unless the percentage coverage was less than 15%, in which case it was left "unassigned" (see Figure 1.1 for diagram). Additionally, taxonomic groups represented by < 1% of the data were removed.

Due to the high number of MinION reads per pollen load, a large proportion of reads were classified as "unassigned". This is because RevMet reverses the traditional metagenomics paradigm and maps the reference skim reads against the sample (pollen) read. As reads are matched, they are effectively prevented from mapping against another pollen read, thus reducing the potential matches available to each subsequent pollen read. In order to increase the proportion of assigned reads and reduce possible bias in the alignment step, a random set of 5,000 reads were subsampled from each MinION pollen sample and used to map percentage coverage. Five samples contained fewer than 1,000 assigned reads (<20% of reads) and were removed from the statistical analyses.

2.2.6 Statistical analysis

All statistical analyses were conducted in R Version 4.0.2. (R Core Team, 2021).

2.2.6.1 Qualitative comparison

To compare genus richness between sampling methods (RevMet and microscopy), rarefaction curves were produced using the R package iNEXT (Hsieh et al., 2016). Alpha diversity was estimated using phyloseq to calculate observed richness, Shannon diversity and Simpson's richness and tested for significant differences with a Kruskal-Wallis rank sum test.

2.2.6.2 Quantitative comparison

The quantitative abilities of RevMet and microscopy were compared by testing the relationship between the number of RevMet reads and the number of pollen grains assigned to the plant taxa. Kendall's correlation test was used to test the overall correlation for all taxa in the collected samples due to the non-normal distribution of residuals. The relationship between microscope counts and RevMet reads in the three most abundant taxa (*Vaccinium, Salix and Prunus*) were also tested separately using Kendall's correlation test.

2.2.6.3 Spatio-temporal variation in pollen composition

To test the difference in bee foraging patterns between early bloom (March-April) and peak bloom (April-May) of the *Vaccinium* crop, plant richness and Shannon diversity were calculated for the pollen loads, using the RevMet data with the phyloseq function 'estimate_richness'; effects of *Vaccinium* flowering period were tested using Kruskall-Wallace rank sum tests. In order to compare colony-level foraging, the pollen species richness and Shannon diversity for each colony were calculated using the RevMet data.

Seasonal and locational differences in the taxonomic composition of pollen loads were analysed using a MGLM with the package mvabund (Wang et al., 2012). The model best fit a

negative binomial distribution due to the strong mean-variance relationship (Supplementary Figure 2.1). In order to take into account the nested design of the colonies within farm, colony was included as a block treatment. Counts were converted into relative abundance, multiplied by 1000 and converted to integers (Brennan et al., 2019) to account for the variation in read number between samples. Differences in pollen community composition within RevMet and microscopy communities were visualised by performing non-metric multidimensional scaling (NMDS) on a Bray-Curtis dissimilarity index over 1000 permutations, using the vegan package in R (Dixon, 2003). T

2.2.6.4 Unassigned reads

We first attempted to assign all nanopore reads in each sample, however a high proportion were classified as unassigned. As there were significantly higher nanopore read counts than earlier RevMet work (Peel et al., 2019), we reasoned that randomly subsampling the nanopore reads would negate this problem. The original dataset was subsampled at 30,000, 20,000, 10,000, 5,000 and 2,000 reads and processed through the RevMet pipeline using a custom Python script so that the proportion of unassigned reads could be compared. We investigated the relationship between subsampling depth and percentage of unassigned reads using a Friedman test, which is a non-parametric method that tests the difference between groups where there is repeated sampling. Post-hoc analysis was conducted using a pairwise Wilcoxon rank sum test with a Bonferroni correction.

In order to investigate the origin of the unassigned reads, we took those classified as unassigned from the subsampled (5,000 reads) and original datasets from the same nine samples and ran them through RevMet for a second time (Figure 2.1). If the unassigned reads remained unassigned, we hypothesised that these reads belonged to a plant taxon that was not included in our reference database, or were unassignable for a different reason. If the reads could be reassigned, then it is likely the plant skims can be depleted in the assignment process.

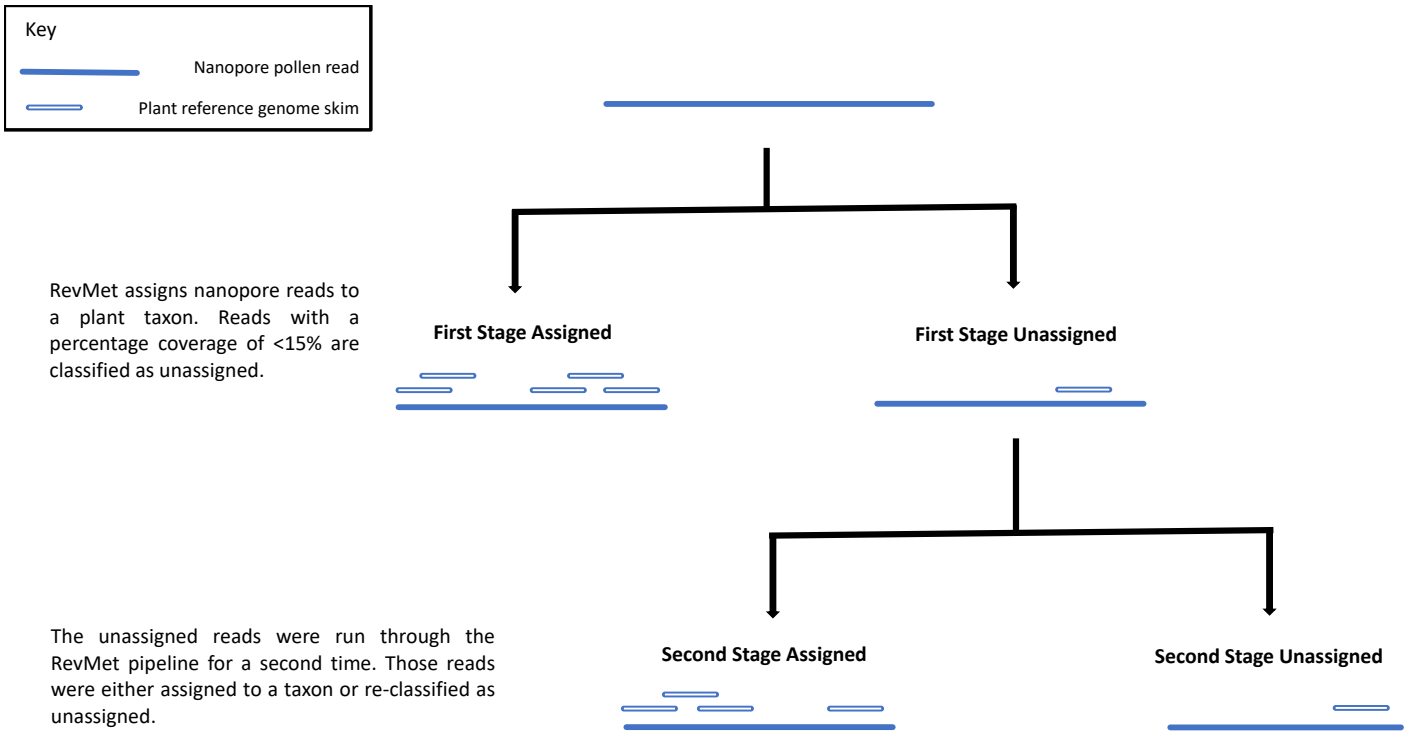


Figure 2.1 The reads classified as unassigned were re-analysed for a subset of pollen samples (n=9). The assigned reads were compared to unassigned reads at the first unassignment step and the second unassignment step to calculate the percentage that could be reassigned.

2.3 Results

2.3.1 Sequencing reads

Plant genome reference skims were generated for all 75 plant species to create an estimated mean coverage of 1.03x (0.3x to 3.6x) based on an average insert size of 600 bp (Supplementary Table 2.1). In each sampling period (early and peak) 12 pollen loads were collected from two colonies on each of the farms, producing a total of 96 samples. The 96 pollen loads yielded DNA quantities ranging from 65 to 5,754 ng. After MinION sequencing and demultiplexing, the mean number of pollen reads per sample was 88,773 and the average read length was 4,927 bp (longest 52,311 bp). A sampling rarefaction curve showed asymptote for the different sampling seasons (Supplementary Figure 2.2).

2.3.2 Qualitative comparison

RevMet identified plant taxa to a lower taxonomic level than microscopy (Table 2.1). All RevMet identifications were made at species-level, but for the purpose of comparison to microscopy, analysis was conducted at the genus-level. Members of the Rosaceae family were difficult to distinguish by morphology using microscopy, which RevMet was better able to do. RevMet identified 11 plant families and 20 genera, whereas microscopy identified nine families and 10 genera (Table 2.2).

There were some discrepancies between the identification of plant taxa between microscopy and RevMet. Three of the pollen samples (samples 38, 75 and 76) were recorded as *Crataegus* in the microscopy results, but yielded very low numbers of assigned reads by RevMet (Figure 2.2). Similarly, samples 52 and 54 were characterised as *Pyrus* in the microscopy results but produced few assigned RevMet reads.

		Early (March / April)								Peak (April / May)							
		Hive A		Hive B		Hive C		Hive D		Hive A		Hive B		Hive C		Hive D	
Family	**Taxa**	**LM**	**RM**	**LM**	**RM**	**LM**	**RM**	**LM**	**RM**	**LM**	**RM**	**LM**	**RM**	**LM**	**RM**	**LM**	**RM**
Adoxaceae	Sambucus nigra																
	Vibernum lantana														0.2		
Apiaceae	Anthriscus sylvestris														0.1		
Asteraceae	Taraxacum		0.5	15.0	15.0	0.2	0.5			0.1		6.6	9.4				
Brassicaceae	Brassica									10.6		0.6		4.2			
	Brassica napus										8.5		1.2				
	Brassica rapa										1.4		3.3		7.5		
	Hirschfeldia incarna										1.2		0.2				
Ericaceae	Vaccinium corymbosum	64.2	57.8	19.1	19.3	2.4	1.9	25.6	24.7	55.7	58.9	34.9	38.1	26.2	32.2	100	99.7
Fabaceae	Ulex europus		1.1				0.1					15.9	15.2	8.2	10.0		
Lamiaceae	Lamium album													8.3	8.7		
	Lamium purpuruem														1.3		
Ranunculaceae	Caltha palustris						0.3				0.3				0.1		
	Ranunculus													8.3			
	Ranunculus acris												0.2		2.6		
	Ranunculus repens										0.1		0.3		7.7		
Rosaceae		0.2										0.9		7.9			
	Crataegus monogyna										1.1	8.3	1.2	8.2	0.4		
	Fragaria		6.6														
	Prunus spinosa		8.1	24.4	23.1	26.6	22.9	1.9	1.3						10.1		
	Pyrus		0.1				0.5			13.6	9.1	32.7	30.5	20.8	9.9		
	Rubus fruticosus		0.2														
Salicaceae	Salix	35.5	26	41.5	42.3	70.8	73.8	72.5	74.0	19.9	19.5			7.8	9.2		0.3

Table 2.1 Percentage of taxa present at over 1% abundance across all sample dates and colonies for light microscopy (LM) and RevMet (RM). Some taxa were only identified to genus or family level for microscopy.

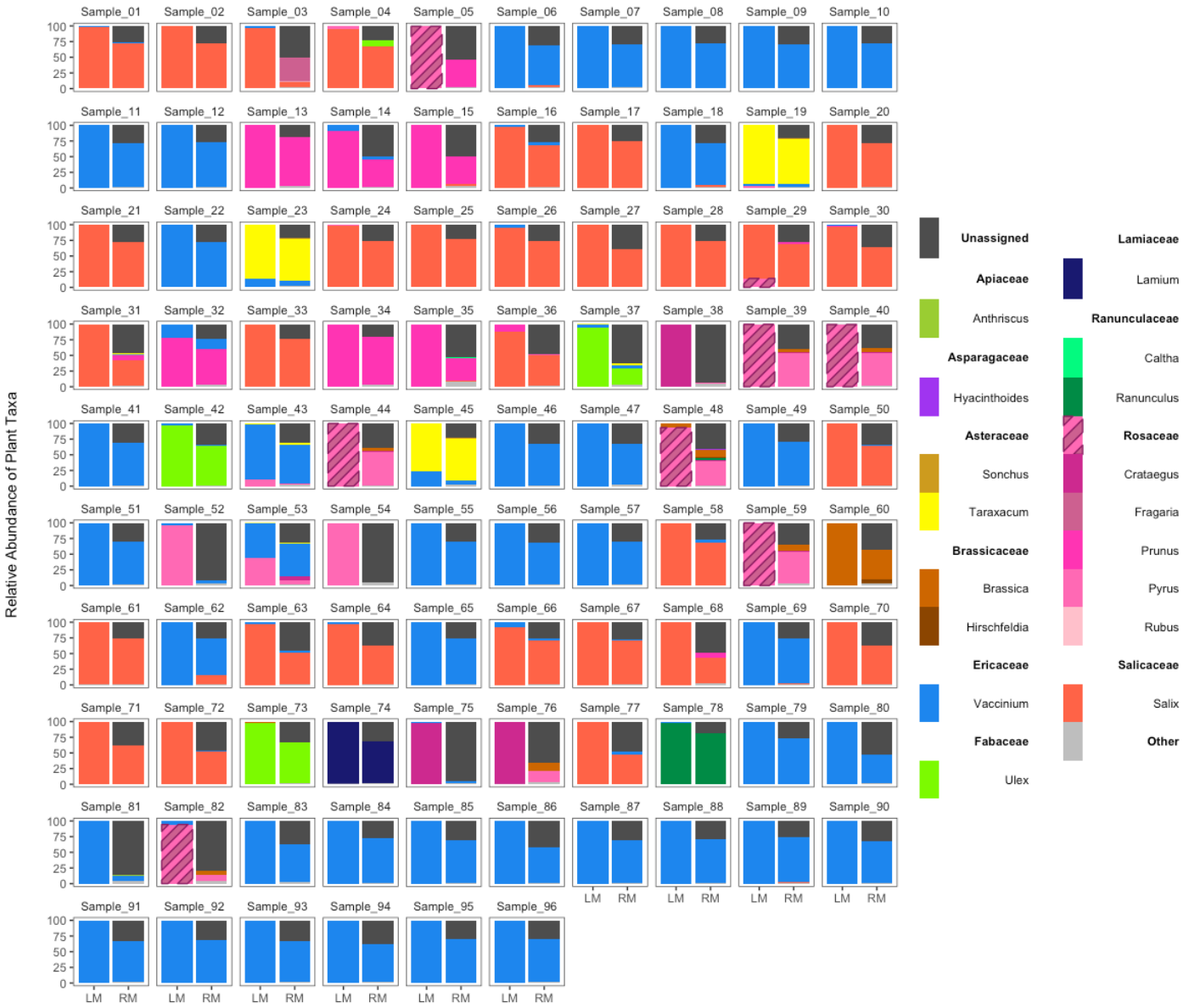


Figure 2.2 Plant taxa in bee-collected 96 pollen loads identified by microscopy (LM) and RevMet (RM). Striped bars represent family-level identification for Rosaceae in the microscopy results, while all other taxa have been identified to genus. Genera present at <1% have been grouped into an "Other" category.

RevMet identified more taxa per pollen load than microscopy ($\chi^2 = 19.5$, $p < 0.001$; Table 2.2, Figure 2.3), and revealed a higher diversity when calculated using Shannon indices ($\chi^2 = 83.0$, $p < 0.001$) and Simpson's index ($\chi^2 = 69.2$, $p < 0.001$; Supplementary Figure 2.3). Rarefaction curves are available for the microscopy and RevMet datasets, split into early and peak crop flowering (Supplementary Figure 2.2).

	Mean number of genera per pollen load (SD)		Number of Genera		Number of Families	
Colony	**RevMet**	**Microscopy**	**RevMet**	**Microscopy**	**RevMet**	**Microscopy**
A	1.91 (1.15)	1.23 (0.53)	14	4	7	4
B	2.35 (1.58)	1.57 (0.79)	13	5	8	5
C	2.18 (1.74)	1.55 (0.8)	14	10	10	8
D	1.42 (0.58)	1.12 (0.34)	4	3	4	3
All	1.96 (1.36)	1.36 (0.66)	20	10	11	9

Table 2.2 The mean number (SD) of plant genera per pollen load, number of genera and number of families identified using RevMet and microscopy. Samples have been pooled across dates to provide colony-level data. Plant taxa with an abundance of <1% in a single pollen load were removed.

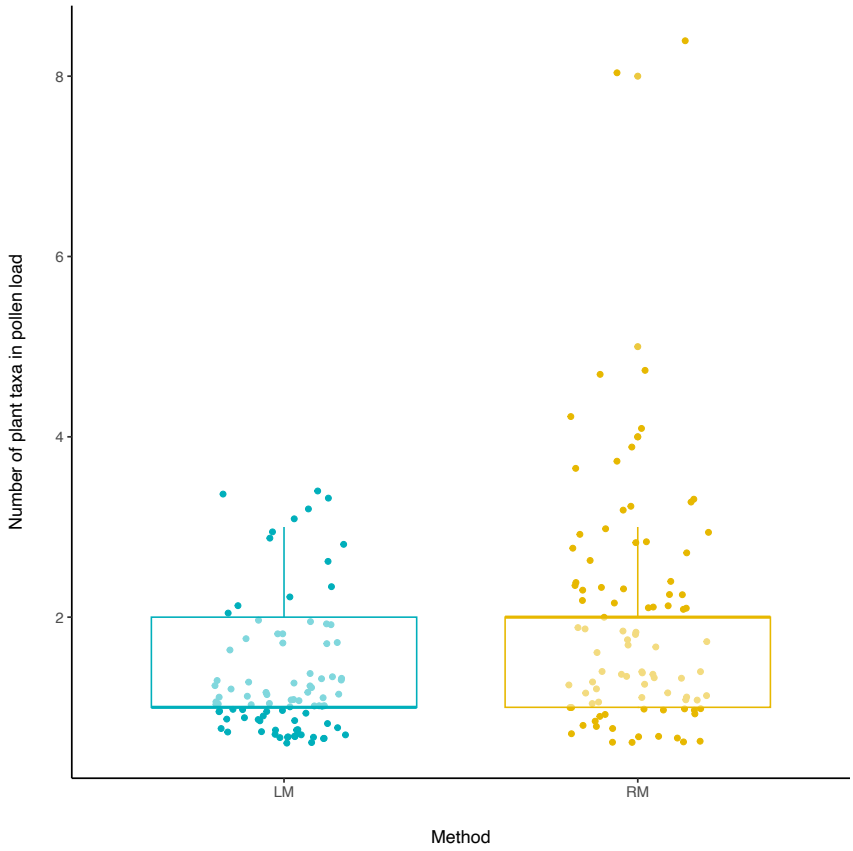


Figure 2.3 The average number of plant taxa per pollen load estimated by microscopy (LM) and RevMet (RM). The medians and interquartile ranges are shown. Data points are jittered around the box plots to visualise the spread of data.

2.3.3 Quantitative comparison

Quantitatively, there was a positive association between the number of reads assigned to each plant genus by RevMet and the number of pollen grains identified using microscopy (τ = 0.59, p < 0.001; Figure 2.4). Significant positive correlations were also obtained by Kendall's correlation test for the top three taxa: *Vaccinium* (τ = 0.74, p-value < 0.001), *Salix* (τ = 0.49, p < 0.001) and *Prunus* (τ = 0.80, p-value < 0.001; Supplementary Figure 2.4). Both methods identified *Salix* as most abundant in the early flowering period, followed by *Vaccinium* and *Prunus*, whilst *Vaccinium* was most frequently identified in the peak flowering period, followed by members of the Rosaceae family and *Salix* (Table 2.3).

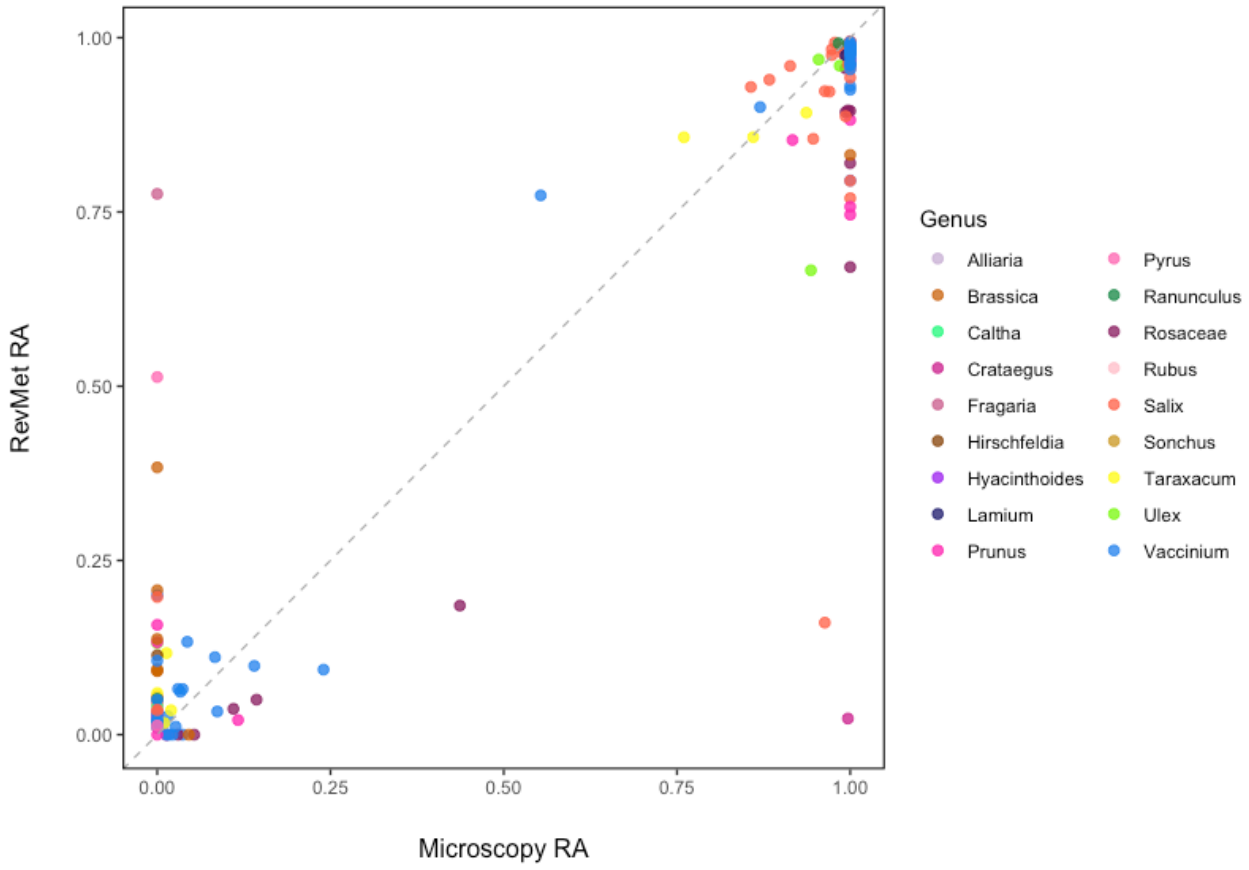


Figure 2.4 Relationship between the estimated relative abundance of plant genera in pollen loads by RevMet and microscopy. Pollen loads are pooled across all sample dates and locations (n = 91). The grey dashed line represents a 1:1 relationship.

Table 2.3 Top five most abundant plant taxa for each sampling period (Early and Peak) and method (RevMet and microscopy) with percentage abundance. Taxa are genus-level except Rosaceae, which was the highest level of taxonomic resolution for some pollen grains identified using microscopy.

	Early				**Peak**			
	RevMet		**Microscopy**		**RevMet**		**Microscopy**	
1	Salix	52.8 %	Salix	44.9 %	Vaccinium	60.8 %	Vaccinium	56.0 %
2	Vaccinium	27.1%	Vaccinium	26.6 %	Pyrus	10.3 %	Rosaceae	17.4 %
3	Prunus	12.5%	Prunus	18.4 %	Salix	6.7 %	Crataegus	10.2 %
4	Taraxacum	4.5%	Ficaria	3.8 %	Ulex	5.6 %	Salix	4.2 %
5	Fragaria	1.2%	Taraxacum	3.7 %	Taraxacum	2.7 %	Ulex	4.1 %

2.3.4 Spatio-temporal variation in pollen composition

The pollen diet of workers displayed differences in composition between colonies, sampling periods and farms (Figure 2.5). As detected by RevMet, *Vaccinium corymbosum* was found in 74 out of a total of 96 (77%) pollen loads, whilst *Salix caprea* was present in 93 (97%), *Prunus spinosa* in 88 (92%) and *Taraxacum officinale.* in 73 (76%).

There was no difference in richness ($\chi^2 = 1.5$, $p = 0.69$) or Shannon diversity ($\chi^2 = 6.9$, $p = 0.07$) of the pollen loads between the colonies. When foraging was compared between the early and peak *Vaccinium* flowering periods, both taxon richness ($\chi^2 = 3.2$, $p < 0.001$), and Shannon diversity ($\chi^2 = 4$, $p < 0.05$) of pollen taxa were higher on the second sampling round than the first (Supplementary Figure 2.5).

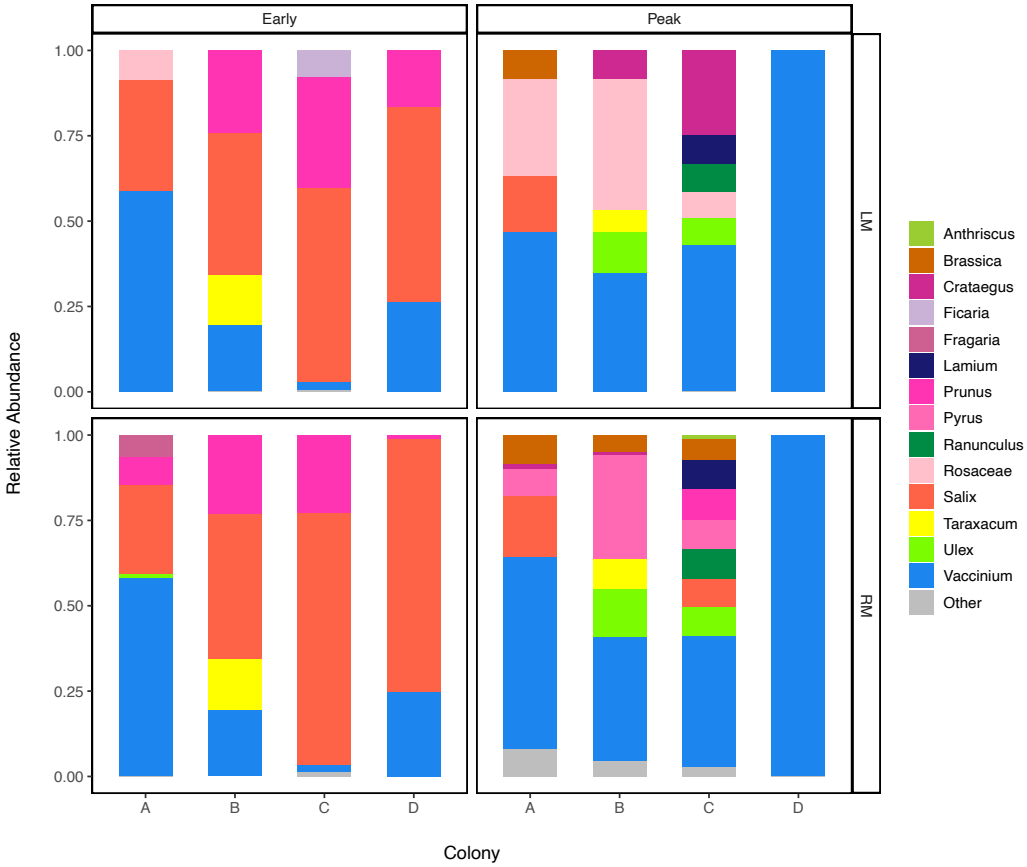


Figure 2.5 Pollen diets for the colonies (A-D) in the early and peak blueberry season collection periods. The pollen has been identified using microscopy (LM; top row) and RevMet (RM; bottom row). Taxa have been identified to genus level, except pollen identified by microscopy in the Rosaceae family. Pollen taxa present at < 1 % abundance have been included in "Other".

When comparing individual bee foragers, the number of plant taxa carried by the workers did not significantly change over the flowering period (Mann-Whitney, $p = 0.87$). The bees collected pollen from an average 1.62 (SD = 0.82) taxa in the early season and 1.73 (SD = 1.28) in peak *Vaccinium* flowering period. Community composition of pollen loads characterised by RevMet changed during the *Vaccinium* flowering period (MGLM, $LR_{188,1} = 433.5$, $p = 0.005$), but did not differ between the two farms (MGLM, $LR_{189,1} = 137.1$, $p = 0.335$). NMDS results show a distinct difference between the pollen samples collected in the early and peak crop flowering periods, which was also observed in the microscopy data (Figure 2.6).

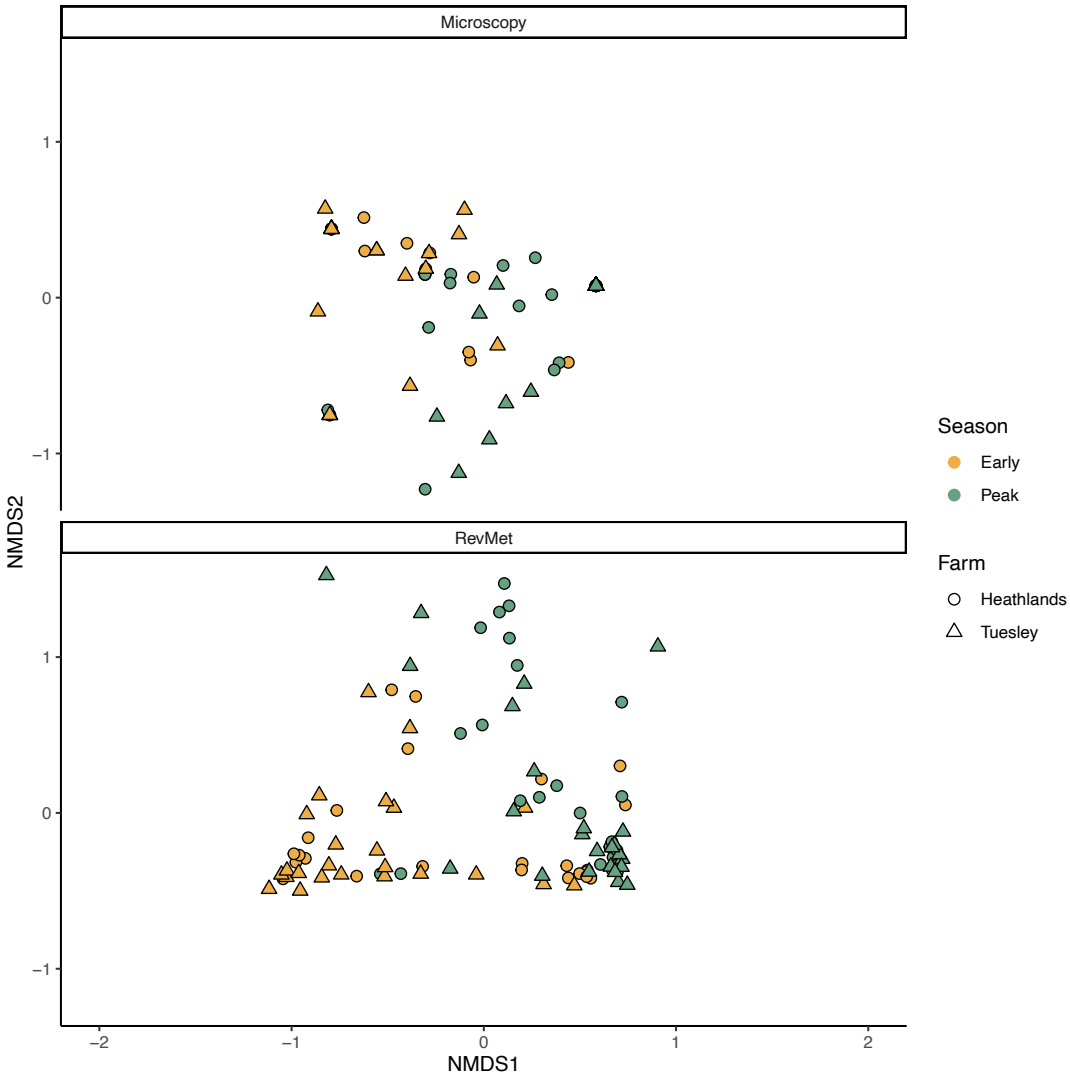


Figure 2.6 Community composition of pollen from early and peak blueberry flowering season for pollen identified using microscopy and RevMet. Each point represents an indiviudal bee's pollen sample (N = 91). NMDS plot based on Bray-Curtis index with 1000 permutations (Microscopy stress = 0.07, RevMet stress = 0.16)

2.3.5 Unassigned reads

We wanted to test the effect of subsampling on the proportion of unassigned reads per sample, to investigate the origin of unassigned reads. The number of reads that were unassigned in the original dataset of pollen loads was on average 41,434 reads per pollen sample (70.2%) and the nine pollen loads selected for subsampling had an average of 58,116 reads per sample (data in Supplementary Table 2.2). There was a significant effect of subsample group on the percentage of unassigned reads ($\chi^2 = 44.5$, $p < 0.001$; Figure 2.7). The pairwise comparisons found a statistically significant rate of assignment between the original dataset with all other subsamples ($p < 0.001$), and the 2,000 and 5,000 subsample groups had a significantly lower unassignment rate than the 30,000 group ($p < 0.05$).

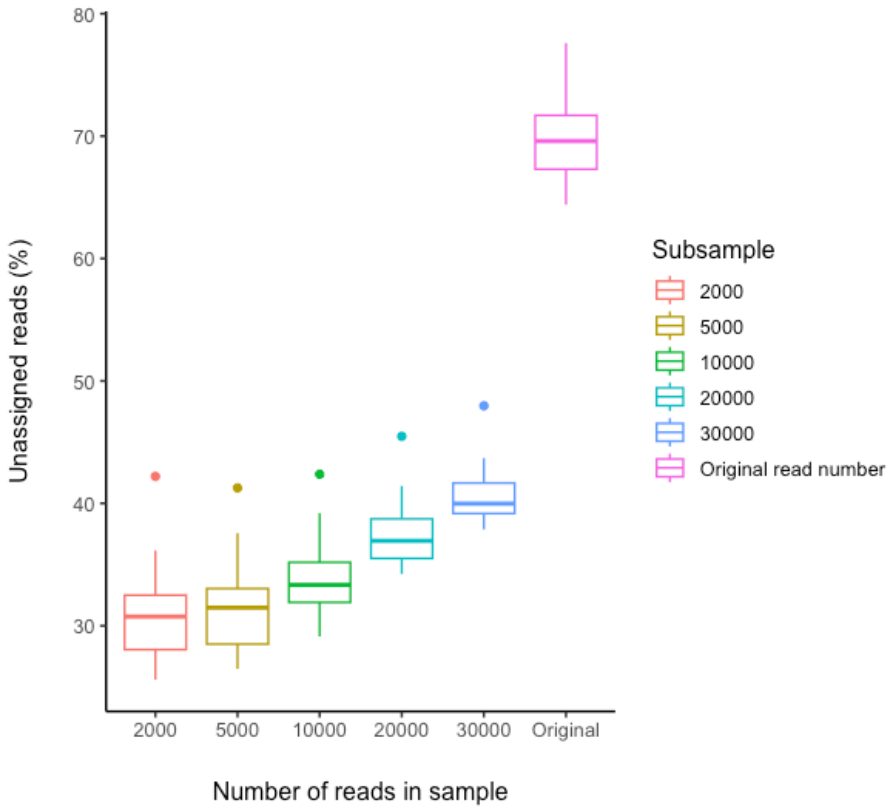


Figure 2.7 The percentage of unassigned reads in each of the read groups subsampled from the original datasets of nine samples. The original dataset was subsampled to 30000, 20000, 10000, 5000 and 2000 reads. Median, IQR and outliers are shown.

To further investigate the unassigned reads, we took the reads classified as unassigned and re-run them through the RevMet pipeline for the original dataset and the 5,000 read subsample group. In the subsampled group the percentage of unassigned reads was 32%, which decreased to 15% when reanalysed using RevMet. In the original (not subsampled) dataset 70% of the reads were unassigned, which decreased to 56% when run through the pipeline for a second time (Figure 2.8, Supplementary Table 2.3). The fact that unassigned pollen reads could be assigned in a re-run suggests that mapping a plant genome skim to a pollen nanopore read effectively removes the skim from the pool available for mapping to subsequent pollen reads, and unassigned reads do not originate from plant taxa missing from the reference database.

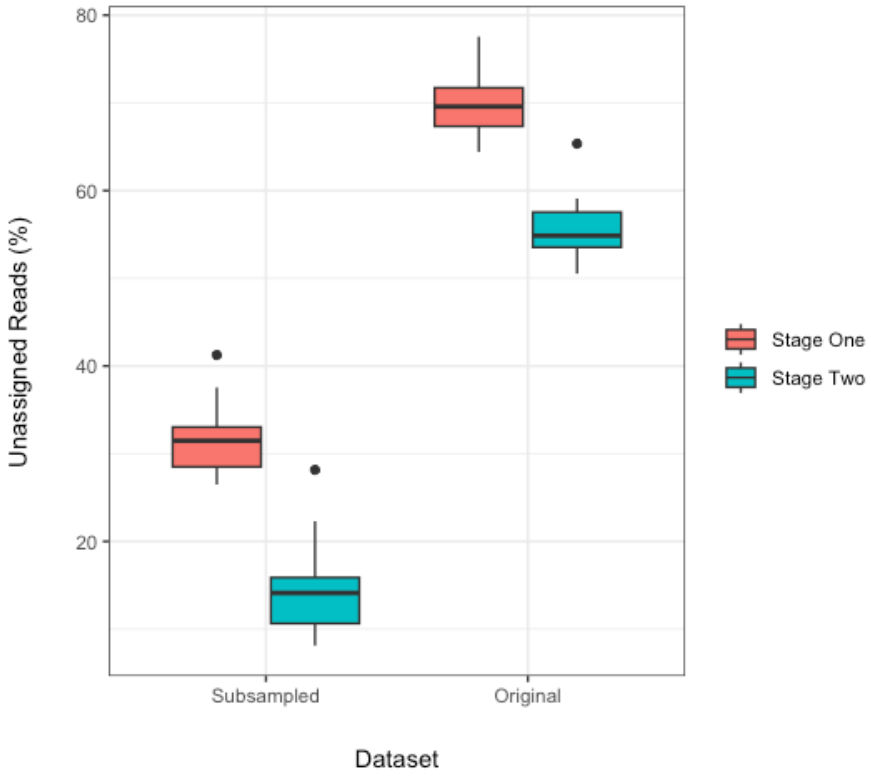


Figure 2.8 The proportion of unassigned reads in the subsampled (5,000 reads) and original dataset (Stage One). The reads classified as unassigned were re-run through the RevMet pipeline to find out what percentage would be reclassified as unassigned (Stage Two). Median, IQR and outliers are shown.

2.4 Discussion

In this study we characterised the diversity and relative abundances of plant taxa in bumblebee-collected pollen loads on two farms growing *Vaccinium corymbosum* crops in the UK, using both the RevMet approach and light microscopy. In comparison to microscopy, RevMet provided a finer taxonomic resolution in terms of genus richness and a semi-quantitative characterisation of pollen loads. There was a significant association between the number of pollen grains identified using microscopy and the number of sequences assigned to taxa via the RevMet approach, providing evidence of RevMet's quantitative abilities. Here we recommend the RevMet approach as a whole genome sequencing (WGS) alternative to metabarcoding in the characterisation of pollen loads. Using this method, we found that foraging bumblebees situated within polytunnels collected the majority of their pollen from non-crop plants in early spring but collected mostly *Vaccinium corymbosum* in the second collection period. Large amounts of pollen originated from woody species commonly found in hedgerows, which is in agreement with other studies (Bänsch et al., 2020; Bertrand et al., 2019), and highlights the importance of these species in bee foraging diets in the spring and early summer. The pollen communities at the start of the *Vaccinium corymbosum* flowering period were different to those sampled approximately one month later, mostly likely due to a shift in the available flowering plant species in the farm landscape.

2.4.1 Qualitative ability of RevMet

RevMet revealed a higher number of plant genera (20) than microscopy (10) and all taxa identified by microscopy were also identified by RevMet. Furthermore, RevMet was able to identify all taxa to species level, whilst only one taxon, *Vaccinium corymbosum*, could be confidently identified to species level using light microscopy. This result suggests RevMet has an advantage over morphological identification in its ability to characterise pollen to a lower taxonomic level (Peel et al., 2019). The recommended number of grains to count for microscopic analysis is 200, which is two-thirds of the number (300) used in this study (Lau

et al., 2017). It is suggested that for pollen taxa present at low abundance 500 pollen grains should be counted, which may have revealed a higher number of taxa in our microscopy samples, although they would still be unlikely to be characterised at species level. Species-level identification for pollen grains is difficult, hence most analyses are performed at genus level or higher (Kraaijeveld et al., 2015). Members of the Rosaceae family in this study were either only able to be identified at family level, using microscopy, or were classified as a different genus to those identified by RevMet. Five samples were classified as either *Crataegus* or *Pyrus* by microscopy but yielded low numbers of assigned reads by RevMet. This could be due to missing reference species in the genome skim database, which would have led to a high number of unassigned reads in the RevMet results. Rosaceae pollen grains have highly similar morphologies which are difficult to distinguish by size and shape alone (Sawyer, 2018). The complex nature of palynology may have led to misallocation of pollen grains to taxa, particularly where similarities in pollen morphology of related species and genera make them difficult to distinguish. Using a molecular approach reduces the likelihood of misidentification, although it is still possible in closely related plants (Peel et al., 2019). We tried to overcome the possibility of false positives by creating reference genome skims targeting a 1x coverage. The genome skims used in Peel et al. (2019) had an average coverage of 0.6x, whereas the skims used in our approach had an average coverage of 1.03x. A higher coverage represents a larger proportion of the genome, which allows a higher degree of discrimination between closely related species. It's possible that identification of pollen grains using microscopy could have been more reliable if the work was conducted by a very experienced palynologist. However, in this study, we aimed to match the level of effort and palynological expertise that is usually employed in studies of pollinator diets based on bee-collected pollen. In most such studies, ecologists rather than palynologists identify the pollen grains, using available keys, images and/or a reference collection of locally occurring taxa (Richardson et al., 2015b; Wood et al., 2017).

2.4.2 Quantitative ability of RevMet

There was a strong positive correlation between the number of RevMet sequences and the number of grains identified using microscopy when all samples were pooled together. *Vaccinium* (651.7 Mbp)*, Salix* (392.0 Mbp)*, Prunus* (637.0 Mbp)*, Taraxacum* (1549.2 Mbp)

and *Ulex* (3822.0 Mbp) were the five most abundant taxa for RevMet and microscopy over the two sampling periods. There was a variation in the genome sizes of the five most abundant plant species present, with *Ulex* (3822 Mbp) having a genome size almost ten times greater than *Salix* (392 Mbp).

A small group of plant taxa dominated the pollen diet, with a high number present at low abundances. We analysed the three most common plant taxa (*Vaccinium corymbosum, Salix caprea* and *Prunus spinosa)* separately to determine whether there was correlation between pollen grain number and RevMet reads. All three taxa showed a strong positive relationship between the read count and pollen grain abundance, providing evidence for RevMet's correlative abilities (Supplementary Figure 2.4).

An advantage of RevMet lies in its PCR-free approach, which removes the biases present in an amplification step. Even so, several factors could still influence the relationship between pollen grain abundance and sequence count, such as interspecific differences in the structure of the pollen wall, the size of the pollen grain and genome size. Other PCR-free whole genome skimming (WGS) methods have also demonstrated more accurate quantitative abilities than metabarcoding (Bell et al., 2021; Lang et al., 2019). Bell et al. (2021) compared results obtained using WGS and metabarcoding, finding the number of reads from the PCR-free approach to be more strongly correlated to pollen grain counts. The current disadvantage of using WGS approaches is the lack of reference sequence databases, but as sequencing costs decrease WGS methods will become more feasible. The RevMet approach demonstrated here has the quantitative advantages of a WGS method, without relying on pre-existing reference databases.

2.4.3 Bee foraging over the *Vaccinium corymbosum* flowering period

Taxonomic identification of bee-collected pollen provides valuable information about bee foraging preferences and the surrounding vegetation. On soft fruit farms where bees are crucial in providing pollination services, the amount of pollen collected from the crop can serve as a useful estimate of how often they are visiting the crop flower. In the early crop

flowering period the bees collected the majority of pollen from *Salix caprea*, and the second most frequently collected pollen originated from *Vaccinium corymbosum*. In the second round of sampling approximately one month later the majority of pollen originated from *Vaccinium*, with pollen from the Rosaceae family the second most frequent overall. *Salix* flowers in early spring and as the crop approaches peak bloom there is a wider availability of *Vaccinium* flowers in late April to early May. A high proportion of plant taxa identified by RevMet in the pollen loads was from Salicaceae and Rosaceae families; predominantly from the genera *Salix, Prunus* and *Pyrus*. The first two of these are mass-blooming taxa often found in hedgerows and semi-natural habitats. Woody and hedgerow plant taxa are important components of pollen diet for bumblebees, even when there is a high diversity of plants to forage from (Bertrand et al., 2019; Kämper et al., 2016). Bumblebees can fly several kilometres to forage for high quality resources, but since this is energetically costly they are likely to forage closer to the colony if there are resources available, highlighting the importance of hedgerow taxa in agricultural landscapes (Redhead et al., 2016).

The average number of plant taxa carried in one pollen load was between one and two, as predicted by both methods. Our results suggest individual bees forage from a few key taxa in large proportions and several others in much smaller amounts, which is in agreement with other studies (Hawkins, de Vere, et al., 2015; Smart et al., 2017). This result supports evidence of flower constancy exhibited by foraging bees, whereby they predominantly visit one species of flower in a single foraging trip (Gruter & Ratnieks, 2011). Bumblebees fly longer distances in landscapes that are simple and less florally diverse in order to fulfil foraging requirements, so it is likely they will encounter more patches of flowering resources on a single foraging trip (Westphal et al.,2006). In a florally abundant landscape, such as a mass-flowering *Vaccinium corymbosum* crop, the requirement to fly longer distances is reduced, which could explain the low taxa richness of the pollen loads.

Bumblebee foraging strategies might be based on the nutritional quality of the pollen, including the levels of protein, lipid, and amino acids (Moerman, Roger, et al., 2016; Somme et al., 2016; Vanderplanck et al., 2014). Pollen that has a high protein content, considered to be over 25%, has been found to promote the growth and development of the larvae (Vaudo, Patch, et al., 2016; Williams et al., 2012). *Vaccinium corymbosum* pollen has an estimated

13.9% protein (Somerville, 2001), which is considered a low-quality pollen, therefore we might expect bumblebees to forage off crop to meet the nutritional requirements of the colony. *Salix* pollen was found to increase pupal mass of *Bombus terrestris* in comparison to micro-colonies fed *Erica* pollen (Ericaceae family), indicating its potential as a higher quality, more attractive resource than *Vaccinium* (Moerman, Roger, et al., 2016). Foraging has a high energetic cost, so there is a trade-off between collecting highly nutritious sources of pollen that are further away and foraging from lower-quality *Vaccinium corymbosum* flowers in the immediate surrounding (Toshack & Elle, 2019).

We endeavoured to collect a high enough sample of foraging bees from the colonies to capture the foraging diet on a single day, but without disrupting the pollen provision of the colony. Pollen loads from bumblebees are collected by actively catching the bee at the entrance to the colony, which requires a higher sampling effort than placing a pollen trap on honeybee hives. We collected pollen samples from 12 bees per colony in order to provide a more complete picture of the colony foraging preferences, which is a larger sample size than other studies have used (Bänsch et al., 2020; Piko et al., 2021). Pollen diversity at the colony level showed a high degree of variability, particularly in late spring, so we recommend for future studies that a greater number of bees samples are used. However, there is a limit to how long a colony can be closed before food restrictions may harm its growth and development, and in periods of low foraging activity the number of bees caught may be less than 12 within a one-hour timeframe.

Sampling bees from the entrance of the colony is more likely to give a realistic estimation of the colony-level foraging, in comparison to sampling pollen from foraging bees caught at patches of flowers or along transects. Bees are more likely to be collecting pollen from the flowers on which they are found, and our understanding of pollen diets based on foraging bees sampled in this way will be biased towards plant taxa that are more likely found on transects, and accessible to observers (Peel et al., 2019; Potter et al., 2019; Wood et al., 2017). Sampling from the colony will include bees that have visited flowers less easily observed from transect walks, for example larger tree species (Allen & Davies, 2023; Donkersley, 2019).

2.4.4 Unassigned reads

Even after removing the five pollen samples that were above the 80% threshold, the proportion of unassigned reads was unusually high (70.2%) across the dataset. The pollen reads generated by MinION could be given an “unassigned” status in the alignment step, where *minimap2* is used to align plant genome reference skims to the long-read pollen sequences. The plant reference genome skim with the highest percent coverage is binned to the pollen nanopore read, unless the coverage is below 15% in which case the read is classified as unassigned.

To test the origin of unassigned reads we repeated the RevMet pipeline on a selection of samples’ unassigned reads, resulting in over one third of the reads being reclassified as ‘assigned’ to a plant reference (see Figure 2.1). The fact that unassigned pollen reads could be assigned in a re-run alerted us that mapping a skim read to a pollen read effectively removes it from the pool of skim reads available for mapping to subsequent pollen reads. Noticing that we had much higher nanopore read counts than earlier RevMet work (Peel et al., 2019), we reasoned that randomly subsampling the nanopore reads would negate this problem. When the original samples were subsampled at different levels, a decreasing proportion of reads were classified as unassigned, from 70% in the original dataset to 32% in the 2,000 read subsampled group. There was little difference between the levels of assignment in the 2,000 and 5,000 subsample groups, so we reasoned the 5,000 subsampling step should be applied to the full pollen dataset because it would be more likely to capture the complete pollen community. From this, we conclude that the high proportion of unassigned reads is an artefact of the RevMet pipeline when there are a high number of reads, which can be reduced by subsampling. The pollen loads we collected had a relatively low species richness, so by subsampling we were less likely to generate false negatives. However, if this method is applied to more diverse samples, such as honey, there should be steps taken to ensure the full breadth of plant taxa are captured.

It is also possible that a proportion of the unassigned reads encountered in the RevMet originated from microbial genomes and other environmental DNA (eDNA) sources (animal, plant etc) that were not included in the reference library. Pollen grains gathered by bees into corbicular loads are moistened using saliva and nectar and transported back to the colony. There is evidence of communities of microorganisms including bacteria and fungi that are present in fresh bee pollen and are present even after an extended period of freezing at -20°C (Mauriello et al., 2017; Pelka et al., 2021). In NGS studies, all DNA is sequenced in a non-targeted approach, which makes it likely that microorganismal DNA and eDNA will be sequenced alongside plant DNA. In our study it's likely that these external sources of DNA present in the pollen loads were sequenced with the pollen DNA, and categorised as "unassigned" in the RevMet alignment process because their genomes were not present in the reference library. To mitigate against the impact of eDNA in the RevMet results, it is possible that a higher number of subsampled reads are required, and this step should be further investigated. Approximately 20% of the reads in Peel et al. (2019) were classified as unassigned in constructed mock DNA mixes, which suggests a proportion of reads will be unassigned that cannot be attributed to having an incomplete reference database. An interesting future step would be to modify the RevMet pipeline into aligning the genome skim pool against each nanopore read individually, which may serve to reduce the level of unassigned reads but could be computationally expensive.

The advantage of subsampling the MinION reads lies in the time efficiency. The time taken for the RevMet pipeline to process nine samples on a High Performance Computing cluster was reduced from 170 minutes for the original dataset to 80 minutes for the 5,000 read subsample dataset, which is an advantage when processing a large number of samples or working on a local network.

2.4.5 Limitations

There are limitations to both RevMet and light microscopy as methods of identifying and quantifying mixed pollen communities. First, there is no true measure of the pollen taxa collected by the bees, because both RevMet and microscopy can miss interactions. Some of

the mismatches that occurred between the two methods were likely due to the misidentification of pollen taxa via microscopy. For example, it is difficult to distinguish between species of the same genera or family using microscopy, particularly for the Rosaceae family. In this study there was a high proportion of pollen from Rosaceae, which could have originated from several genera indistinguishable from pollen grain morphology, such as *Pyrus* and *Prunus*.

Plant leaf samples were collected from a maximum distance of 1000m from the centre of the farms during the *Vaccinium corymbosum* flowering period, a distance that captures most bee foraging journeys (Osborne et al., 1999, 2008). The 75 plants species consisted of crop and wildflowers that were found predominantly on the farm, either in semi-natural areas, woodland, field margins or road margins. The remaining plant genome skims were from the database generated by Peel et al., (2019) and were included to align the pollen sequences against a larger reference database and reduce the possibility of false negatives. We identified five pollen samples with a high proportion (>80%) of unassigned reads after the subsampling step which were subsequently removed from the dataset. The results from morphological identification using microscopy labelled these samples as Rosaceae, which is likely due to their absence from the reference database and therefore absence from the list of identified plant taxa flowering during this period. Garden plants and ornamentals were not included in the collection and therefore could not be included in the RevMet results, even if they were present in the samples. Urban areas, which include gardens and allotments, have been described as pollinator 'hotspots' due to high abundance and richness of flowering plant taxa (Baldock et al., 2019; Samnegård et al., 2011; Tew et al., 2021). In these circumstances, we considered that pollen loads showing a very high proportion of unassigned reads (>80%) were likely to contain a plant species missing from the reference database of genome skims. As costs for sequencing decrease, we can expect whole genome sequencing to become more prevalent and plant reference genomes to increase in availability. Until that point, there is a limitation to these studies that arises from the gaps in the plant genome database.

Another source of bias may arise from the difference in sequencing coverage in the plant reference skims. Although we targeted and achieved an average coverage of 1x, the skim

coverage ranged from 0.28x to 3.55x, which could result in differing powers of discrimination. Peel et al., (2019) found that coverage as low as 0.05x was good enough for species detection, although we are unsure of the effects of closely related species having large differences in coverage. The variation in the genome size of the plant skim library may also have impacted the results of the quantification of the pollen reads. A plant with a large genome will generate a larger skim library than a plant with a small genome, if both are targeting a 1x coverage. In this scenario it is possible that the large genome skim set will produce a higher number of alignments to the pollen reads, therefore inflating its importance in the RevMet results. If this were the case, we might have expected plants with a small genome size, such as *Salix*, to be under-represented in the proportional results. We did not find this to be the case, but the possibility of genome size as a confounding factor deserves further investigation.

The RevMet results had a 1% minimum abundance filter applied in order to remove plant taxa present at low abundances. Without the 1% minimum abundance filter, the species richness of the pollen loads would have been very high, and included plant taxa in the reference library that were not flowering at the time of sample collection. However, it would be preferable to calculate a minimum abundance threshold modelled on sequencing results derived from mock communities and negative controls (Drake et al., 2021).

2.5 Conclusions

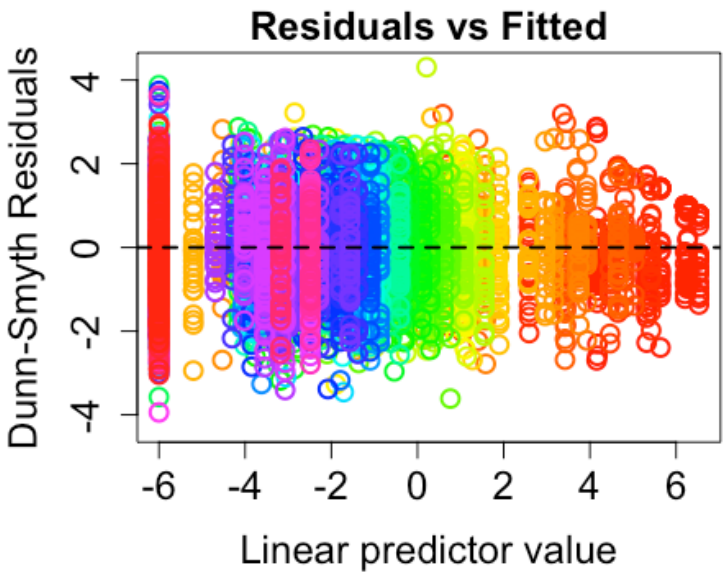
We demonstrate here that RevMet is a suitable alternative to metabarcoding in terms of pollen characterisation and quantification. Successful identification of pollen grains via light microscopy is reliant on a high level of expertise, whereas molecular approaches require little knowledge of the taxa and steadily decreasing sequencing costs. The associations between pollen grain count and sequence counts were variable between taxa, so we recommend further investigation into RevMet as a tool for estimating relative abundance of pollen taxa. This could be achieved by making pollen grain mock mixes of known volumes, a step that has been demonstrated in metabarcoding and whole genome sequencing but not for RevMet (Bell et al., 2019; Lang et al., 2019; Baksay et al., 2020). In ecological applications,

RevMet will be a valuable tool in building knowledge of pollinator foraging and pollen preferences in agricultural systems and wider landscapes.

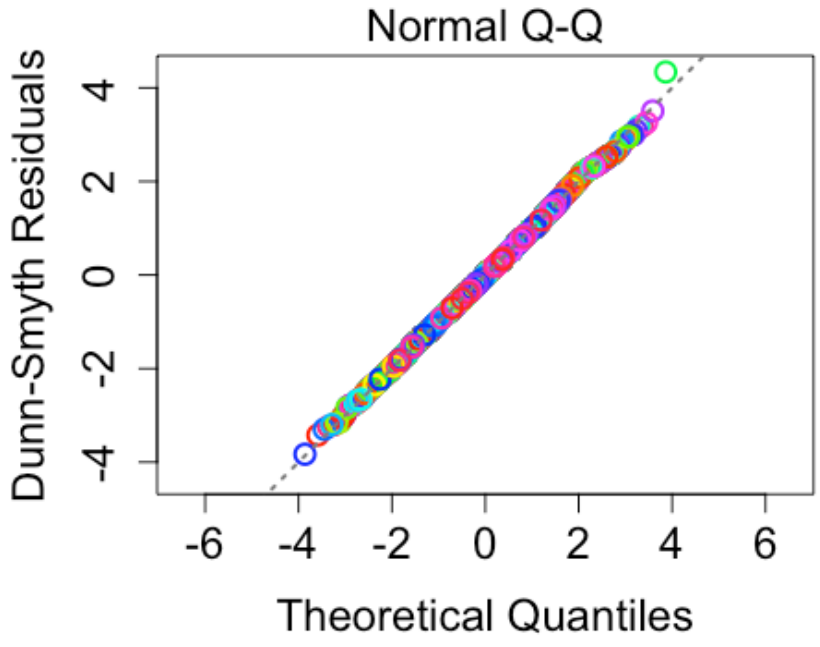
2.6 Supplementary Material

Supplementary Figure 2.1 a) Scatter plot of the Dunn-Smyth residuals from the model and b) Scatter plot of theoretical quantile values and the residuals output from the model used to analyse the RevMet data, using a negative binomial distribution which best fit the data. The plot was produced using the mvabund package in R.

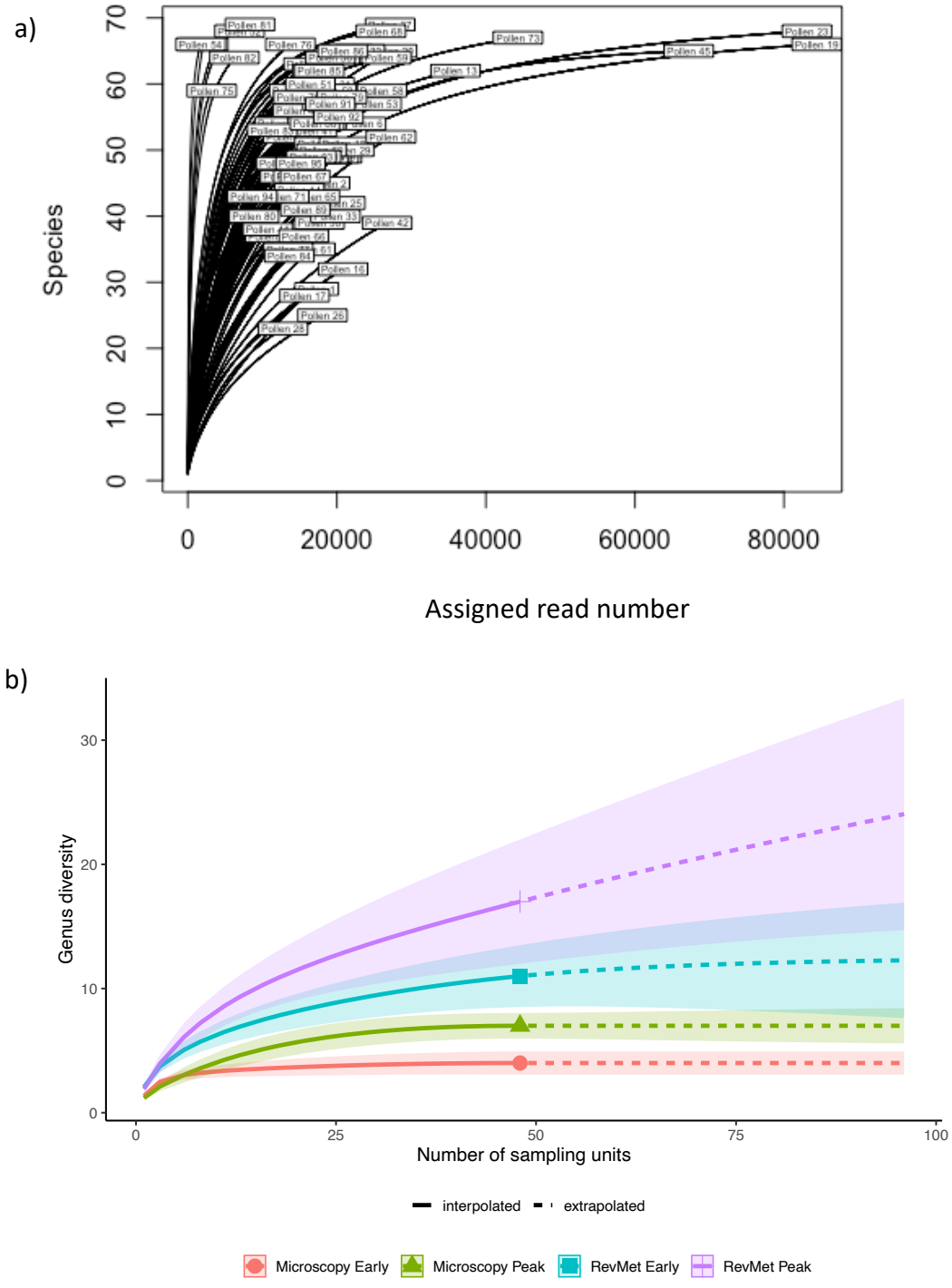
a)



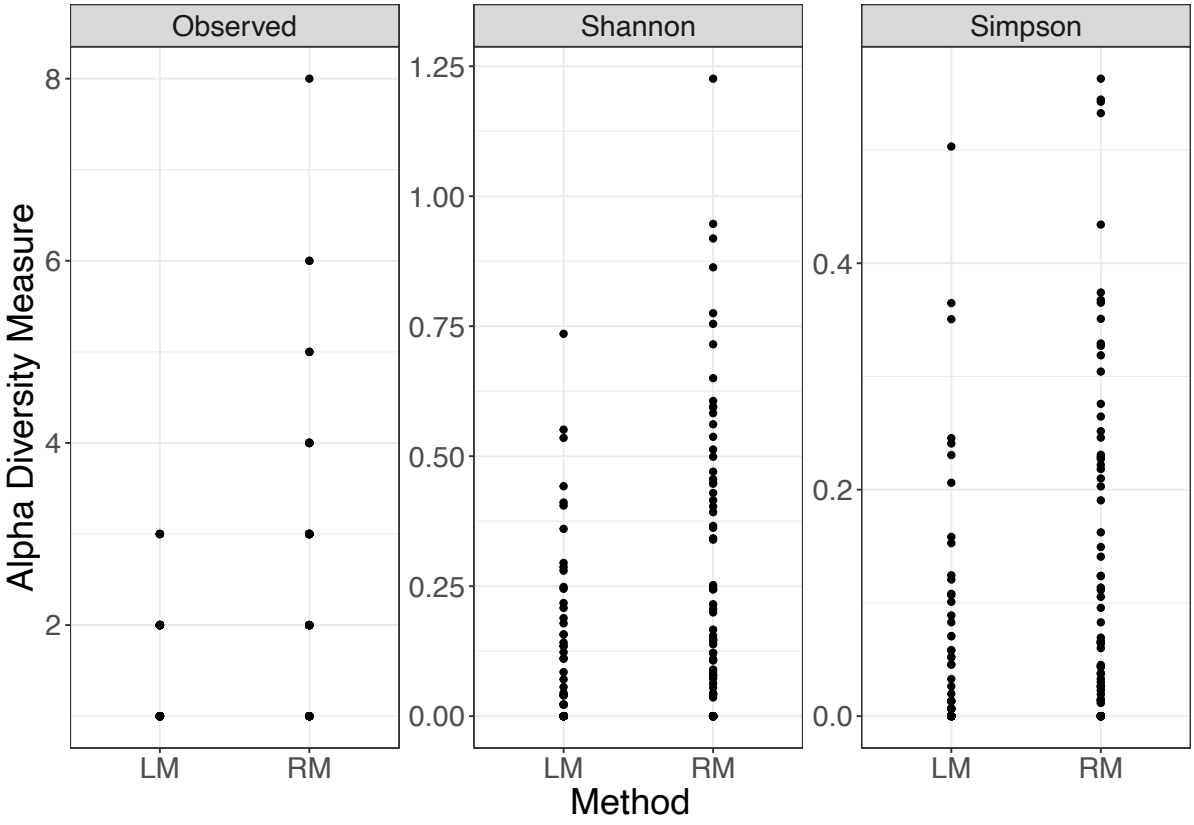
b)



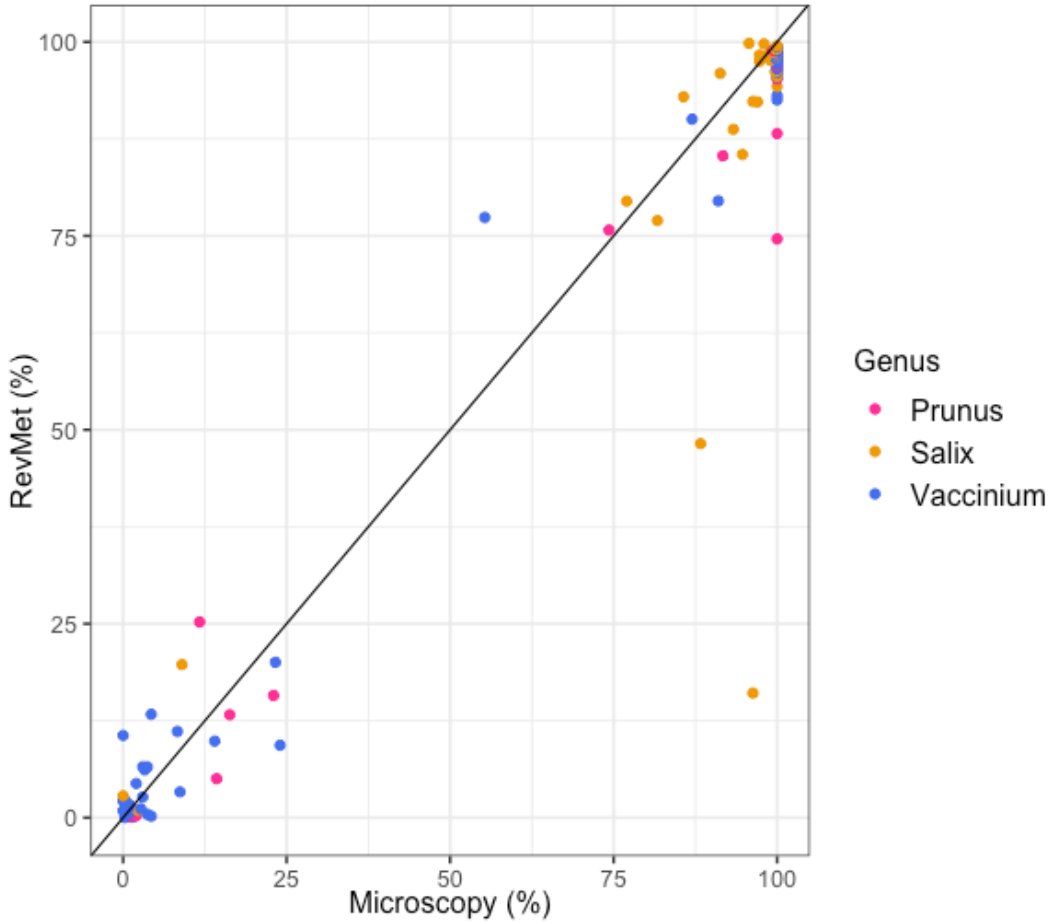
Supplementary Figure 2.2 a) Sample curves for the RevMet sequencing data. Five samples with low sample numbers (< 1000 reads) were removed from the dataset, b) Rarefaction / extrapolation curves for the number of sampling units (pollen loads) analysed using Microscopy and RevMet, separated into Early (n=48) and Peak (n=48) sampling periods. Curves created using iNEXT (Hsieh et al., 2016).



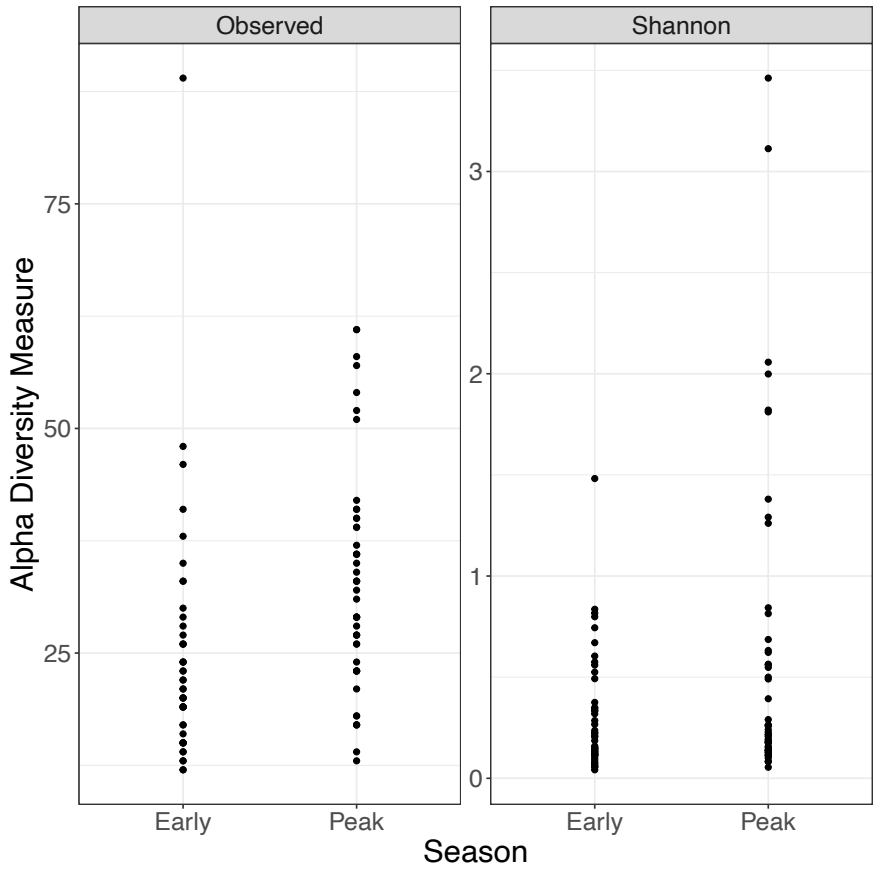
Supplementary Figure 2.3 Taxa richness, Shannon and Simpsons diversity were compared between estimates generated by microscopy (LM) and RevMet (RM). RevMet showed significantly higher levels of observed richness, Shannon diversity and Simpson diversity.



Supplementary Figure 2.4 Relationship between pollen grain abundance and RevMet read count for the top three most commonly occurring taxa: *Prunus*, *Salix* and *Vaccinium*. Correlations were calculated using Kendall's t. The black line represents a 1:1 relationship.



Supplementary Figure 2.5 Taxon richness (observed) and Shannon diversity for RevMet results comparing early and peak *Vaccinium corymbosum* crop flowering pollen taxa diversity.



Supplementary Table 2.1 Genome skim species used for the RevMet reference genome database with associated genome size, read count (pre- and post-processing) and genome coverage (x)

Plant species list	Genome size (Mbp)	Raw PE read counts	PE read counts (post-processing)	Genome coverage (x)
Achillea millefolium	6161.8	16,637,553	13,619,103	1.33
Agrimonia eupatoria	1455.3	2,496,741	2,138,922	0.88
Alliaria petiolata	1323.0	3,703,354	2,496,869	1.13
Anchusa arvensis	4459.0	8,378,088	6,146,665	0.83
Anthriscus sylvestris	2254.0	3,614,850	3,199,879	0.85
Arctium lappa	975.1	1,880,864	1,596,981	0.98
Barbaria vulgaris	270.0	519,947	413,243	0.92
Bellis perennis	1545.0	1,893,338	1,638,548	0.64
Borago officinalis	1666.0	1,989,964	1,804,743	0.65
Brassica napus	1078.0	1,379,322	1,162,704	0.65
Brassica rapa	784.0	988,157	813,097	0.62
Caltha palustris	16170.0	31,251,388	22,484,617	0.83
Calystegia sepium	739.9	1,205,404	1,032,163	0.84
Calystegia silvatica	739.9	1,463,025	1,127,996	0.91
Calystegia silvatica	739.9	434,744	409,418	0.33
Cardamine flexuosa	882.0	478,269	418,436	0.28
Cerastium arvense	970.2	4,545,755	3,974,418	2.46
Cirsium arvense	1391.6	5,523,752	4,816,299	2.08
Clinopodium vulgare	436.1	676,938	584,433	0.80
Convolvulus arvensis	651.7	386,765	328,266	0.30
Crataegus monogyna	744.8	526,499	474,969	0.38
Crepis capillaris	2058.0	3,529,832	3,108,873	0.91
Dipsacus fullonum	3216.4	5,779,097	5,119,088	0.95
Epilobium hirsutum	294.0	717,935	623,378	1.27
Ficaria verna	14217.0	28,047,198	22,774,999	0.96
Fragaria x ananassa	324.6	449,049	331,324	0.61
Geranium dissectum	1722.7	2,139,609	1,988,382	0.69
Geranium molle	1283.0	3,383,910	3,028,989	1.42
Geranium robertianum	1283.0	2,735,627	2,330,592	1.09
Glechoma hederacea	882.0	2,460,127	1,953,307	1.33
Heracleum sphondylium	2141.3	6,729,963	5,688,693	1.59
Hirschfeldia incarna	509.6	715,401	576,853	0.68
Hyacinthoides non-scripta	20776.0	47,677,329	24,605,325	0.71
Hypericum perforatum	574.8	660,219	606,701	0.63
Impatiens glandulifera	1176.0	1,803,807	1,437,975	0.73

Lamium album	1078.0	1,857,373	1,637,701	0.91
Leontodon saxatilis	1024.0	2,308,628	1,578,874	0.93
Linaria vulgaris	852.6	1,257,224	1,136,257	0.80
Lotus corniculatus	1161.3	2,509,094	2,188,446	1.13
Malus sylvestris	759.5	1,096,698	801,308	0.63
Malva sylvestris	1470.0	4,020,090	3,066,540	1.25
Narcissus pseudonarcissus	22344.0	43,784,371	24,972,541	0.67
Pentaglottis sempervirens	1599.0	1,231,120	1,014,266	0.38
Persicaria maculosa	1911.0	1,605,845	1,451,746	0.46
Picris echioides	1176.0	3,030,276	2,668,549	1.36
Plantago lanceolata	1421.0	2,464,210	2,090,044	0.88
Primula veris	475.3	623,476	532,654	0.67
Prunus spinosa	637.0	4,717,213	3,771,932	3.55
Pulicaria dysenterica	965.3	3,602,840	3,111,451	1.93
Pyrus communis	591.7	1,648,288	1,426,446	1.45
Pyrus communis	591.7	1,877,119	1,636,700	1.66
Pyrus communis	591.7	3,479,562	3,000,850	3.04
Rubus fruticosus	438.7	1,606,343	1,440,161	1.97
Rubus idaeus	296.5	427,889	358,029	0.72
Salix caprea	392.0	952,620	783,426	0.82
Salix caprea	392.0	837,756	656,317	0.68
Senescio jacobaea	2254.0	2,148,444	1,894,211	0.50
Silene dioica	2646.0	2,065,158	1,794,485	0.41
Sonchus arvensis	3038.0	6,990,806	5,774,732	1.14
Stellaria graminea	940.8	2,518,664	2,144,891	1.37
Stellaria holostea	1470.0	4,552,429	3,390,618	1.38
Stellaria media	980.0	1,967,621	1,452,226	0.89
Succisa pratensis	2721.5	3,099,242	2,727,800	0.60
Symphyton officinale	1904.5	4,079,216	3,675,251	1.16
Taraxacum officinale	1549.2	2,455,616	2,203,591	0.85
Trifolium repens	1038.8	1,626,689	1,248,162	0.72
Tripleurospermium inodorum	3521.3	8,251,500	7,245,072	1.23
Ulex europaeus	3822.0	15,197,202	12,846,912	2.02
Vaccinium corymbosum	651.7	923,294	786,447	0.72
Veronica chamaedrys	1460.2	2,348,359	2,054,322	0.84
Vibernum lantana	4003.3	4,407,925	3,955,435	0.59
Vicia sativa	2254.0	2,394,498	1,985,433	0.53
Viola arvensis	1500.0	4,604,717	3,216,534	1.29
Viola riviniana	1500.0	2,415,928	2,135,367	0.85

Supplementary Table 2.2 The number of unassigned reads (and percentage of subsample number) for nine barcodes at different subsampling depths. The original dataset is the number of reads sequenced by MinION (with organelle reads removed). No subsampling refers to the number of unassigned reads from the original dataset. Reads were subsampled at five different levels to investigate the number and percentage of unassigned reads.

Subsampled read no.	sample_11	sample_12	sample_13	sample_14	sample_15	sample_16	sample_17	sample_18	sample_19	Average % unassigned
2000	723 (36.2)	520 (26.0)	615 (30.8)	844 (42.2)	594 (29.7)	561 (28.1)	512 (25.6)	650 (32.5)	616 (30.8)	31.3
5000	1878 (37.6)	1331 (26.6)	1574 (31.5)	2063 (41.3)	1509 (30.2)	1425 (28.5)	1324 (26.5)	1640 (32.8)	1652 (33.0)	32.0
10000	3920 (39.2)	2913 (29.1)	3334 (33.3)	4238 (42.4)	3246 (32.5)	3191 (31.9)	2923 (29.2)	3518 (35.2)	3484 (34.8)	34.2
20000	8284 (41.4)	6859 (34.3)	7388 (36.9)	9095 (45.5)	7366 (36.8)	7101 (35.5)	6845 (34.2)	7555 (37.8)	7746 (38.7)	37.9
30000	13110 (43.7)	11705 (39.0)	12046 (40.2)	14391 (48.0)	11850 (39.5)	11752 (39.2)	11362 (37.9)	11993 (40.0)	12498 (41.7)	41.0
No subsampling	20506 (64.4)	29240 (68.2)	44748 (71.7)	71820 (77.6)	40306 (69.6)	39375 (69.8)	32491 (67.3)	35204 (67.2)	59218 (75.6)	70.2
Original dataset (read no.)	31837	42858	62393	92597	57931	56423	48257	52415	78330	

Supplementary Table 2.3 The number and percentage of reads from the original (unsubsampled) and subsampled (5,000 reads) that were classified as unassigned in the first RevMet analysis (unassigned round = “first”), and we ran the unassigned reads through the pipeline for a second time (unassigned round = “second”).

Subsampled	Unassigned Stage	Read number (± SE)	Percentage (± SE)
Original	First	41434 (5220)	70.2 (1.4)
Original	Second	23557 (3668)	55.7 (1.5)
Subsampled	First	1600 (82)	32.0 (1.6)
Subsampled	Second	756 (108)	15.1 (2.2)

Chapter Three

Experimental qualification and quantification of pollen mock mixtures using nanopore sequencing of ITS2 regions and the RevMet approach

3.1 Abstract

Studies on insect-collected pollen provide important information about the foraging behaviour of pollinators that are crucial in providing pollination services in agriculture and for wild plants. A commonly used method in characterising mixed pollen samples is DNA metabarcoding, which utilises PCR-amplification and high-throughput sequencing. The PCR step in metabarcoding introduces bias, which can lead to inaccurate quantification of the relative abundance of taxa within the pollen loads. PCR-free approaches that use whole genome sequencing, such as the recently developed RevMet technique, could offer promising alternatives with the potential to better predict taxon proportions in mixed pollen loads. Nanopore sequencing offers a low-cost and portable alternative to Illumina platforms but has the disadvantage of higher reported sequencing errors. We applied nanopore sequencing in two approaches, RevMet and ITS2 amplicon sequencing, to seven mock mixtures of bee-collected pollen. We examined the qualitative and quantitative accuracies of each method and found ITS2-derived relative abundances to have a stronger relationship to relative abundances of pollen grains estimated by light microscopy than RevMet. We concluded the whole genome sequencing approach that RevMet uses infers a bias that is caused by the amount of DNA yielded by pollen grains of different plant taxa. Therefore, we calculated correction factors based on the genome sizes of the included mock mix taxa and applied them to the RevMet relative read abundances. The quantitative relationship was greatly improved by the correction factor and produced a stronger correlation than ITS2 when compared to pollen grain relative abundances, therefore we recommend further investigation into genome correction factors with a larger range of input taxa. We conclude that nanopore sequencing provides a feasible alternative to other high-throughput sequencing platforms for both amplicon and long-read sequencing in characterising and quantifying the taxa in mixed pollen loads. RevMet is a new approach that can be applied to improve our understanding of pollinator interactions, but should be further investigated to identify potential sources of quantitative biases.

3.2 Introduction

Pollinator declines have been widely reported in the literature, which is of concern due to the ecological importance demonstrated by insect pollinators and their contribution to pollination services (Potts et al., 2016; Powney et al., 2019; Stout & Dicks, 2022). Key drivers of loss include habitat reduction and agricultural intensification, whereby floral resources and suitable habitats are decreasing in quality and abundance (Dicks et al., 2021; Senapathi et al., 2017). There has been an increase in efforts to monitor pollinators and their interactions with plants in order to better understand their floral preferences and provide suitable resources with which to support them (IPBES, 2016). One such monitoring approach involves identifying the pollen taxa collected by insects or present on their bodies, which provides information about the plants most frequently visited by different insect groups. Such studies have been performed on moths (Macgregor et al., 2018), hoverflies (Lucas et al., 2018), solitary bees (Gresty et al., 2018; Sickel et al., 2015b), honey bees (Richardson, Lin, Quijia, et al., 2015a; Richardson, Lin, Sponsler, et al., 2015b) and bumblebees (Bänsch et al., 2020; Leidenfrost et al., 2020).

There are several methods of identifying bee-collected pollen resources. Traditional techniques for pollen identification include light microscopy, a time-consuming method that requires a high level of expertise (Rahl, 2008). Often taxa cannot be identified to species-level, and closely related species can be mis-identified unless using a trained palynologist (Khansari et al., 2012). Over the past decade, molecular techniques such as DNA metabarcoding have become more common, which utilise short regions of the genome to identify species in mixed samples from a reference database of known sequences (Hebert et al., 2003; Taberlet et al., 2012). Multiple barcode markers have been used to identify pollen sources including plastid derived gene markers (*rbcL, matK*) and nuclear regions (ITS2). These have been applied in pollen metabarcoding studies that have investigated airborne pollen allergen sources (Brennan et al., 2019; Kraaijeveld et al., 2015), and plant taxa present in honey samples and bee-collected pollen loads (Bänsch et al., 2020; Hawkins, De Vere, et al., 2015; Jones, Brennan, et al., 2021; Lowe et al., 2022). Metabarcoding is an accurate method for qualitative identification of mixed pollen taxa samples (i.e., the species present), although the results are less reliable for quantifying the relative abundances of different

plant species in the samples (Baksay et al., 2020; Bell et al., 2019; Keller et al., 2015; Smart, 2017). When metabarcoding sequence proportions were compared to pollen grain proportions estimated using microscopy, the relationship was rarely 1:1, with taxon-specific biases observed in the relative abundance results.

Sources of quantitative bias include primer mismatch (divergence in priming sites affect primer-template stability), amplification bias, where some DNA regions are amplified at a different rate than others, and copy number variation, where the number of copies of the target region varies between and among taxa (Stadhouders et al., 2010). Additional bias can occur when pollen grains from different taxa vary in DNA yield at the extraction stage, which could be in part due to pollen structure (e.g. exine wall) or genome size, resulting in a large range of DNA yields between species for a set number of pollen grains (Swenson & Gemeinholzer, 2021).

Plant genomes, and particularly those of angiosperms, vary hugely in terms of size with nuclear haploid genomes ('C-values'), with a more than 2,400 fold difference between the smallest and the largest (Pellicer et al., 2010). The extensive variation in the genome size of plants is a source of continuing research, made easier by developments in sequencing technologies, and there are C-value estimates for over 12,000 plant species (Pellicer & Leitch, 2020). The variation in C-value is likely to affect whole-genome sequencing (WGS) approaches to characterising pollen loads, because species with a large genome would be expected to produce higher yields of DNA in the extraction step, which translates to a higher number of reads in the sequencing step. This potential source of bias could be corrected for using C-values, just as gene copy number has been used to correct for bias in metabarcoding studies (Garrido-Sanz et al., 2022; Vasselon et al., 2018). Correction factors based on C-values have been demonstrated as a method of reducing the quantitative bias exhibited in metabarcoding studies in arthropods (Krehenwinkel et al., 2017; Gerrido-Sanz 2021) and plants (Pawluczyk et al., 2015). On application they adjust the relative read abundance (RRA) estimated by sequencing to improve the strength of the relationship between biomass (e.g. pollen grains) and sequence reads for each taxon (Krehenwinkel et al., 2017; Thomas et al., 2016). The application of correction factors is a relatively recent development, and

depending on the sequencing methodology and scientific question an appropriate correction factor must be chosen, which is a subject of exploration.

The application of WGS approaches could solve some of the problems related to metabarcoding quantification biases. WGS uses the entire genome, which contains many more loci and therefore variation to better discriminate between species. The lack of a PCR step removes the associated amplification bias, which might also serve to reduce copy number bias (Bista et al., 2018). Studies using WGS to identify and quantify mixed pollen samples have reported promising qualitative and quantitative success (Peel et al., 2019, Lang et al., 2019, Bell et al., 2021). Bell et al. (2021) and Lang et al. (2019) used the Illumina platform to generate pollen reads, which are limited by short DNA fragment lengths (150 – 250 bp PE), whereas Peel et al. (2019) used nanopore sequencing to produce long read pollen sequences.

The Oxford Nanopore Technologies (ONT) MinION platform offers longer sequencing lengths than Illumina, and has the added advantage of being low cost and portable. Nanopore sequencing has shown promise in ecological applications for taxonomic identification, which has been demonstrated in water samples (Reddington et al., 2020), microbial samples (Gonçalves et al., 2020) and bee-collected pollen (Leidenfrost et al., 2020; Peel et al., 2019). Although this platform has been continuously developed over recent years, there are still concerns over the sequencing error rate, which on first release was over 10% (Ip et al., 2015). However, as the sequencing and basecalling technology has improved the error rate has decreased, with recent published reports of 5% global error rates (Delahaye & Nicholas, 2021), while ONT claim even lower.

Reverse Metagenomics (RevMet) is a method that maps short read genome skims to long read nanopore sequences for the identification of pollen taxa in mixed species samples (Peel et al., 2019). Thus far, RevMet has been tested quantitatively using mock mixtures of extracted DNA, which removes sources of extraction bias, but has not been tested using pollen grains as the mock mixture input. Previous studies characterising mixed pollen samples have compared amplicon sequencing generated by nanopore to Illumina (Leidenfrost et al., 2020), and Illumina metabarcoding to whole genome shotgun sequencing

(Bell et al., 2021). Leidenfrost et al. (2020) found comparable results when using Illumina and MinION to sequence the ITS2 region, and Bell et al., (2021) found WGS provided a stronger correlation with the microscopic grain counts than *rbcL* and ITS2 barcodes. To our knowledge, no studies have compared pollen amplicon sequences to WGS long reads using nanopore for both datasets.

In this study we compare a WGS method (RevMet) to a metabarcoding approach (ITS2 marker) in order to compare their abilities to identify plant species and quantify relative abundances against pollen grain relative abundances estimated using microscopy. First, we compare the qualitative abilities of RevMet and metabarcoding (using the ITS2 marker) by counting the number of false positives and false negatives produced by each approach. Second, we compare the quantitative correlations of the microscope pollen grains counts with each of the RevMet and ITS2 datasets at different taxonomic levels. Third, we calculate a correction factor estimated using genome size (GCF: Genome Correction Factor) and apply it to the RevMet results. We hypothesise that a correction factor adjusts the RevMet DNA read proportions, providing quantification of species' relative abundances that resembles a stronger positive relationship with microscopy than the original RevMet data. The GCF was applied to RevMet only, as genome size is most likely to affect a WGS approach and there are several examples of correction factors that have been applied to metabarcoding studies (Krehenwinkel et al., 2017; Thomas et al., 2016; Vasselon et al., 2017).

3.3 Materials and methods

3.3.1 Pollen samples

Corbicular loads were sampled from bees in 2019 and 2021, following methods described in Chapters Two and Four. Pollen samples were chosen that had over 70% read assignment, and over 97% of the reads were assigned to a single species, as estimated by RevMet analyses in Chapters Two and Four (sample metadata in Supplementary Table 3.1). As previously discussed in Chapter Two, there is an average unassigned rate of ~30% among all pollen samples, and with a higher unassigned rate there is a possibility of the sample containing reads that are not included in our reference database, and therefore less likely to be a monofloral sample. We selected six taxa that were present in multiple samples and exhibited a range of genome sizes to test the effect of C-value on the quantitative abilities of RevMet (Table 3.1).

Pollen samples were verified using microscopy, in order to confirm the species identity and ensure there were no secondary taxa. Pollen grains were suspended in a mixture of 1:3 glycerol:ethanol, which has been found to possess the best viscosity for pipetting and preventing pollen grains from quickly sinking to the bottom of the tube (Bell et al., 2019). 10 μl of the glycerol/ethanol mixture containing the suspended pollen load were pipetted onto a microscope slide and topped with a coverslip. A Zeiss Axioplan 2ie Motorized Microscope was used to identify pollen grains at 400 – 1000x magnification in order to count all visible grains and determine the taxonomic identities. In order to create the mock mixtures, we first tried to measure the number of pollen grains per μl. The high density of pollen grains made the intended proportions of taxa in the mock mixtures difficult to create accurately using a micropipette because there was a large variation in pollen grains per μl. Instead, we created the mock mixtures of the different species first, and then subsequently counted the number of pollen grains belonging to each taxon (Bell et al., 2018; Supplementary Table 3.2). Each mock mix had six replicates in total, which were averaged and converted into relative abundances for the taxon mock mixture estimates (Table 3.2).

Table 3.1 Taxonomic identities of mock mixture taxa and their associated genome sizes.

Family	Species	Genome Size (Mbp)
Asteraceae	*Taraxacum officinale*	1549.2
Ericaceae	*Vaccinium corymbosum*	651.7
Fabaceae	*Ulex europaeus*	3822.0
Ranunculaceae	*Ficaria verna*	14217.0
Rosaceae	*Prunus spinosa*	637.0
Salicaceae	*Salix caprea*	392.0

Table 3.2 The relative abundances of pollen grain taxa in the mock mixtures. Pollen grains were counted after the mixtures were made, to reduce biases caused by pipetting inaccuracies.

Mock Mixture	Ficaria verna	Prunus spinosa	Salix caprea	Taraxacum agg.	Ulex europaeus	Vaccinium corymbosum
1	31.9	20.7	33.9	0	0	13.6
2	31.5	17.8	37.2	0	0	13.4
3	28.7	21.6	37.3	0	0	12.4
4	0	0	41.5	0	43.8	14.7
5	0	0	36.4	0	63.6	0
6	10.6	3.96	82.8	2.21	0	0.39
7	18.6	8.04	68.7	4.03	0	0.67

3.3.2 DNA extraction

Once the relative abundances of different pollen species in the mock mixtures had been counted, the suspended pollen mixture was transferred to a new collection tube and spun down on a centrifuge at 15,000 rpm for 1 minute so that the ethanol could be removed without disturbing the pollen grains. DNA was extracted using a modified bead washing protocol (Heavens et al., 2021). Using reagents from the DNeasy PowerSoil Pro Kit (Qiagen), 150 mg of beads and 400 ul of CD1 lysis buffer were added to the pollen tube and disrupted using the GenoGrinder for 5 minutes at 15,000 rpm. The tubes were spun in a centrifuge for 1 minute at 15,000 rpm and the supernatant was transferred to a fresh 1.5 ml Lobind Eppendorf tube. 400 µl of KAPA Pure Beads (Roche Diagnostics Ltd, West Sussex, UK) were added to 400 µl of the lysed and bead beaten pollen grains, which were vortexed and incubated at room temperature for 5 minutes. The tubes were pulse spun in a microcentrifuge and placed on a magnetic particle concentrator (MPC) for five minutes, to allow the beads to concentrate. The supernatant was discarded and the beads were washed twice with fresh 70% ethanol. Care was taken to remove all the ethanol and the tube removed from the MPC. The beads were resuspended in 10 µl of Qiagen CD6 buffer and incubated at room temperature for 2 minutes. The tube was then pulse spun in a microfuge then placed in a magnetic particle concentrator for 5 minutes to allow the beads to concentrate. The supernatant containing the DNA was then transferred to a fresh 1.5 ml Lobind Eppendorf tube. The concentration of eluted DNA was assessed on a Qubit fluorometer using the dsDNA BR assay kit (Life Technologies, Loughborough, UK).

3.3.3 Amplicon sequencing

ITS2 regions were amplified using a one-step protocol and region-specific primers (Chen et al., 2010)(Supplementary Table 3.3). For the ITS2 amplicon sequencing we created three separate reactions which has been found to reduce PCR bias (Sickel et al., 2015). A negative control was also included in each round of PCR. The PCRs were 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), 7.2 µl of molecular grade H_2O, and 1 µl of

template DNA. The ITS2 PCR cycle was as follows; initial denaturation at 94°C for 5 minutes, followed by 40 cycles of denaturation at 94°C for 30 seconds, annealing at 56°C for 30 seconds and elongation at 72 °C for 45 seconds. Final extension was performed at 72°C for 10 minutes. The PCR products were cleaned using Agincourt AMPure XP beads (Beckman Coulter) using a 1:1 ratio. Products were quantified using a Qubit 3.0 (Thermo Fisher Scientific) and pooled at equal concentrations. The negative extraction and PCR controls from each plate were sequenced with the pollen samples (Supplementary Table 3.4).

3.3.4 MinION nanopore sequencing

The ITS2 amplicon products were sequenced on a single MinION flow cell, with three replicates for each mock mix. The samples for RevMet analysis were pooled on a separate flow cell with two replicates per mock mix. WGS sequencing requires a higher volume of eluted DNA than metabarcoding because there is no DNA amplification step, therefore we had to use fewer replicates for the RevMet analysis.

The pooled libraries for both RevMet and ITS2 sequencing were generated with the SQK-LSK109 kit and Native Barcoding Expansion kit (EXP-NBD104) according to the manufacturer's instructions, with an input of 50 ng of extracted DNA for the RevMet samples and 150 ng for the ITS2 samples. The pooled library was sequenced on a GridION using a FLO-MIN106D flow cell with the standard 72 hour run script (MinKNOW v21.05.12). Reads were basecalled on-instrument by Guppy (v5.0.12) with the High Accuracy (HAC) model and a minimum pass read Q score of 9.

3.3.5 Bioinformatics

The RevMet samples were processed based on the pipeline written by Peel et al. (2019). Briefly, a library of short read Illumina skims are created from plant tissue to form the reference genomes. Pollen sequences generated by ONT MinION are long read sequences that form the query reads. The RevMet pipeline maps the short read plant skims against

each nanopore long read to calculate a percent coverage. The plant reference species with the highest coverage is assigned to the individual nanopore read, which is the reverse of the standard metagenomic approach of mapping query reads to assembled reference genomes.

The ITS2 raw sequences were analysed using MARTi (Metagenomic Analysis in Real-Time, https://marti.cyverseuk.org), a newly developed open-source tool for analysis and visualisation of metagenomic sequencing data (Peel et al., in prep). MARTi processes FASTQ or FASTA sequence files through an engine, presenting results through a web browser interface known as the MARTi GUI. First, the reads are filtered to remove short and low-quality reads, the parameters of which can be decided by the user. We chose to filter out reads shorter than 200 bp, as these could represent nanopore adapter sequences and short sequences, which might produce false assignments. Next, MARTi classifies reads to assign plant taxa using a combination of BLAST and a lowest common ancestor (LCA) algorithm. The LCA algorithm assigns taxa based on several parameters, including BLAST bit score, length of match, percent identity of match and the maximum number of hits, all of which were configured within the MARTi engine. Our chosen parameters were reads matching at least 100 bp, with a similarity of over 90% up to 20 hits. The LCA algorithm classifies a read to the lowest taxonomic level consistent with all "good" hits (i.e. fulfilling bit score, identity, length etc. parameters). The reads were blasted against a custom database created from ITS2 plant sequences, compiled from the same list of species used in the RevMet analysis. Sequences for the ITS2 region were searched for using *NCBIEntrez* and used to create a custom library of 1,884 FASTA sequences. We used SeqKit to identify and remove duplicate sequences, but none were identified (Shen et al., 2016). MARTi output was analysed at the family-, genus- and species-taxonomic levels.

3.3.6 Statistical analysis

All statistical analyses were carried out at species-, genus- and family-level taxonomies for the ITS2 data, and species-level for the RevMet data, which could be used to infer genus- and family-level results.

The false negatives detected in ITS2 and RevMet results were recorded for all replicates to determine their ability to identify the taxa present in the mock mixtures.

To determine the false positive rate we calculated the taxon richness of each mock mixture for the RevMet samples and the ITS2 samples. The number of assigned sequences for each taxon and replicate were summed, and the richness was calculated for each mock mixture. A 1% minimum abundance filter was also applied, as is common in eDNA analyses, because a large number of taxa are commonly assigned low levels of reads that represent false positive results.

The assigned reads for each taxon were summed across replicates to provide a single relative read abundance (RRA) for each mock mixture. Using RRA for sequence abundances allows for comparison to the relative abundance of pollen grains estimated via microscopy and was calculated as the proportion of reads for each taxon present out of the total number of assigned reads for the sample.

To determine which method had the strongest relationship to the mock mixture relative abundances, we investigated the relationship between sequence proportions and the relative abundance of pollen grains estimated using microscopy. We used linear regression to measure the relationship between RRA for each plant taxon and the relative abundance of grains in the mock mix measured using microscopy, providing a statistic (R^2, or percentage of variance explained) for overall quantitative ability. The regressions were calculated using the 'lm' function in base R (version 4.2.2; Team R. Core, 2013).

Given the improved ability of WGS approaches over amplicon sequencing to quantify plant relative abundances of pollen shown by others (Bell et al., 2021), it was surprising that RevMet's quantitative ability was weaker than ITS2 (Fig. 3.3). Considering our plant taxa had a range of genome sizes, we looked at the relationship between genome size (Mbp) and the over- or under-estimation of RevMet proportions in comparison to the microscopy proportions. This 'percent change' was calculated as the RevMet proportion subtracted by the microscopy proportion. Using the genome sizes, we created a correction factor for the

RevMet proportions to see if it would improve the quantitative relationship between RRA and relative abundances estimated used microscopy.

We created a genome correction factor (GCF) for each of the six plant taxa included in the mock mixtures, using Mbp values extracted from the Kew Plant C-values DNA database (Pellicer & Leitch, 2020):

$$GCF_k = \frac{C_k}{\sum C_n}$$

$$GCF\ RevMet\ RA_k = \frac{RevMet\ RA_k}{GCF_k} \times \frac{1}{\sum GCF_n}$$

To calculate the GCF for a single taxon *k* the genome size *C* (in Mbp) of taxon *k* is divided by the sum of all genome sizes (*n*) for the taxa present in the mock mixture. This number GCF_k is then used to divide the original estimated RevMet proportion for the specific taxon $RevMet\ RA_k$ and calculated as a proportion of the sum of all GCF_n in the mock mixture. The recalculated proportions were used in regressions to measure the strength of the relationship between GCF proportions and the proportions of mock mixture pollen grains. We used the 'lm' function in R to calculate slope values and the variance (R^2).

3.4 Results

For the RevMet samples, we obtained 1,058,818 reads, ranging from 10,633 to 146,234 per sample. The average read length was 1834 bp, and maximum read length was 161,653 bp**.** The ITS2 samples generated 1,224,013 reads, ranging from 30,593 to 104,510 per sample with an average read length of 596 bp.

We calculated the different read lengths for the six taxa from the RevMet dataset (Figure 3.1). The mean fragment lengths were *Vaccinium* (1890 SE ± 39.7), *Ficaria* (1854 ± 8.8), *Salix* (1341 ± 17.3), *Prunus* (1318 ± 32.4), *Ulex* (1218 ± 15.2) and *Taraxacum* (953 ± 40.9).

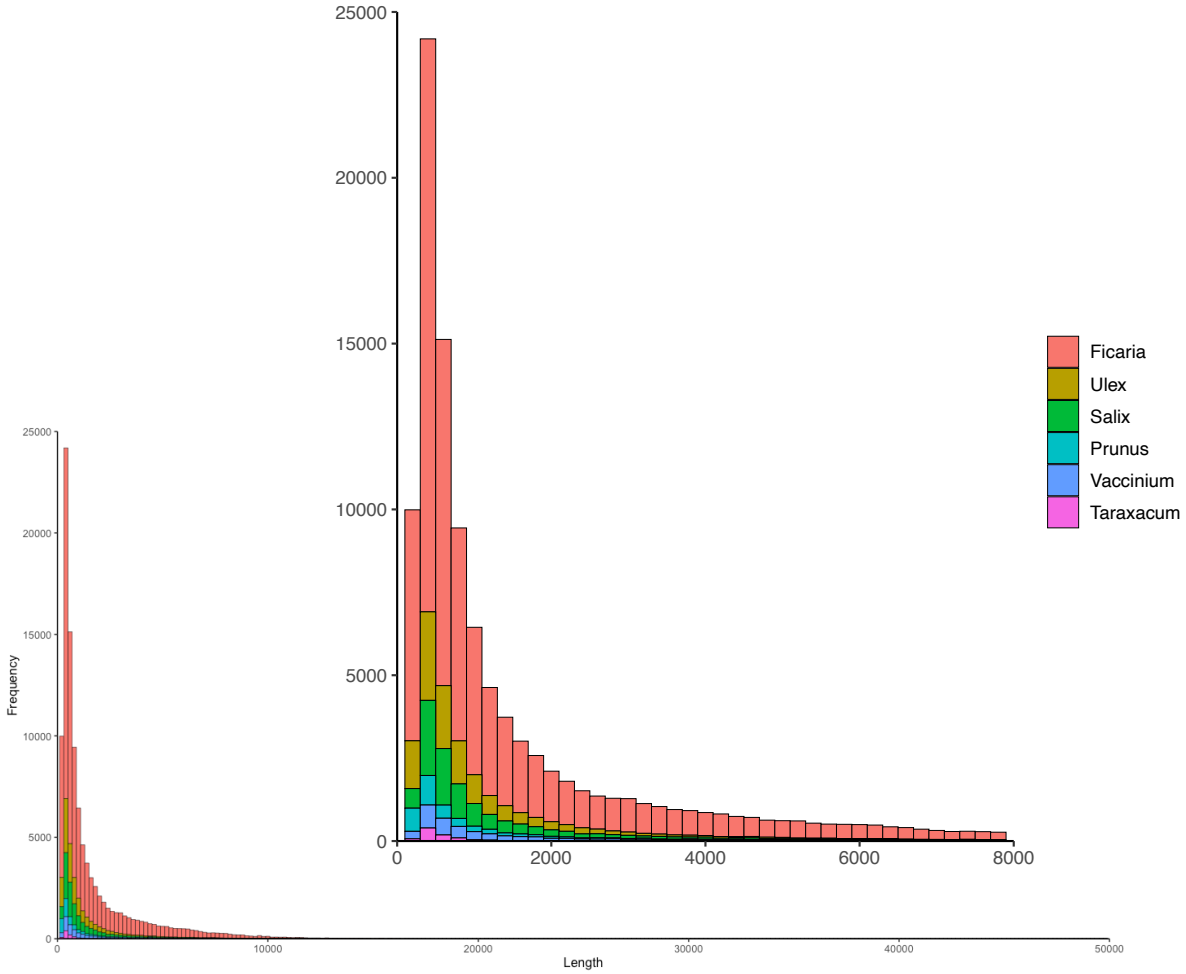


Figure 3.1 Histograms of read lengths for each of the mock mixture taxa.

3.4.1 Qualitative comparison

To test the qualitative ability of each method to identify plant species, we calculated the number of false negatives and false positives estimated. The RevMet approach had zero false negative detections, and identified all taxa present in the seven mock mixtures correctly at species, genus and family levels. ITS2 sequencing also identified all taxa at each of the taxonomic levels, producing zero false negative results (Supplementary Table 3.5).

There were false positive detections in the mock mixture samples for both RevMet and ITS2 approaches in the filtered and unfiltered datasets (Table 3.3). When no filter was applied, RevMet produced a high number of false positives, with a range of 32 – 49 species present in the mock mixtures. When a 1% minimum abundance filter was applied, RevMet accurately predicted the richness for mocks 1,2,3 and 5. Mock mixtures 6 and 7 contained *Vaccinium corymbosum* pollen grains at relative abundances of 0.39% and 0.67%, respectively. RevMet accurately predicted the proportions at less than 1%, and as a result when the 1% filter was applied *Vaccinium corymbosum* was removed from the dataset. Mock mixture 4 contained three species, *Salix caprea, Ulex europaeus* and *Vaccinium corymbosum* that were all correctly identified. However, *Prunus spinosa* was also identified at a relative abundance of 1.19%, and was therefore not removed by the 1% filter.

ITS2 richness was calculated for the species, genus and family taxonomic levels (Table 3.3). With no filter, the richness ranged from 4 – 7 for the genus and family results, and 3 – 6 for the species results. Applying a 1% filter only improved the accuracy for mock mixture 6 at the genus- and family-levels, by removing a low abundance of *Geranium* and Geraniaceae reads. Mock mixture 7 was the only mixture that ITS2 results correctly predicted the taxon richness for with and without the 1% filter. Mixtures 1 to 5 were over-estimated in richness, with sequences assigned to *Calystegia* (family Convolulaceae) and *Geranium* (family Geraniaceae) that were not removed by the 1% abundance filter. The species-level results for ITS2 did not identify *Geranium* to this taxonomic level, but did assign reads to *Calystegia sepium*. Similarly to the RevMet results, a low proportion of *Prunus spinosa* was identified in mock mixture 4, alerting us to the possibility of contamination.

Table 3.3 The species richness of the mock mixtures estimated by the RevMet and ITS2 approaches. The mock mixture taxa belonged to separate families, therefore family, genus and species richness are the same. ITS2 results were calculated at the species-, genus- and family-levels, whereas RevMet produced species-level identification. A 1% minimum abundance filter was applied to the datasets to remove taxa present at low abundances that likely represent false positives assignments.

Mock Mixture Number	Mock Mixture Richness	RevMet		ITS2 - Species		ITS2 - Genus		ITS2 - Family	
		No filter	1% filter	No filter	1% filter	No filter	1% filter	No filter	1% filter
1	4	32	4	5	5	6	6	6	6
2	4	34	4	5	5	6	6	6	6
3	4	33	4	5	5	6	6	6	5
4	3	49	4	4	4	4	4	4	4
5	2	40	2	3	3	4	4	4	4
6	5	48	4	6	6	7	5	7	5
7	5	36	4	5	5	5	5	5	5

1

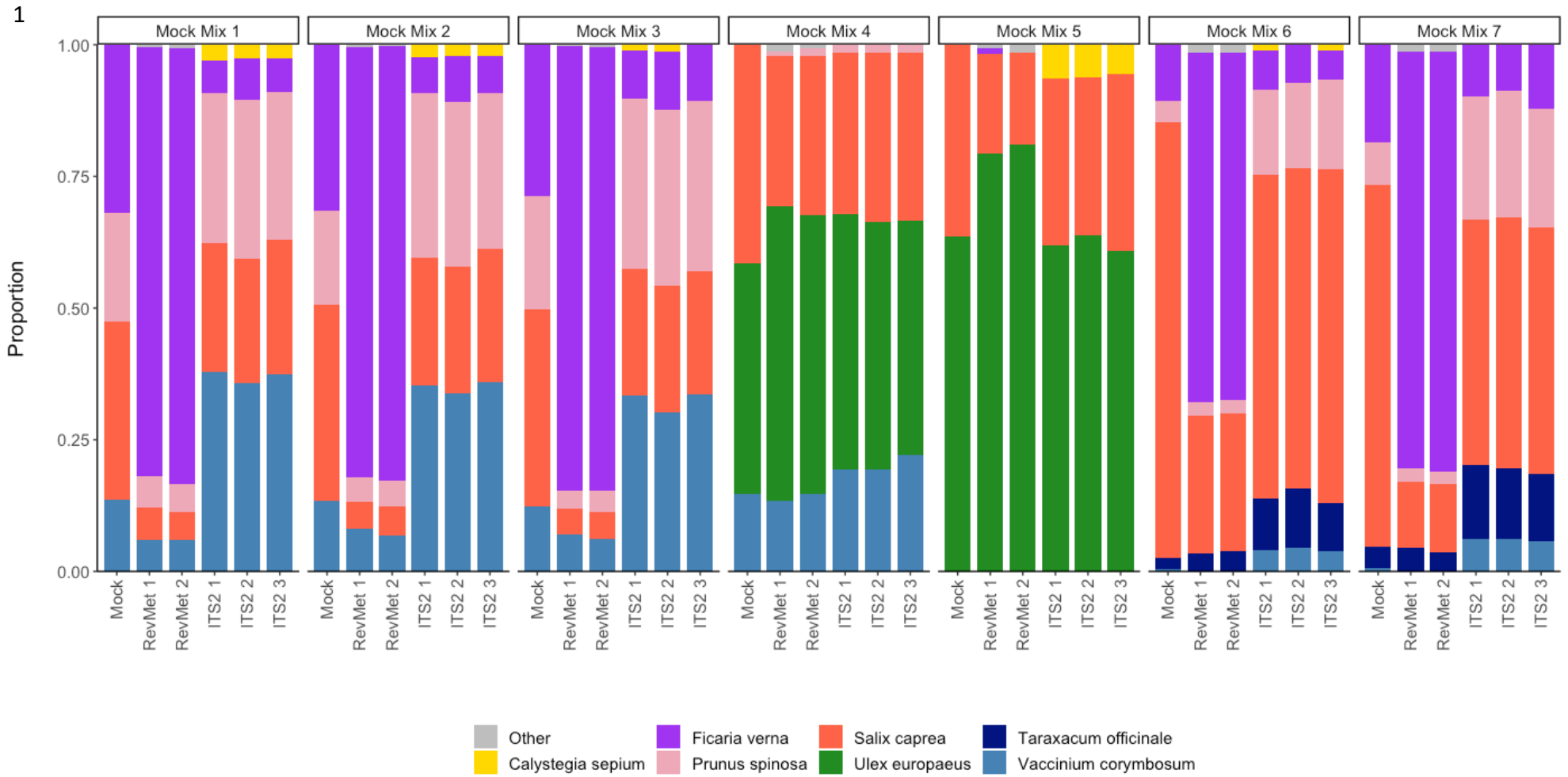


Figure 3.2 Mock mixture quantitative results for RevMet (two replicates) and ITS2 sequences (three replicates). The ITS2 results are shown for species-level identifications. Genus- and family-level assignments are shown in the Supplementary Information (Figures S3.1 and S3.2).

3.4.2 Quantification

When comparing the quantitative abilities of RevMet and ITS2, both produced significant positive relationships between read proportions and pollen grain proportions estimated using microscopy (Figure 3.2). ITS2 results were calculated at the species (slope = 0.59, R^2 = 0.56, p < 0.001), genus (slope = 0.69, R^2 = 0.69, p < 0.001) and family level (slope = 0.69, R^2 = 0.69, p < 0.001). The RevMet results were not normally distributed (Kolmogorov-Smirnov test, d = 0.33, p = 0.003), therefore we measured the relationship using a non-parametric Spearman's Rank correlation (*rho* = 0.66, p < 0.001).

RevMet showed considerable quantitative bias, with certain taxa over- or under-represented in the sequence data (Figure 3.2). RevMet results consistently over-represented *Ficaria verna* and *Ulex europaeus,* while *Salix caprea* and *Prunus spinosa* were under-represented in all of the mock mixtures (Figure 3.3). When comparing the percentage change between microscope estimated counts and RevMet proportions, the two taxa with the largest genomes, *F. verna* (14217 Mbp) and *U. europaeus* (3822 Mbp) also exhibited the highest over-estimation in the RevMet data. The two taxa with the smallest genome sizes, *S. caprea* (392 Mbp) and *P. spinosa* (637 Mbp) were the two taxa with the largest degree of under-estimation. *Vaccinium corymbosum* (651.7 Mbp) and *Taraxacum officinale* (1549 Mbp) exhibited close to a 1:1 relationship and were the median taxa in terms of genome sizes.

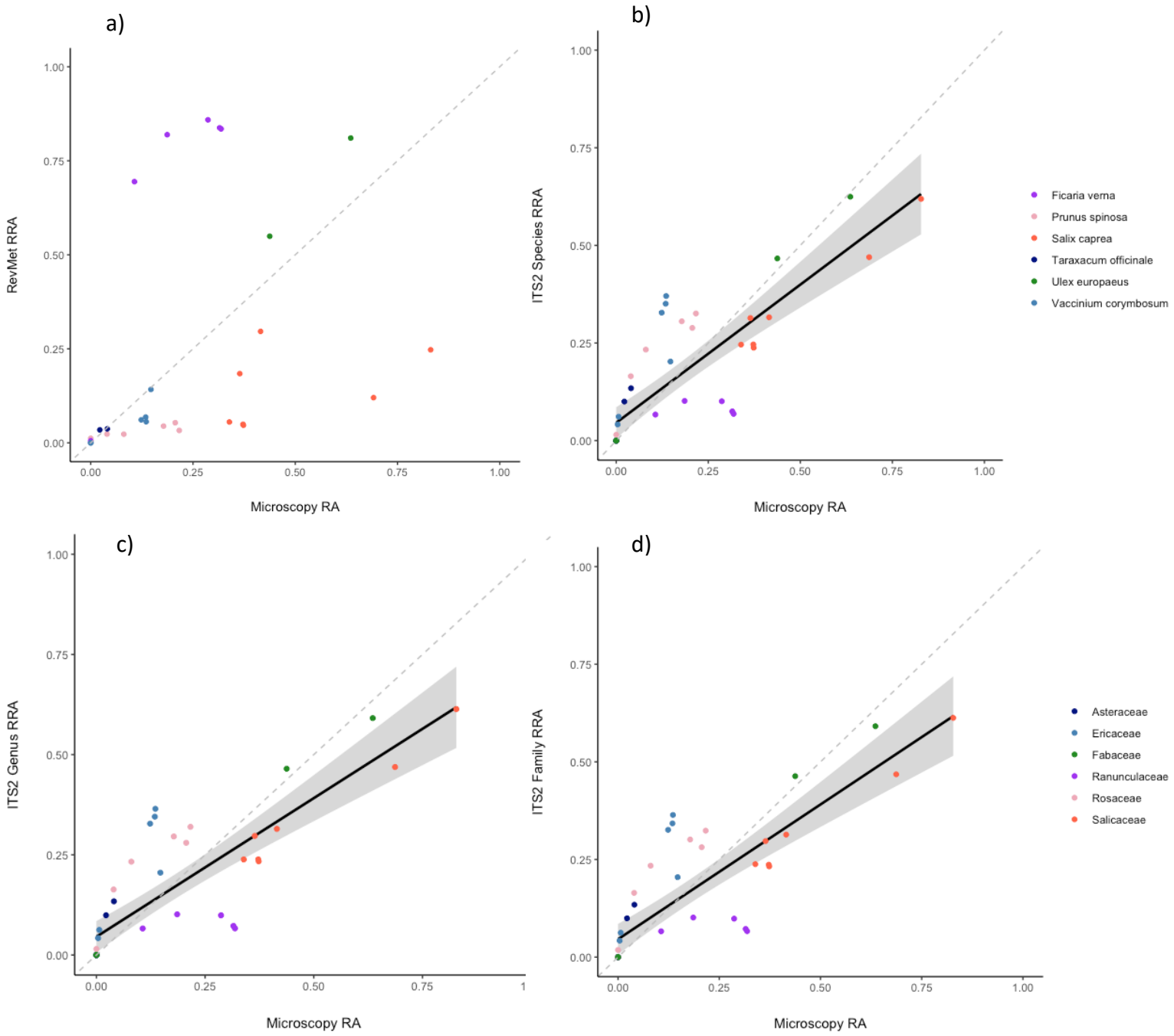


Figure 3.3 The quantitative relationship between mock mixture pollen grain counts and sequence reads for a) RevMet, b) ITS2 species-level, c) ITS2 genus-level, d) ITS2 family-level. The grey dashed line represents a 1:1 relationship and the black lines represents the linear regression, with 95% confidence intervals shown in grey. RevMet residuals were non-normal therefore correlations were calculated using Spearman's Rank correlation coefficient. The same colour is used for each taxon at the different taxonomic levels.

3.4.3 Correction factors

We created a correction factor using the plant genome sizes (genome correction factor; GCF) and applied it to the proportions produced calculated using the RevMet results (Table 3.4; Figure 3.4). We found a statistically significant relationship for the GCF-corrected reads and mock mixture estimated proportions (slope = 0.87, R^2 = 0.66, $p < 0.001$; Figure 3.5). The correction factor improved the 1:1 relationship to mock mixture proportions of the RevMet results and resulted in a stronger relationship than ITS2 (Figure 3.6).

When each taxon's relationship to the mock mixture pollen abundances were calculated separately, the taxa that were over-represented (*Ficaria verna* and *Ulex europaeus*) or under-represented (*Salix caprea* and *Prunus spinosa*) in the RevMet sequences were improved in their quantitative relationship to pollen mock mixture proportions. Species are shown on separate plots in Figure 3.7, to enable easy distinction between sequencing and calculation methods although they were all analysed together in the linear model (Figure 3.6).

Taxon	Mock Mixture %	RevMet %	Percentage change	Genome size (Mbp)	GCF
Vaccinium	13.5	10.0	-3.49	651.7	0.031
Salix	*48.3*	27.4	-20.9	392.0	0.019
Prunus	*14.4*	7.26	-7.16	637.0	0.030
Ficaria	*24.3*	54.7	30.4	14217.0	0.678
Taraxacum	*3.12*	3.25	0.12	1549.2	0.059
Ulex	*53.7*	60.4	6.78	3822.0	0.182

Table 3.4 The difference in percentages of each taxon estimated by the Mock Mixtures and RevMet, the plant genome sizes and the genome correction factor for each taxa.

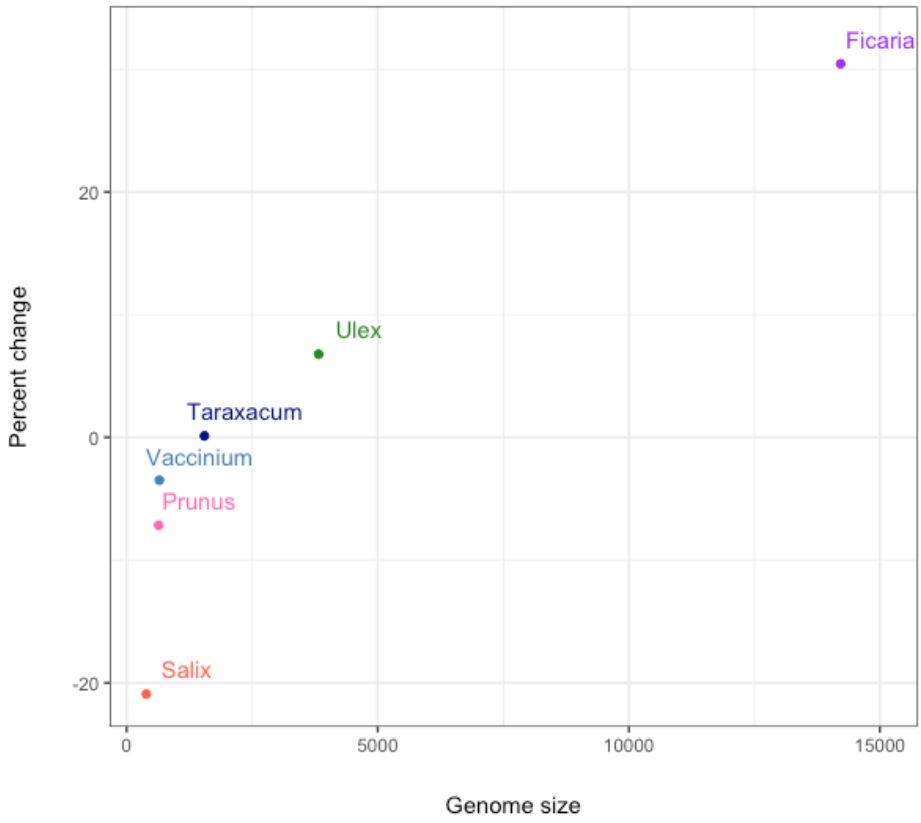


Figure 3.4 The relationship between the percentage change (calculated as the difference between RevMet and mock mixture-estimated proportions) and the genome sizes of the taxa included in the mock mixtures.

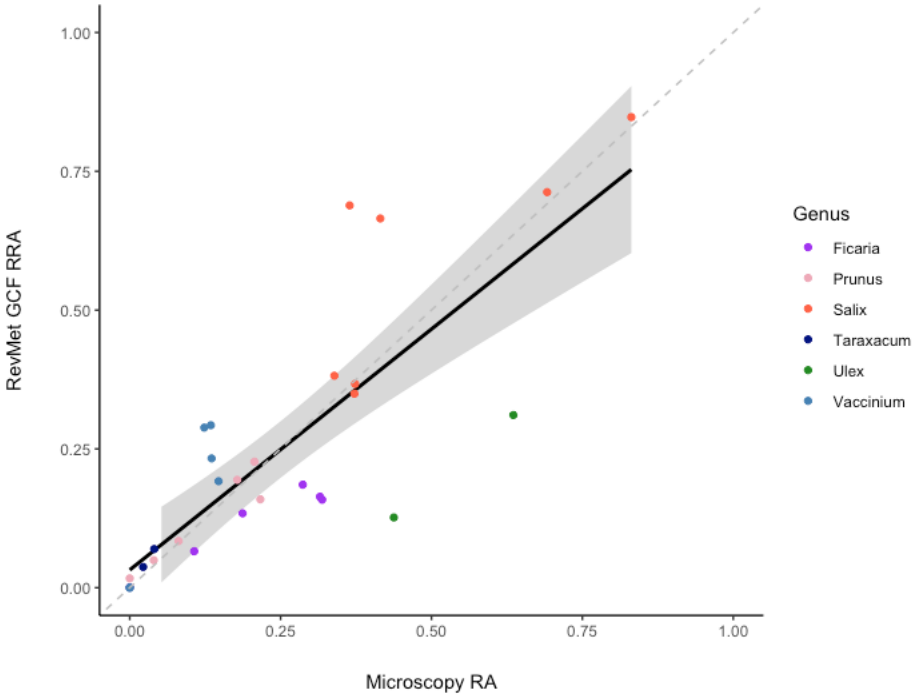


Figure 3.5 The relationship between the mock mixture proportions and the GCF-adjusted RevMet proportions. The linear regression and 95% confidence intervals are displayed, as well as the 1:1 relationship represented by a grey dashed line.

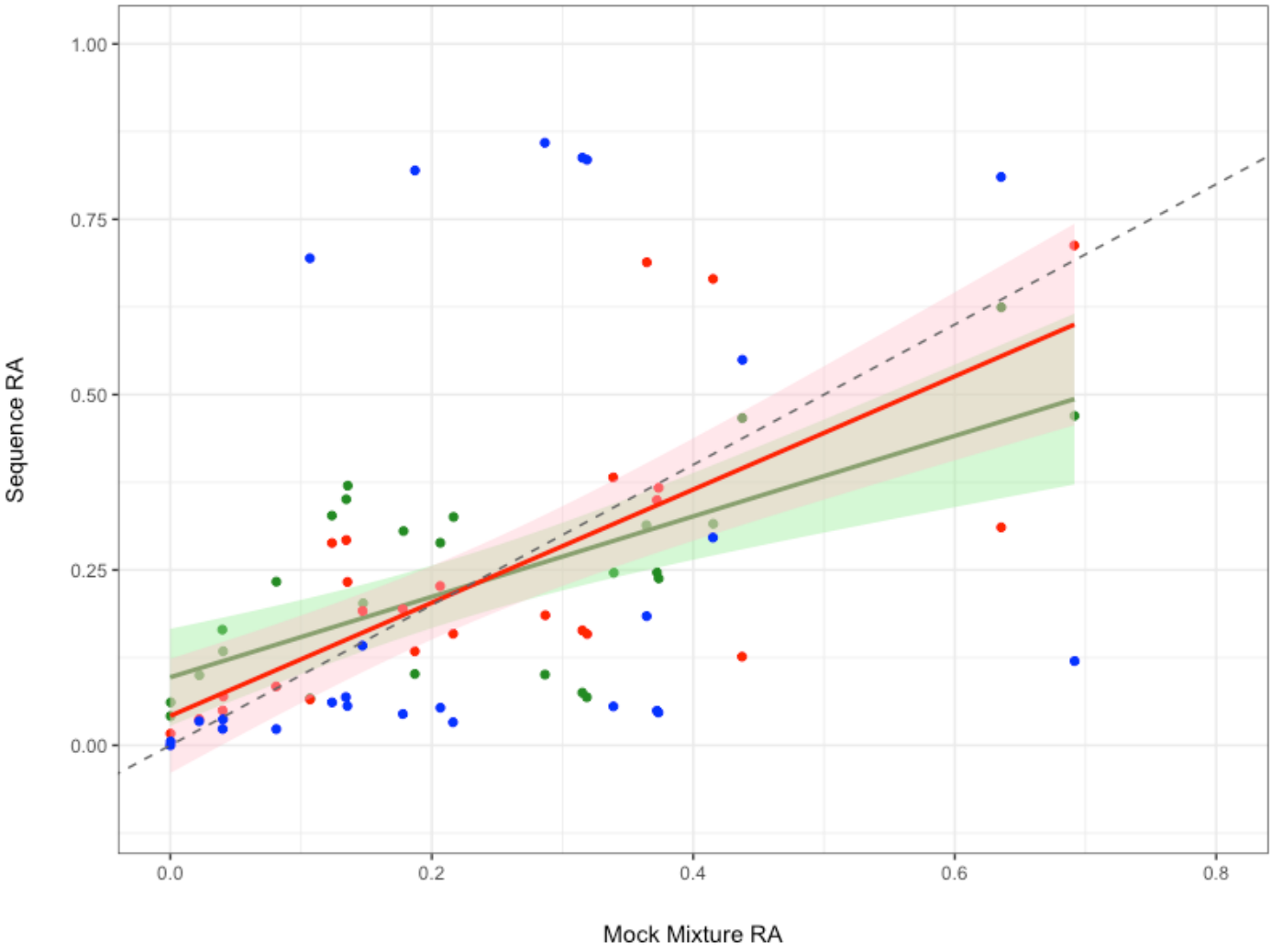


Figure 3.6 The relationship between the mock mixture proportions and sequence proportions for ITS2 (green), RevMet (blue), and the genome correction factor-adjusted RevMet results (red). ITS2 (R^2 = 0.56, p < 0.001) and GCF (R^2 = 0.66, p < 0.001) results showed strong positive relationships with the mock mixture proportions, with linear regressions and 95% confidence intervals shown above. The RevMet residuals were non-normal and so could not be used in a linear model, but showed a positive correlation using Spearman's Rank correlation coefficient (*rho* = 0.66, p < 0.001). The grey dashed line represents a 1:1 relationship.

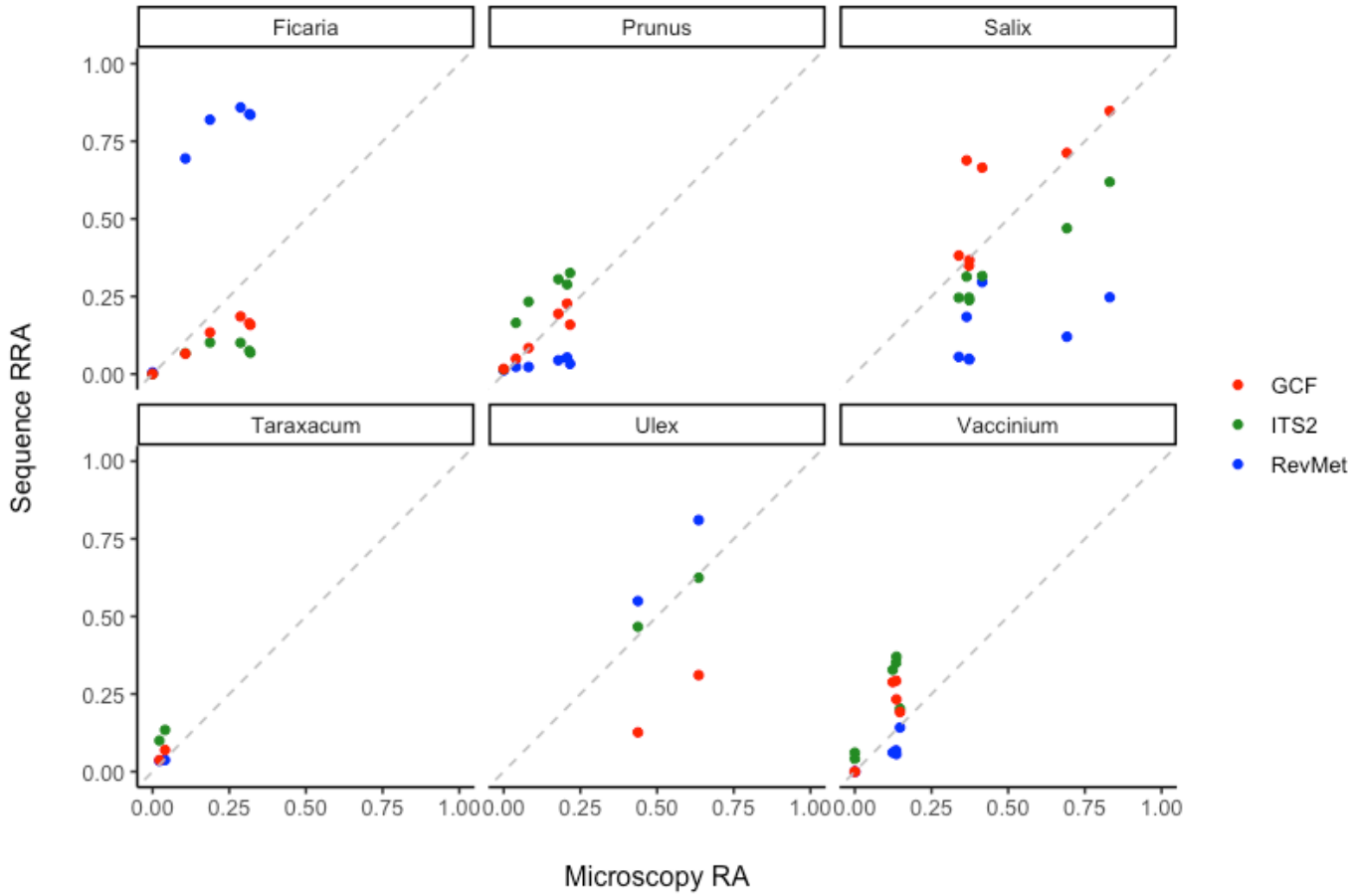


Figure 3.7 The relationship between the sequencing relative abundances and mock mixture abundances for each of the mock mixture taxa. The three methods are shown: ITS2, RevMet and the RevMet results that have been adjusted using a genome correction factor (GCF). The purpose of this figure is to demonstrate the taxon-specific differences between the three methods. For details on the specific linear regressions, see Figure 3.6

3.5 Discussion

In this study, we found a newly developed WGS approach, RevMet, was better able to identify plant species present in mixed pollen samples, with a lower rate of false positive identification than the metabarcoding approach using the ITS2 barcode. However, this was only true when a 1% minimum abundance filter was applied to the RevMet reads, which also removed plant taxa present (true positives) that were present at < 1% in the samples. Metabarcoding with ITS2 demonstrated a stronger positive relationship between the proportions of reads allocated to each plant taxon and the relative abundances of pollen grains counted in the mock mixes than RevMet. Examination of the patterns suggested this was likely due to the large variation in genome sizes in the plant taxa. We applied a correction factor to the RevMet results, derived from the plant genome sizes, and this substantially improved the quantitative relationship, so the relative abundances of pollen grains explained more of the variance in relative abundances of sequence reads than was the case for ITS2.

3.5.1 Detecting species presence

Both RevMet and ITS2 sequencing correctly identified all taxa present in the mock mixtures, with zero false negative detections at all taxonomic levels. Previous studies have found generic or familiar taxonomic assignments to be better than species-level, due to gaps in barcode reference libraries and the similarity in sequences of closely related taxa producing false positive assignments (Bänsch et al., 2020; Bell et al., 2019; Leidenfrost et al., 2020; Richardson, Lin, Sponsler, et al., 2015). One possible explanation for our high rate of assignment at the species level is the reduced genome reference library used in this study. In order to make the RevMet and ITS2 analyses comparable, the database used for the ITS2 sequences only contained species present in the RevMet database. In ecological studies a much larger reference dataset is likely to be used, containing a wider breadth of taxa which include a higher number of conspecifics. To further test this theory, we created an ITS2 database containing a near-complete list of flowering plant species in the UK (Jones, Twyford, et al., 2021). The results for the ITS2 mock mixtures contained a lower number of

assignments at the species level, with only two correctly identified taxa (Supplementary Table 3.6, Supplementary Table 3.7).

MARTi uses a mixture of BLAST and lowest common ancestor (LCA) to identify sequences with the highest similarity score from taxa uploaded on a global reference database. If there are matches to multiple species within a genus, the LCA approach assigns the read to genus, and similarly if there are multiple genus matches, the sequence will be assigned to the common family. The low species-level identification for the ITS2 barcode marker when used in a larger database could be due to sequence similarities in the barcode regions for the targeted species, which would have created multiple species hits for a sequence read, and therefore classified a genus-level identification. With few publicly available WGS reference genomes, we couldn't test the effect of a larger reference database on the RevMet analysis. Peel et al. (2019) found that conspecifics could be correctly identified using RevMet, and this is likely because genome skim references provide a higher coverage of the variable loci and therefore ability to discriminate between species.

Few pollen studies have used nanopore technology for sequencing amplicons, but there is evidence of higher error rates in the basecalling step (Delahaye & Nicolas, 2021). The MinION error rate is always improving, but rates of around 6% were reported in recent literature, whilst the Illumina MiSeq error rate is less than 0.2%. Leidenfrost (2020) sequenced ITS2 amplicons using the MinION and Illumina platform, and although Illumina sequences were of a higher quality and had a higher rate of assignment, the taxonomic results using the two approaches were comparable. Similar results were found when comparing assignments of several taxonomic groups using nanopore, Illumina and PacBio sequencers and it was concluded that taxonomic group (e.g. plant, fungi, bacteria) was more likely to affect the classification accuracy than the sequencing method or classifier (e.g. BLAST) (Pearman et al., 2020).

We also tested the false positive rates of the methods, identifying a number of taxa that were not present in the mock mixes. There were 14 species identified in the RevMet samples at <1% abundance that were not present in the mock mixtures, which supports a recommendation of applying a 1% minimum abundance threshold to RevMet results. The

higher number of false positives generated when there was no filter applied to RevMet could be partially due to the WGS approach spanning regions of the genome with little inter-specific variability, therefore the likelihood of providing an incorrect alignment is higher. It is also possible that RevMet was able to detect a low abundance of many plant taxa, which would result in the 1% abundance filter removing a large number of true positives. The positive identification of plant taxa could have arisen from contamination in the field or lab, or represent small numbers of pollen grains that were transferred to the pollen load from the body of the bee or were deposited on a flower by another insect. Further investigation could test whether a high number of plant taxa present in low numbers is an artifact of the RevMet alignment, or true positives of pollen grains that cannot be identified using microscopy due to low abundance. Using pollen from flower anthers that have dehisced in a sterile environment to make mock mixtures would reduce the possibility of contaminant pollen grains being present in the samples (Lobaton et al., 2021). The 1% abundance filter would be improved by calculating a minimum abundance threshold modelled on sequencing results derived from mock communities and negative controls (Drake et al., 2021).

RevMet was able to correctly detect low proportions of *Vaccinium corymbosum* at less than 1% in two of the mock mixtures. Applying a 1% abundance filter would remove this detection, so care must be taken when removing taxa present at low abundances because rare species that are present in the mix could be removed (Table 3.1). Filtering is a necessary step, otherwise the sample species richness and therefore generalism of the pollinator diet is overestimated, but filtering strategies are yet to be standardised (Tommasi et al., 2021). This is not of concern when analysing the pollen loads of individual bumblebees because the pollen richness is relatively low, therefore even with a minimum abundance filter the range of taxa are likely to be identified. If RevMet is used for applications that look at a much higher richness of taxa, such as honey bee diets using pollen traps or honey analysis, this could be a limitation that requires further investigation.

Other studies using WGS have also found high levels of species detection and low false positives rates, agreeing with our results here (Bell et al., 2021; Lang et al., 2019; Peel et al., 2019). Bell et al. (2021) identified several false positives occurring at <1% in the WGS data.

Lang et al. (2019) and Peel et al. (2019) implemented a 1% minimum sequence filter, so we can't assess the number of false positives identified that were below this threshold. but if they occurred at <1% we can assume it was at similarly low levels to those found in our study.

In the ITS2 results, the genera *Calystegia* and *Geranium* were identified in five of the mock mixtures and were not below 1% abundance, therefore would not be affected by a filtering step. The presence of *Calystegia* and *Geranium* could be either a misidentification by the MARTi alignment process, or a contamination when samples were processed. We did not find these genera in the negative controls or the RevMet results, but it is possible that there were very low levels of *Calystegia* or *Geranium* pollen grains in the mock mixtures that were amplified at a higher rate than the other taxa, and therefore present in the results.

Both RevMet and ITS2 revealed low levels of *Prunus spinosa* in mock mixture 4. It's possible that *P. spinosa* pollen grains were not identified in the creation of the mock mixtures and low levels were identified by both sequencing techniques. The detection of rare species supports evidence of metabarcoding and molecular approaches being more sensitive to rare taxa than microscopic techniques (Keller et al., 2015; Smart et al., 2017).

3.5.2 Quantifying relative abundances

If a sequencing approach were accurate in estimating the relative abundance of each taxon in a mixture, we would expect to see a 1:1 relationship with the relative abundance of pollen grains. We predicted that the WGS approach, RevMet, would be more accurate quantitatively, but ITS2 predicted a stronger quantitative relationship at all taxonomic levels. ITS2 is commonly used in pollen metabarcoding, although multiple studies have presented a weak correlation when ITS2 relative abundance is compared to the microscopic relative abundance (Bell et al., 2019; Keller et al., 2015; Richardson, Lin, Sponsler, et al., 2015b; Smart et al., 2017) or no relationship at all (Richardson et al., 2019b). In our study the ITS2 marker provided a strong positive correlation, which has been demonstrated to a lesser

degree elsewhere (Bänsch et al., 2020; Polling et al., 2022; Richardson et al., 2021). The studies that have demonstrated a strong relationship have been taxon-dependent (Bänsch et al., 2020), used on a relatively small number of taxa (Polling et al., 2022) or at a low level of taxonomic discrimination (Richardson et al., 2021). The relatively small number of taxa used in our mock mixtures and reference database could explain the improved quantitative abilities of ITS2 and should be further tested with a larger number of samples and range of taxa. The positive relationship found in our study adds to the evidence of the quantitative abilities of ITS2 sequencing, although we recommend further investigation due to the contrasting evidence shown in the literature.

Our WGS approach, RevMet, produced a positive relationship between pollen grain proportions and sequence read relative abundance, although the variance could not be estimated due to the non-normal residuals of the data. The relationship for RevMet was less strong than ITS2, providing evidence of pitfalls relating to whole genome sequencing. The taxa used for the mock mixes had a large range of genome sizes, which likely contributed to the deviations from a 1:1 relationship with the microscopic data. *F. verna* had the largest genome size and was consistently over-represented in the quantitative analyses, while *S. caprea* had the smallest genome and was under-represented, leading us to believe genome size could be a contributing factor to quantitative bias in the RevMet approach. Compositional data suffers from the problem of non-independent values; as one taxon increases in proportion, by definition the others must decrease (Gloor et al., 2017). There are statistical approaches that can be adopted to overcome this problem, such as using an offset, which includes the total number of reads for each sample in order to control for differences in sequencing depths (Pendegraft et al., 2019). The total number of sequencing reads has no ecological inference, i.e. it is not comparable to count data, but the use of an offset accounts for the variability in library size.

There are few studies that have used WGS to quantify pollen abundance, but there is evidence of an improved quantitative relationship (Bell et al., 2021; Lang et al., 2019; Peel et al., 2019). Peel et al. (2019) applied the RevMet approach to mock mixes of extracted DNA and achieved improved accuracy in comparison to our results. The plant taxa used in Peel et al. (2019) had genomes ranging from 465 to 14,915 Mbp, but using extracted DNA to create

the mock mixes rather than pollen grains removed the source of the bias originating from genome size. There may also be an effect of bicellular or tricellular pollen on the amount of DNA released by the pollen grain, which could have an effect on the RevMet results independent of plant genome size. Around 60% of angiosperm pollen is two-celled when released from the anther (one vegetative cell and one generative cell), and the remaining 40% are three-celled (one vegetative cell and two generative cells) (Thomas et al., 2003). The difference in cell number is likely to affect the DNA content per pollen grain in different species, which should be further investigated as a potential impact on the accuracy of the RevMet analysis.

3.5.3 Correction Factors

Correction factors have been applied to improve the quantitative accuracy of metabarcoding sequencing data. Species-specific correction factors were created from 50/50 mock mixtures of target organisms in dietary analyses using 16S mtDNA metabarcoding, and once applied they significantly improved the quantitative relationship between sequence proportions and biomass proportions (Thomas et al., 2016). Previous studies have also looked at correcting biases caused by gene copy number variation in microbial communities (Kembel et al., 2012) diatoms (Vasselon et al., 2018) and macroinvertebrates (Elbrecht et al., 2017).

Metabarcoding studies are more likely to encounter biases derived from gene copy variation than genome size, which is why we only applied a correction factor to the RevMet samples and not to the ITS2 samples. Our correction factor was derived from the genome sizes of the plants used in the study to adjust the RevMet proportions, which resulted in the slope of the relationship moving closer to 1, and more of the variance in relative abundance of reads being explained by the relative abundance of the pollen grains. The genome correction factor (GCF) improved RevMet's quantitative relationship from points highly dispersed around the 1:1 line, to a slope of 0.87 and R^2 value of 66%.

The GCF was calculated from the published genome sizes of the plant taxa present in the mock mixtures (Pellicer & Leitch, 2020). The correction factor was successful in reducing the over-estimation displayed by *Ficaria* and also produced a more realistic estimate for *Salix.*

The advantage of GCF is that the calculations can be achieved without prior sequencing of mock mixtures to create a "standard" (Krehenwinkel et al., 2017). The ability of RevMet to detect major species is very good, so once taxa have been identified, their genomes can be searched for in published databases and the GCF applied to the results. This effect has only been found in a small selection of taxa, so it would be interesting to further test this correction factor with a wider range of genome sizes.

Whilst taxon-specific biases can be corrected in a small sample of pollen, we are not addressing the cause of the bias. The genome correction factors produced the highest level of congruence between sequence abundance and pollen grain abundance, but the relationship was still not 1:1. There are likely to be multiple factors influencing the bias in sequencing results, such as differences in DNA extraction between taxa (e.g. different pollen exine wall thickness) and ploidy number. In amplicon sequencing, biases can arise from more possible sources than RevMet, because the WGS approach removes the requirement for PCR-amplification.

3.5.4 Sources of quantification bias

Plant genome sizes can vary hugely between plant taxa, and the taxa included in our mock mixtures ranged from 392 Mbp to 14217 Mbp. This difference was reflected in our results, as the plant taxon with the largest genome, *Ficaria verna*, was overestimated in our RevMet analyses. The DNA content of *Ficaria* pollen grains is much greater than the other taxa used in the mock mixtures, which could explain why the relative read abundance did not reflect the pollen grain proportions. When the percentage change in abundance between RevMet read proportion and microscopic proportions for the six plant taxa was calculated, there was a correlation between over-estimation of RRA and genome size. *Salix* has the smallest genome and was underestimated by the largest degree.

Another source of potential bias lies in the physical characteristics of the pollen grains. Pollen grains from different taxa have variable pollen exines and volumes, which contributes to interspecific differences in DNA extraction efficiency (Swenson & Gemeinholzer, 2021).

There is a trade-off between the yield of DNA that can be extracted and the length of DNA fragments, because a longer mechanical homogenisation step will break the cell walls to a higher degree, but that will simultaneously break DNA fragments into smaller pieces.

3.5.5 Suitability of bee-collected pollen

In this study, we investigated the qualitative and quantitative accuracy of two molecular approaches to characterising mixed species pollen loads collected by foraging bumblebees. Several previous studies have used constructed species pollen mixes to assess the accuracy of the chosen molecular method, but all have used cultivated ornamental or wind-pollinated plants as sources of pollen (Baksay et al., 2020, 2022; Bell et al., 2019, 2021; Lang et al., 2019). There are advantages to using these pollen sources, such as a higher quantity of pollen grains with which to create the mixtures. However, these species are often not representative of the foraging resources used by pollinators. In this study we used pollen originating from corbicular loads collected from foraging bees, providing a more realistic representation of the ecological context in which genetic techniques can be used to identify pollen grains to investigate plant-pollinator interactions. Small amounts of contamination from other pollen sources are possible, for example when a bee visits a flower for nectar small quantities of pollen grains might be picked up, but this could serve as interesting additional information. There are also non-plant sources of DNA that are likely to be included in the pollen loads, such as microbial or bee DNA, which have an unknown effect on the sequencing process.

3.5.6 Suitability of nanopore sequencing for pollen analysis

Here, we demonstrate the suitability of nanopore technology for pollen taxa identification using whole genome or amplicon sequencing. In comparison to Illumina sequencing, the per sample cost is reduced when using a MinION, particularly if there are a large number of samples that can be multiplexed (van der Reis et al., 2023). The newly released Flongle cells make sequencing cheaper still, as they yield similar results with fewer nanopores but at a

fraction of the cost. At a larger scale, there is ONT PromethION, a platform that can sequence up to 48 flow cells at a time, utilise a greater number of pores, and ultimately increase the sequencing output and reduce the per sample cost (Kim et al., 2019). The portability of MinION sequencers is an advantage over other sequencing technologies, and sequencing has the potential to be carried out without the requirement for a laboratory.

There are disadvantages to using nanopore sequencing and the RevMet approach. First, nanopore sequencing has a higher error rate than other methods, estimated at approximately 6%, although continuous improvements in basecalling accuracy are likely to reduce sequencing errors (Zeng et al., 2020). Despite the error rate, there is little evidence to suggest it has a negative impact on taxonomic assignment in comparison to Illumina data (van der Reis, et al., 2023, Leidenfrost et al., 2020). Here, we used high-accuracy basecalling to remove reads of low quality ($Q < 9$) from both ITS2 and RevMet datasets.

Second, there is currently a much higher availability of barcode reference databases than full genome sequences, which is the current limitation to WGS approaches. There is a database that covers > 97% of UK flowering plants for at least one metabarcoding marker of ITS2, *rbcL* or *matK* (Jones, Twyford, et al., 2021). This is not true for all countries, and the costs associated with creating reference libraries is a disadvantage, although this technology is getting steadily cheaper. We used low-cost, low-coverage genome skims generated for Chapters Two and Four to identify our RevMet long read nanopore sequences. Genome skims were targeted at 1x coverage, in order to provide enough information for species-level identification. WGS using full reference genomes has demonstrated a strong quantitative ability in multi-species pollen loads, but there are currently few plant species represented in these databases to be fully utilised (Bell et al., 2021). Ongoing projects aim to provide assembled reference genomes for all eukaryotes, which will make WGS approaches more feasible (Lewin et al., 2018).

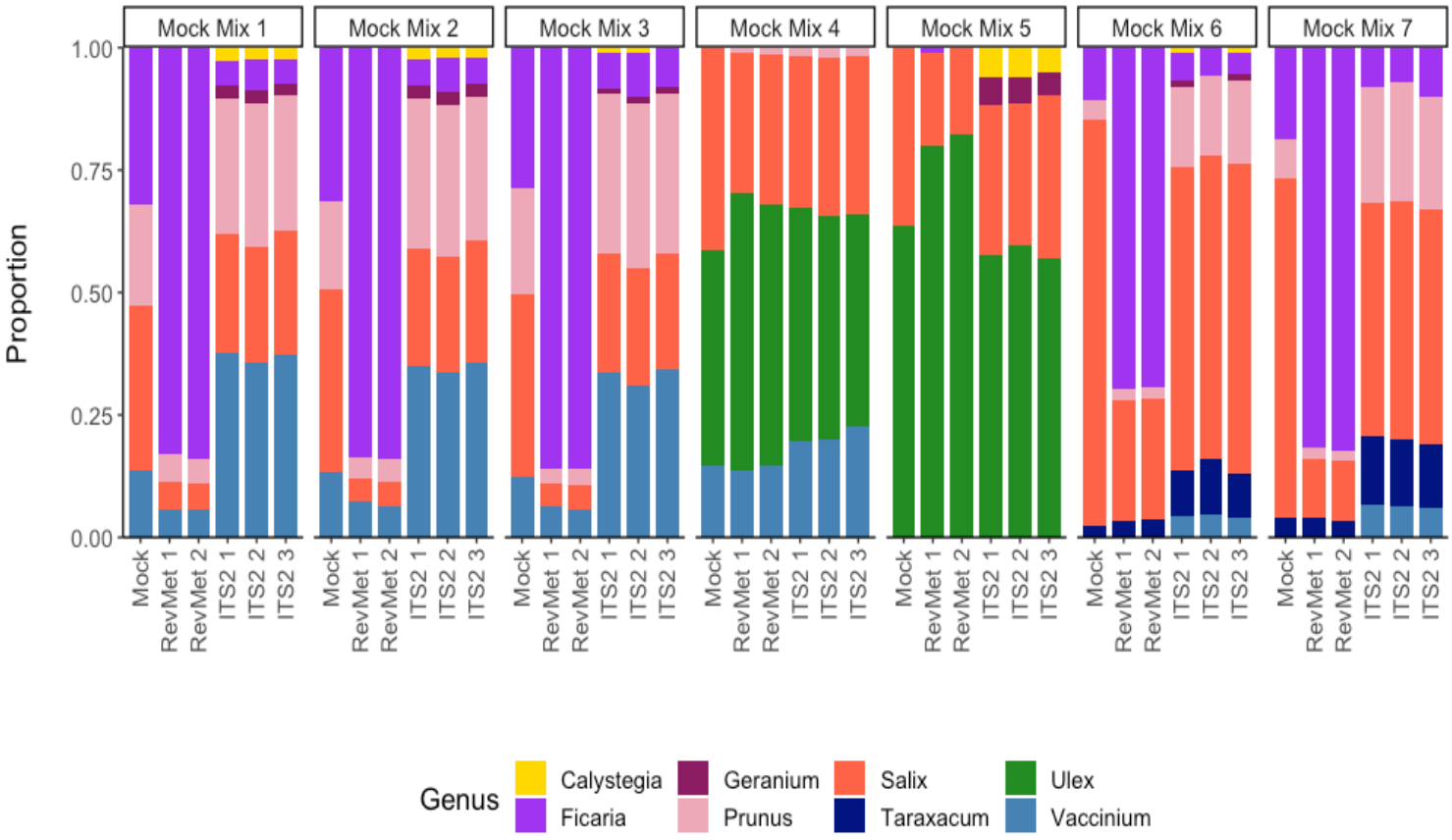
3.6 Conclusions

RevMet shows promise in detecting pollen taxa to a high level of taxonomic discrimination. We found that quantitatively, plant taxa with large genomes could be over-represented and

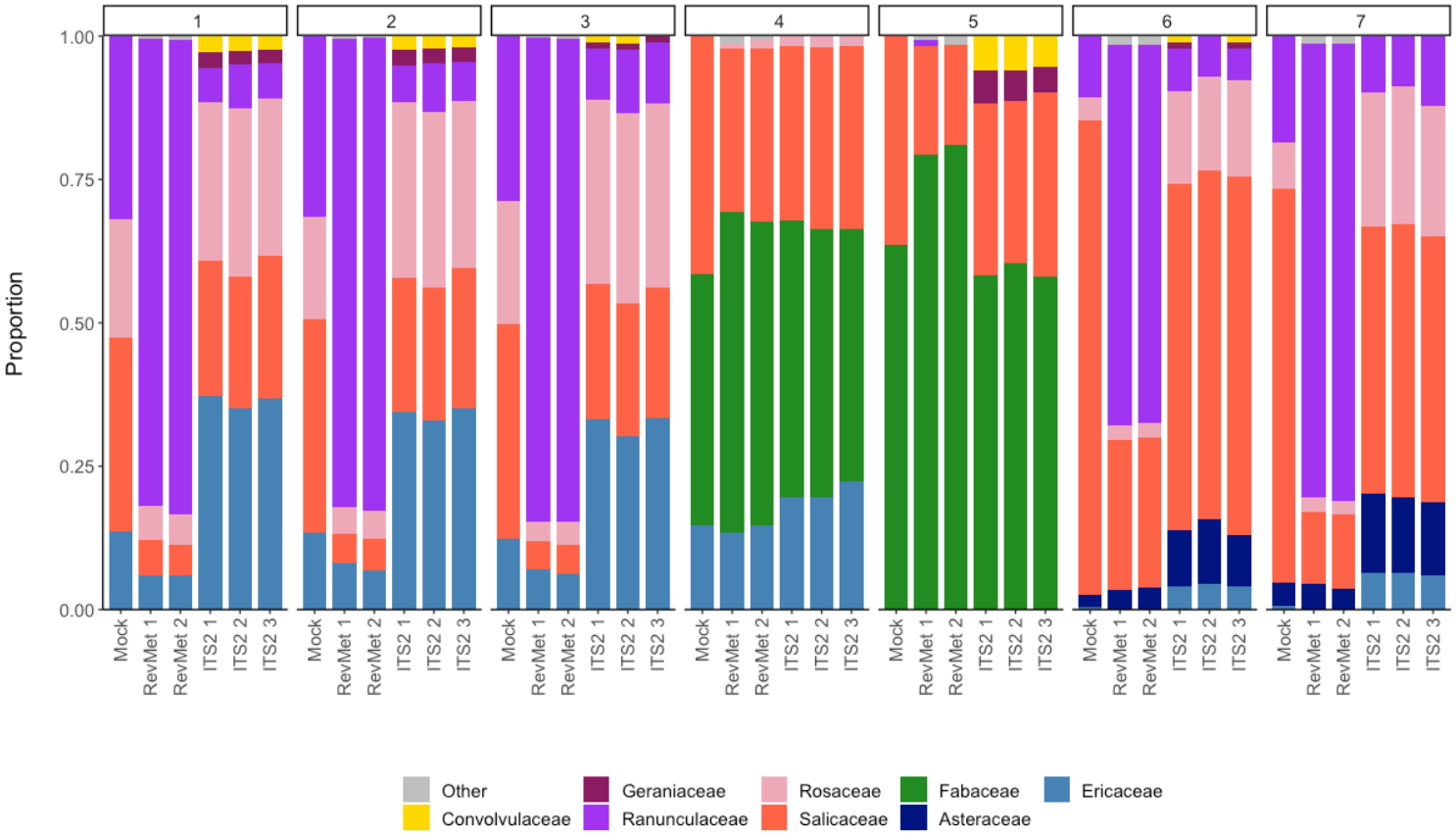
plant taxa with small genomes under-represented in comparison to relative abundances estimated using microscopy. We suggest using a genome correction factor to mitigate against this bias, which improved the quantitative abilities of RevMet in our study but should be further explored with a larger range of taxa and samples. Of the two approaches, ITS2 exhibited the strongest quantitative relationship with the mock mixture proportions. This result is somewhat out of place with the literature where ITS2 amplicons have shown inconsistent quantitative abilities, and we think this is in part due to the reduced reference database used here so we recommend further investigation. We suggest that nanopore sequencing, with its portability, low cost and increasing accuracy, presents an attractive alternative to Illumina sequencing for characterising and quantifying mixed species pollen loads.

3.7 Supplementary Information

Supplementary Figure 3.1 Mock mixture quantitative results for RevMet (two replicates) and ITS2 sequences (three replicates). The ITS2 results are shown for genus-level identifications.



Supplementary Figure 3.2 Mock mixture quantitative results for RevMet (two replicates) and ITS2 sequences (three replicates). The ITS2 results are shown for family-level identifications.



Supplementary Table 3.1 Sources of pollen for mock mixtures

Farm	Date collected	Sample period	Sample number	Barcode	Unassigned	Total read number	Assigned	Assigned (%)	Focal Taxa
Winterwood	19/04/2021	2	198	barcode06	783	4952	4169	84.2	Prunus spinosa
Winterwood	09/04/2021	1	203	barcode11	628	3563	2935	82.4	Prunus spinosa
Colworth	01/04/2021	1	937	barcode73	925	4899	3974	81.1	Prunus spinosa
Tuesley	22/04/2021	2	553	barcode73	400	4954	4554	91.9	Ulex europaeus
Heathlands	27/04/2021	3	54	Barcode54	463	4987	4524	90.7	Ulex europaeus
Colworth	01/04/2021	1	874	barcode10	567	2738	2171	79.3	Prunus spinosa
Colworth	18/04/2021	2	685	barcode13	450	4997	4547	91.0	Ficaria verna
Colworth	01/04/2021	1	811	barcode43	468	4998	4530	90.6	Ficaria verna
Winterwood	09/04/2021	1	219	barcode27	482	4998	4516	90.4	Ficaria verna
Colworth	01/04/2021	1	852	barcode84	603	3911	3308	84.6	Prunus spinosa
Colworth	18/04/2021	2	882	barcode18	933	4812	3879	80.6	Taraxacum officinale
Winterwood	19/04/2021	2	245	barcode53	975	4975	4000	80.4	Prunus spinosa
Heathlands	27/04/2021	3	55	barcode55	1188	4986	3798	76.2	Vaccinium corymbosum
Colworth	18/04/2021	2	763	barcode91	1275	4976	3701	74.4	Vaccinium corymbosum
Colworth	01/04/2021	1	913	barcode49	1306	4982	3676	73.8	Vaccinium corymbosum
Colworth	18/04/2021	2	740	barcode68	1341	4985	3644	73.1	Vaccinium corymbosum
Colworth	18/04/2021	2	677	barcode5	1347	4972	3625	72.9	Vaccinium corymbosum
Heathlands	11/05/2021	4	295	barcode07	1373	4973	3600	72.4	Vaccinium corymbosum

Supplementary Table 3.2 The estimated pollen grain number, intended proportions and actual proportions of the pollen mock mixtures. The taxon included are *Ficaria verna* (F), *Prunus spinosa* (P*), Salix caprea* (S), *Taraxacum officinale* (T), *Ulex europaeus* (U) and *Vaccinium corymbosum* (V). The DNA yields are shown for each mock mixture, measured after the DNA extraction step.

Mock Mixture	Taxon ratio	Intended ratio	Observed ratio (approximate)	DNA concentration (ng/μl)
1	F:P:S:V	1:1:1:1	3:2:3:1	25.9
2	F:P:S:V	1:1:1:1	4:2:5:2	11.4
3	F:P:S:V	1:1:1:1	3:2:4:1	21.1
4	S:V:U	1:2:2	3:3:1	7.65
5	S:U	1:4	1:2	0.1
6	F:P:S:T:V	2:1:2:1:1	10:4:80:2:1	19.7
7	F:P:S:T:V	2:1:2:1:1	20:10:70:5:1	24.5

Supplementary Table 3.3 The primer sequences used to amplify the ITS2 barcode region.

Primer name	Forward/Reverse	Sequence
ITS2F	Forward	ATGCGATACTTGGTGTGAAT
ITS3R	Reverse	GACGCTTCTCCAGACTACAAT

Supplementary Table 3.4 MARTi reads for the negative controls for ITS2 sequencing.

Taxon	NCBI ID	NCBI Rank	Negative control 1	Negative control 2
Asteraceae	4210	family	0	0
Geranium	4028	genus	0	4
Ulex europaeus	3902	species	6	7
Glechoma hederacea	28509	species	0	0
Silene dioica	39879	species	0	0
Calystegia sepium	47519	species	1	3
Taraxacum officinale	50225	species	1	0
Vaccinium corymbosum	69266	species	7	4
Ficaria verna	79245	species	1	6
Prunus spinosa	114937	species	4	19
Salix caprea	172267	species	18	51
Viola riviniana	214052	species	0	0
Total number of pass reads:			38	94

Supplementary Table 3.5 Qualitative accuracy in constructed mock mixtures for the RevMet and ITS2 results. Each value is out of a total of two (RevMet) or three (ITS2) replicate mocks

Mock mixture	Taxon	RevMet			ITS2		
		Family	Genus	Species	Family	Genus	Species
1	Ficaria verna	2/2	2/2	2/2	3/3	3/3	3/3
	Prunus spinosa	2/2	2/2	2/2	3/3	3/3	3/3
	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Vaccinium corymbosum	2/2	2/2	2/2	3/3	3/3	3/3
2	Ficaria verna	2/2	2/2	2/2	3/3	3/3	3/3
	Prunus spinosa	2/2	2/2	2/2	3/3	3/3	3/3
	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Vaccinium corymbosum	2/2	2/2	2/2	3/3	3/3	3/3
3	Ficaria verna	2/2	2/2	2/2	3/3	3/3	3/3
	Prunus spinosa	2/2	2/2	2/2	3/3	3/3	3/3
	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Vaccinium corymbosum	2/2	2/2	2/2	3/3	3/3	3/3
4	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Ulex europaeus	2/2	2/2	2/2	3/3	3/3	3/3
	Vaccinium corymbosum	2/2	2/2	2/2	3/3	3/3	3/3
5	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Ulex europaeus	2/2	2/2	2/2	3/3	3/3	3/3
6	Ficaria verna	2/2	2/2	2/2	3/3	3/3	3/3
	Prunus spinosa	2/2	2/2	2/2	3/3	3/3	3/3
	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Taraxacum officinale	2/2	2/2	2/2	3/3	3/3	3/3
	Vaccinium corymbosum	2/2	2/2	2/2	3/3	3/3	3/3
7	Ficaria verna	2/2	2/2	2/2	3/3	3/3	3/3
	Prunus spinosa	2/2	2/2	2/2	3/3	3/3	3/3
	Salix caprea	2/2	2/2	2/2	3/3	3/3	3/3
	Taraxacum officinale	2/2	2/2	2/2	3/3	3/3	3/3
	Vaccinium corymbosum	2/2	2/2	2/2	3/3	3/3	3/3

Supplementary Table 3.6. The MARTi assignments for the ITS2 reads at species-, genus- and family- taxonomic levels. The ITS2 database contained the same list of plant taxa as the RevMet database.

Taxon	NCBI ID	NCBI Rank	Mock Mix 1.1	Mock Mix 1.2	Mock Mix 1.3	Mock Mix 2.1	Mock Mix 2.2	Mock Mix 2.3	Mock Mix 3.1	Mock Mix 3.2	Mock Mix 3.3	Mock Mix 4.1	Mock Mix 4.2	Mock Mix 4.3	Mock Mix 5.1	Mock Mix 5.2	Mock Mix 5.3	Mock Mix 6.1	Mock Mix 6.2	Mock Mix 6.3	Mock Mix 7.1	Mock Mix 7.2	Mock Mix 7.3
Ranunculaceae	3440	family	2570	3969	3616	4527	4564	4523	8856	5338	7123	0	0	0	0	0	0	4055	3718	3107	6768	5062	5752
Salicaceae	3688	family	10126	11768	14494	16123	12555	16005	23555	11301	15424	13144	15022	14103	15348	15788	8776	33779	31584	35481	31962	27255	22044
Rosaceae	3745	family	11975	15042	16009	21143	16645	19064	32105	16090	21872	763	899	799	0	0	0	9011	8490	9576	16058	13829	10750
Fabaceae	3803	family	0	0	0	0	0	0	0	0	0	20794	22034	19609	29834	33682	15899	0	0	0	0	0	0
Geraniaceae	4027	family	1116	1251	1313	1829	1399	1598	1124	614	774	0	0	0	2874	2936	1266	641	0	647	0	0	0
Convolvulaceae	4118	family	1247	1298	1445	1659	1165	1321	1126	586	0	0	0	0	3112	3354	1447	623	0	625	0	0	0
Asteraceae	4210	family	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5373	5858	5120	9569	7659	6057
Ericaceae	4345	family	16034	18063	21568	23754	17934	22977	33170	14647	22698	8418	9292	9881	0	0	0	2284	2374	2267	4403	3646	2796
Prunus	3754	genus	11894	14914	15908	20611	16188	18616	31491	15824	21489	611	719	684	0	0	0	8952	8410	9504	15931	13718	10681
Ulex	3901	genus	0	0	0	0	0	0	0	0	0	20791	22029	19605	29828	33674	15898	0	0	0	0	0	0
Geranium	4028	genus	1116	1251	1313	1829	1399	1598	1124	614	774	0	0	0	2874	2936	1266	641	0	647	0	0	0
Vaccinium	13749	genus	16034	18063	21568	23754	17934	22977	33170	14647	22698	8418	9292	9881	0	0	0	2284	2374	2267	4403	3646	2796
Salix	40685	genus	10126	11768	14494	16123	12555	16005	23555	11301	15424	13144	15022	14103	15348	15788	8776	33779	31584	35481	31962	27255	22044
Calystegia	47518	genus	1247	1298	1445	1659	1165	1321	1126	586	0	0	0	0	3112	3354	1447	623	0	625	0	0	0
Taraxacum	49743	genus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5353	5844	5104	9534	7630	6042
Ficaria	168009	genus	2568	3967	3615	4514	4555	4516	8848	5336	7121	0	0	0	0	0	0	4055	3716	3106	6767	5061	5752
Ulex europaeus	3902	species	0	0	0	0	0	0	0	0	0	20791	22029	19605	29828	33674	15898	0	0	0	0	0	0
Calystegia sepium	47519	species	1223	1275	1429	1637	1145	1306	1090	575	0	0	0	0	3051	3324	1436	611	0	617	0	0	0
Taraxacum officinale	50225	species	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5353	5844	5104	9534	7630	6042
Vaccinium corymbosum	69266	species	15747	17763	21277	23353	17612	22631	32473	14368	22388	8264	9123	9721	0	0	0	2231	2328	2205	4297	3565	2735
Ficaria verna	79245	species	2568	3967	3615	4514	4555	4516	8848	5336	7121	0	0	0	0	0	0	4055	3716	3106	6767	5061	5752
Prunus spinosa	114937	species	11894	14914	15908	20611	16188	18616	31491	15824	21489	611	719	684	0	0	0	8952	8410	9504	15931	13718	10681
Salix caprea	172267	species	10126	11768	14494	16123	12555	16005	23555	11301	15424	13144	15022	14103	15348	15788	8776	33779	31584	35481	31962	27255	22044
Total number of pass reads:			46418	72649	72825	58843	104450	53529	60541	57313	54516	45224	50345	60641	49608	68566	49488	71022	53789	46451	57967	58525	30576

Supplementary Table 3.7. The MARTi assignments for the ITS2 reads at species-, genus- and family- taxonomic levels. The reference database contained all ITS2 sequences for the UK. The number of species-level assignments decreased with a larger reference database

Taxon	NCBI ID	NCBI Rank	Mock Mix 1.1	Mock Mix 1.2	Mock Mix 1.3	Mock Mix 2.1	Mock Mix 2.2	Mock Mix 2.3	Mock Mix 3.1	Mock Mix 3.2	Mock Mix 3.3	Mock Mix 4.1	Mock Mix 4.2	Mock Mix 4.3	Mock Mix 5.1	Mock Mix 5.2	Mock Mix 5.3	Mock Mix 6.1	Mock Mix 6.2	Mock Mix 6.3	Mock Mix 7.1	Mock Mix 7.2	Mock Mix 7.3
Ranunculaceae	3440	family	0	0	4042	4898	3247	4140	4200	6463	8046	0	5079	6090	3350	4570	2824	3575	0	3664	2314	0	0
Salicaceae	3688	family	8384	12414	15024	10688	13630	15313	11953	14538	22319	14464	20605	30057	30132	25888	33871	11175	14345	32163	9657	13454	15102
Rosaceae	3745	family	0	731	7944	7031	6173	9879	7366	9826	15364	0	4120	6560	4144	6051	4140	5781	879	4194	5613	753	0
Fabaceae	3803	family	15595	20144	0	0	0	0	0	0	0	28911	0	0	0	0	0	0	21483	0	0	19100	32928
Geraniaceae	4027	family	1241	0	1566	596	1270	1792	1358	752	1072	2806	0	609	469	535	639	1205	0	634	1074	0	2876
Convolvulaceae	4118	family	1368	0	1229	561	1342	1547	1112	582	1024	2933	0	623	0	529	590	1209	0	581	1168	0	3193
Asteraceae	4210	family	0	0	0	0	0	0	0	0	0	0	5637	8901	5541	7163	4779	0	0	4914	0	0	0
Ericaceae	4345	family	0	8078	21978	13947	20637	22791	17270	21703	31641	0	2615	4130	2272	3454	2123	17322	8943	2164	15239	9519	0
Prunus	3754	genus	0	725	7802	7014	6158	9707	7226	9803	15327	0	4115	6552	4133	6043	4130	5766	874	4188	5603	752	0
Ulex	3901	genus	15212	19669	0	0	0	0	0	0	0	28210	0	0	0	0	0	0	20956	0	0	18613	32152
Geranium	4028	genus	1239	0	1565	596	1270	1792	1358	752	1072	2806	0	609	469	535	639	1205	0	634	1074	0	2876
Vaccinium	13749	genus	0	1591	4210	2716	3522	4342	3321	4089	6118	0	492	753	0	668	0	3165	1753	0	2794	1749	0
Salix	40685	genus	8383	12412	15018	10683	13627	15311	11947	14530	22308	14461	20601	30048	30126	25878	33865	11170	14342	32155	9652	13448	15100
Calystegia	47518	genus	1368	0	1205	560	1341	1514	1091	580	1023	2930	0	623	0	529	590	1207	0	581	1167	0	3192
Taraxacum	49743	genus	0	0	0	0	0	0	0	0	0	0	5629	8884	5533	7154	4771	0	0	4909	0	0	0
Ficaria	168009	genus	0	0	4020	4886	3229	4118	4177	6436	8023	0	5059	6070	3330	4550	2814	3557	0	3650	2304	0	0
Ficaria verna	79245	species	0	0	4020	4886	3229	4118	4177	6436	8023	0	5059	6070	3330	4550	2814	3557	0	3650	2304	0	0
Geranium pusillum	122182	species	1229	0	1542	593	1263	1771	1330	748	1068	2793	0	606	468	533	637	1199	0	631	1066	0	2861
Taraxacum sp. CF-2016	1844647	species	0	0	0	0	0	0	0	0	0	0	5629	8884	5533	7154	4771	0	0	4909	0	0	0
Total number of pass reads:			30440	44914	68253	50021	60361	72349	56971	70674	103538	53373	49346	72088	54159	60176	58201	53141	49250	58466	46077	46244	57744

Chapter Four

Nanopore sequencing of pollen reveals how commercial bumblebees supplement their diet from the wider landscape

4.1 Abstract

Agricultural intensification has led to a decrease in wild pollinators, partly due to a lack of suitable forage and habitat. As a result, many pollinator-dependent crops now rely on commercial pollinators to fulfil their pollination requirements. Specific recommendations for supplementary floral resources are required that can attract and sustain wild pollinators and improve synergies between provision of pollination services and pollinator conservation in agriculture. Understanding the foraging behaviours of commercial bees may help close this gap. Here, we demonstrate the use of nanopore long read sequencing, RevMet, to identify plant species that are attractive to foraging social bumblebees (*Bombus terrestris*). Specifically, we characterize and quantify the plant taxa in the pollen loads of foraging commercial *B. terrestris* on four UK blueberry (*Vaccinium corymbosum*) farms during the spring crop-flowering period. Our method is PCR-free and has a proven ability to quantify relative abundances of plant taxa in mixed bee-collected pollen loads. Read frequencies were compared to the abundance of flowering taxa in the surrounding landscape to analyse the extent to which bumblebees displayed preference or avoidance of available plant taxa. Less than half of all assigned sequences (45.4%) originated from the *Vaccinium corymbosum* crop. Goat willow (*Salix caprea*) was a major contributor to the pollen diet (30.6 %), followed, to a lesser extent, by blackthorn (*Prunus spinosa*, 5.2%). Pollen diet composition was associated with seasonal change and the flowering taxa available in the surrounding landscape. There was a higher diversity of pollen taxa in pollen loads later in the season, although most plant taxa were used by the bees in relatively small quantities (< 10%) to supplement their dietary needs. When compared to a null model*, B. terrestris* consistently visited *Salix caprea* more than would be expected based on its abundance in the landscape. *Vaccinium corymbosum* was collected less than expected in the early crop flowering period (March) and more than expected towards the end of the flowering (May). Our data can inform hedgerow and field margin plantings and management for bumblebee conservation for bumblebees, and support a clear recommendation to increase densities of *Salix caprea* in UK hedgerows.

4.2 Introduction

Insect pollinators provide essential pollination services to crops globally, with one third of food being dependent on animal pollination to some degree (Klein et al., 2007). There is evidence of wild insect populations declining, primarily due to loss of suitable habitat, increased use of pesticides and disease (Potts et al., 2010). As human populations increase globally, more pressure is placed on agricultural systems to be able to provide enough food. The consequence of this is reduced biodiversity in farm landscapes, which has a negative impact on crop yields, if they are pollinator dependent (Potts et al., 2016).

Traditionally, soft fruit growers have relied on wild insects to carry out the pollination services required to produce marketable fruit. An insufficient number of pollinators in soft fruit crops can lead to a reduction in qualities related to lower weight, abnormal shape and shorter shelf life (Bartomeus et al., 2014; Benjamin & Winfree, 2014; Wietzke et al., 2018). For example, highbush blueberry (*Vaccinium corymbosum* L.; Ericales: Ericaceae) is an economically valuable crop that is self-fertile to a degree but requires some cross-pollination to produce high quality and marketable fruit (Gibbs et al., 2016; Kendall et al., 2020). Wild bees, including bumblebees, can significantly reduce pollination deficit on *Vaccinium corymbosum* farms, indicating a potential yield increase of 33%, should the sufficient level of pollination be present (Benjamin & Winfree, 2014; Button & Elle, 2014; Nicholson et al., 2019).

To mitigate against these losses, pollinators such as honey bees (*Apis mellifera*) and buff-tailed bumblebees (*Bombus terrestris*) are used commercially to supplement pollination services in a wide variety of crops (Gibbs et al., 2016; Isaacs & Kirk, 2010). While honey bees are commonly used to pollinate *Vaccinium corymbosum* crops in North America, the flower morphology actually requires buzz-pollination (Buchmann, 1983; De Luca & Vallejo-Marín, 2013). Honey bees are unable to buzz-pollinate, and are therefore considered less effective pollinators of *Vaccinium corymbosum* in comparison to bumblebees (Estravis-Barcala et al., 2021; Javorek et al., 2002; Rogers et al., 2013). Bumblebees also display higher rates of activity at lower temperatures compared to honey bees, and are therefore more suitable for polytunnel-grown *Vaccinium corymbosum* crops that typically start flowering in March in the

UK, when the average temperature is below the threshold for honey bee activity (Corbet et al., 1993; Tuell et al., 2009).

However, increasing densities of commercial bees may be detrimental to wild pollinators. Commercial bumblebees and honey bees can cause displacement of wild bees, increase competition for resources, and facilitate the spread of disease (Furst et al., 2014; Mallinger et al., 2017). The decline in wild pollinators coupled with increasing production of pollinator-dependent crops has created a necessity in using managed bees, which could cause further wild pollinator losses. Therefore, the monitoring of crop visitors is an important step in providing information on the services provisioned by different pollinator groups (Kleijn et al., 2015).

The efficacy of pollination provided by both commercial colonies and wild pollinators is dependent on the farm context and availability of other forage. Pollen is an essential source of protein, lipids, minerals and vitamins that are important for larval growth and reproductive success (Génissel et al., 2002; Moerman et al., 2016). Pollinators are known to change their foraging behaviour in response to the nutritional content of plant pollen, which is thought to contribute to differing visitation rates to particular flowers (Roulston et al., 2000; Somerville, 2001). Plants with a high pollen protein content were more likely to be visited by bumblebees (Hanley et al., 2008; Leonhardt & Blüthgen, 2012; Ruedenauer et al., 2016), although amino acid profiles might be more important than protein content in predicting preferences (Kriesell et al., 2017). This level of nutritional information is available for a relatively select number of plants and needs to be expanded for studies on landscape-scale pollen preferences. However, the relative nutritional importance can be signalled through pollinator preferences.

Mass-flowering crops (MFC) provide a “pulse” of resources for a short period of time, which can boost pollinator populations, provided necessary nesting resources are present in the landscape. MFCs have been found to attract pollinators to the crop from surrounding habitats (Westphal et al., 2006; Kovacs-Hostyanski et al., 2013), and thereby dilute pollinator densities in semi-natural habitats (Holzschuh et al., 2016). The positive effects of MFCs on pollinator populations is reliant on nesting habitats being available within semi-natural

habitat. For examples density of pollinators in MFCs increases with increasing proportion of field margins, hedgerows and forest, surrounding the crop (Diekötter et al., 2014; Gardner et al., 2021).

Alternatively, sometimes crops cannot provide the nutrients required by the pollinators, and they are required to expand their foraging diet by visiting flowers elsewhere. If there is an abundance of resources close to the colony, such as a mass-flowering crop, foraging bees will travel shorter distances than in resource-poor landscapes (Hemberger & Gratton, 2018; Westphal et al., 2006). However, there is a trade-off if there are patches of flowers further away that are higher in density or attractiveness, and bumblebees will travel greater distances to access them (Cresswell et al., 2000). In cases where the MFCs are not fulfilling pollinator diets, supplementary resources become important determinants of pollinator population viability (Beduschi et al., 2018).

The proportion of crop versus non-crop pollen collected by bumblebees, will depend on the relative attractiveness of the crop and other available floral resources. In agricultural contexts, floral and pollen constancy can be used as an indicator of pollination efficiency (Marzinzig et al., 2018; Stubbs & Drummond, 2001). For example, *V. corymbosum* crops have been found to be a less attractive resource than flowers in the surrounding area, which has been demonstrated in pollen diet analyses in wild bumblebees (Toshack & Elle, 2019), honey bees (Colwell et al., 2017) and three species of managed bee (Bobiwash et al., 2018). This could be problematic for growers because a high visitation rate is desired for cross-pollination between conspecific flowers for it to result in a fertilisation event. However, Kendall et al. (2020) suggest fewer than five pollinator visits are required for *V. corymbosum* flowers to produce 90% chance of fruit set, suggesting a lower visitation rate may not present such a problem, especially in farms where the bumblebee stocking densities are very high. Similarly, increasing wild pollinator populations has been shown to increase *V. corymobsum* pollination success (Button & Elle, 2014).

Therefore, an understanding of bumblebee foraging preferences can allow for appropriate resource provision to support pollinator populations and their resulting pollination services. Investigation of floral preferences of foraging bees can be achieved through observation,

where plant-pollinator visitations are recorded, although this approach is likely to miss rare interactions or show bias towards ground-level herbaceous plants (Allen & Davies, 2023; Olesen et al., 2011). A more accurate view of bee foraging might be determined by identifying pollen taxa from the body of the bee, as it represents flower visitations that might have been missed from observations (Arstingstall et al., 2021; Carvell, Westrich, et al., 2006; Pornon et al., 2017). Pollen loads can be morphologically identified using light microscopy, although it is a time-consuming task that requires high levels of expertise, with many taxa unable to be identified past family or genus level (Bell et al., 2017; Hawkins et al., 2015).

DNA metabarcoding, a molecular approach that uses genetic markers to detect taxa in pollen loads, has become a popular alternative to microscopic techniques. Pollen DNA metabarcoding provides a higher taxonomic resolution than microscopy and has a higher throughput when comparing the time taken and costs (Bell et al., 2018; Keller et al., 2015; Richardson, Lin, Sponsler, et al., 2015; Smart et al., 2017). When used in qualitative studies, i.e. the presence/absence of pollen taxa, metabarcoding identifies a higher number of taxa than microscopy, due to its ability to detect rare species (Baksay et al., 2022; Bell et al., 2018). Quantitatively, estimating the relative abundances using metabarcoding has proved to be more challenging. Biases arising from PCR amplification and copy number have cast doubt over the accuracy of pollen quantification, with DNA from certain taxa being amplified more than others, thereby skewing the sequence relative abundances (Baksay et al., 2020; Bell et al., 2017; Lamb et al., 2019; Pornon et al., 2016; Richardson, Lin, Quijia, et al., 2015). As a result, metabarcoding has most successfully been applied in a semi-quantitative, or purely qualitative manner, providing presence-absence type, or binary data for each pollen load, with relative abundances derived from numbers of occurrences at pollen load level, rather than from number of sequence reads. Improvements have been made by using multiple barcodes and reducing PCR cycles, but still present a degree of inaccuracy (Richardson, Lin, Quijia, et al., 2015). PCR-free approaches, such as whole-genome sequencing (WGS), have shown promise in reducing these quantitative biases (Bell et al., 2021; Lang et al., 2019).

Whole Genome Sequencing (WGS) uses the entire genome, containing a high number of variable loci in comparison to the short regions used in metabarcoding, which improves the taxonomic resolution. The Reverse Metagenomics (RevMet) approach is a WGS technique that uses a low-coverage (1x) reference library to identify pollen taxa from long read MinION sequences by mapping a reference dataset of genome skims to long read nanopore pollen sequences (Peel et al., 2019). The 'reverse' nature differs from the standard metagenomic approach, in that usually short query sequences are mapped to a reference library of assembled genomes. However, there are few assembled genomes for eukaryotes due to their large size, and creating a set of low-coverage genome skims is a low-cost alternative. RevMet is a PCR-free technique, removing potential amplification biases, that has shown promising quantitative results when compared to mock mixes of extracted DNA. It has not yet been applied in an ecological study using a large number of bee-collected pollen samples.

Here, we apply RevMet to pollen taken from commercial bumblebees in four farms growing *Vaccinium corymbosum* during the crop flowering period. We expect to find bumblebees feeding on the crop, but also foraging for other sources of pollen in the wider landscape. As the crop season progresses into late spring, we would expect to see a higher flowering plant taxa richness and abundance in the landscape, so we hypothesise that the pollen diet composition will shift and the proportion of *Vaccinium corymbosum* pollen collected by workers will decrease as more attractive resources are available. Specifically, the aims of this study are to 1) investigate how the pollen diet of workers varies over the crop flowering period in different colonies, 2) measure the floral richness of the pollen at individual and colony levels, 3) compare the abundance of plant taxa in the pollen diet to available floral resources in the landscape, and 4) test the effects of landscape floral richness and crop flower cover on the proportion of *Vaccinium* pollen collected by workers.

4.3 Materials and Methods

4.3.1 Sample sites

The study was conducted in spring 2021 at four farms in southern England; Heathlands farm in Berkshire (51°23 34"N 0°49' 06"W), Winterwood farm in Kent (15° 31' 09"N, 0° 36' 00"W), Tuesley farm in Surrey (51° 10' 02"N, 0° 37' 16"W) and Colworth farm in West Sussex (51° 49' 01"N, 0° 42' 07"E). All farms are commercial, small to medium-sized (20 – 161 ha) and grow soft fruit including strawberries, raspberries, blackcurrants, and blueberries. In March, highbush blueberry (*Vaccinium corymbosum,* hereafter referred to as *Vaccinium*) crops established within Spanish polytunnels began flowering. Spanish polytunnels are large, field-scale structures that consist of metal frames covered by polyethylene sheets. These are usually open within daylight and closed to provide thermal protection to the crops at night. The farms also grew strawberry crops (*Fragaria x ananassa*) in Spanish polytunnels, which started flowering in April. In May, new *Vaccinium* crops were planted outside and started flowering towards the end of the month.

The farms stocked commercial buff-tailed bumblebees (*Bombus terrestris*) in March to supplement the pollination services required to successfully pollinate highbush *Vaccinium* crops and produce marketable fruit. Koppert Tripols (Koppert Biological Systems, Haverhill, UK) and Biobest Multi-Hives (supplied by Agralan, Swindon, UK) both contain three large colonies of *B. terrestris*, including a queen, approximately 300 workers and a brood. The lifespan of commercial colonies is approximately 6-8 weeks, which is the duration of the *Vaccinium* flowering period. The farm stocking densities ranged from 18 – 45 colonies per hectare.

All four farms were sampled for bumblebee-collected pollen and flowers in the surrounding landscape, once every two weeks over the *Vaccinium* flowering period (March to May). Floral surveys were conducted either before or after the bee collected pollen samples, as the pollen sampling protocol was weather dependent, but both were always completed within a three-day period. There were four sampling periods in total; the first 29th March – 9th April, the second 13th – 22nd April, the third 29th April – 7th May and the fourth 10th May – 1st June.

The third sampling period for Tuesley farm fell in the fourth block (18th May) but we decided to keep it in the third in order to get a complete time series for this farm. There were only three sampling dates for Winterwood, because in the fourth sampling period there were no foraging workers in the colonies and *Vaccinium* flowering had ended (details of sampling periods in Supplementary Table S4.1).

4.3.2 Floral sampling

Habitat and land use were mapped in a 1km radius from the centre of the selected field using QGIS v3.4 (Google Earth imagery, 2021; Figure 4.1) in order to capture the full foraging range of *B. terrestris* workers (Osborne et al., 2008). The range of habitats surveyed consisted of cereal crop, flowering crop, woodland, grass (mostly improved), field margins, hedgerows, gardens, and manmade structures (e.g., buildings, roads). Due to COVID-19 lockdown measures when fieldwork taking place in March 2021, gardens were not accessible and therefore excluded from floral surveys. Each habitat type was sampled relative to their proportional area for a total of 20 transects (Table 4.1). For example, if a farm had a larger proportion of woodland, an increased number of transects would be placed in that habitat, compared to other less abundant habitats. In addition, six hedgerow transects were also carried out on each farm because although they make up a small proportion of the landscape they are florally diverse habitats (Garratt et al., 2017; van den Berge et al., 2019).

Therefore, in total, 26 transects were carried out on each farm during each sampling round. The transects were randomly allocated in QGIS for each sampling period and were conducted over 50 m, with a 1m x 1m quadrat placed every 5 m (for a total of ten quadrats per transect). The number of flower units were counted in each quadrat, following Baude et al. (2016), which was averaged to provide the number of flower units per m^2 per transect. This value was multiplied by the area of habitat in the 1 km buffer around each farm, to give a landscape-level estimate of flower density for each plant species within bumblebee foraging range of each sampling location.

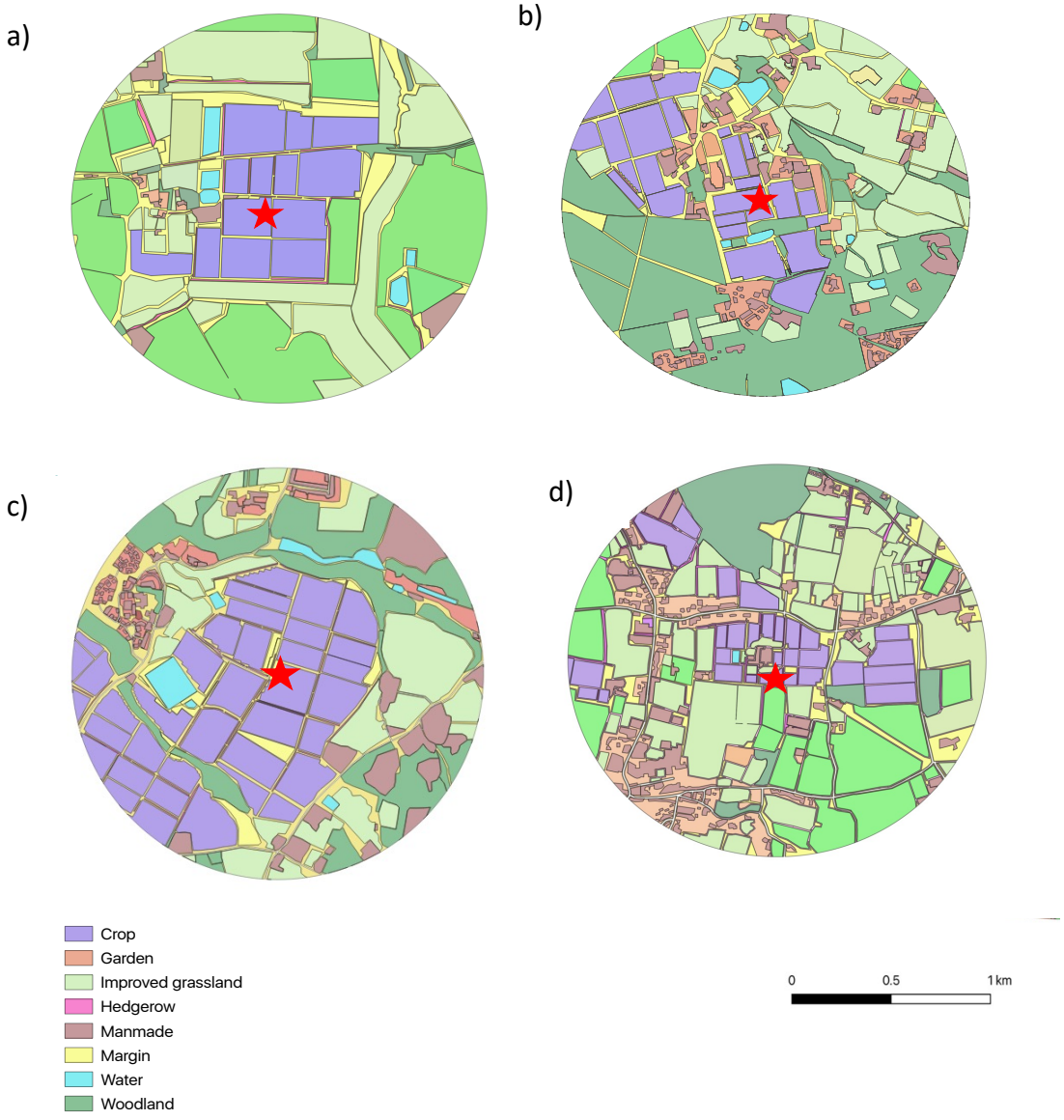


Figure 4.1 Land use in the 1 km radius surrounding the colonies sampled on each farm. The red star denotes the field containing the six bee colonies. The farm identities are a) Colworth, b) Heathlands, c) Tuesley, and d) Winterwood.

Habitat type	*Colworth*	*Heathlands*	*Tuesley*	*Winterwood*
Cereal	32.5%	1.8%	0%	17%
Building	0.2%	6.7%	10.8%	8.4%
Soft fruit crop	6.3%	14.9%	32.3%	9.8%
Garden	0.4%	8.2%	3.4%	7.3%
Improved grass	41.2%	23.2%	16.6%	29.3%
Hedgerow	1.7%	1.0%	2.1%	1.0%
Margins	13%	9.2%	15.1%	9.6%
Water	2.5%	1.4%	2.2%	0.1%
Woodland	2.2%	34.4%	17.5%	9.4%

Table 4.1 Proportions of the different habitats surveyed on each farm. The area measured was a 1km radius around the field closest to the centre of the farm. This field contained the six colonies sampled for the study.

4.3.3 Pollen sampling

A field containing *Vaccinium* crops under polytunnels in the centre of the farm was selected and six commercial colonies (one colony per tripol) were picked at random from the field centre, or as close to the centre as possible, with a maximum distance of 30 m from the field centre. Sampling took place on days where the outside temperature was over 14°C, the entrances to the polytunnels were open and between the hours of 10am and 4pm. During each sampling occasion, the colonies were closed and *B. terrestris* workers carrying pollen loads were intercepted on their return. Using a queen marking tube, the workers were captured and pollen loads on both corbiculae were removed with a mounted needle, which was wiped using alcohol wipes after each sample. Once 12 bees had been captured or an hour had passed, the colony was opened again. This meant foraging bees would not be caught twice on one sampling period, and also avoided disturbing the pollen supply of the colony. The pollen samples were stored in Eppendorf tubes, kept on ice and transferred to a -20°C freezer within eight hours and subsequently kept at – 20°C until DNA extraction.

4.3.4 DNA extraction

The pollen loads from each bee were combined by adding 1 ml of 80% ethanol to Eppendorf tubes and homogenised by pipetting. 350 µl of the suspended pollen mixture was transferred to a new collection tube and spun down on a centrifuge at 15,000 rpm for 1 minute to remove the ethanol with a micropipette without disturbing the pollen grains. The method of DNA extraction used was a modified bead washing protocol (Heavens et al., 2021). Using reagents from the DNeasy PowerSoil Pro Kit (Qiagen), 150 mg of beads and 400 µl of CD1 lysis buffer were added to the pollen tube and disrupted using a Geno/Grinder® for 5 minutes at 15,000 rpm. The tubes were spun in a centrifuge for 1 minute at 15,000 rpm and the supernatant was transferred to a fresh 1.5 ml Lobind Eppendorf tube. Next, 400 µl of KAPA Pure Beads (Roche Diagnostics Ltd, West Sussex, UK) were added to 400 µl of the lysed and bead beaten pollen grains. The pollen grains were vortexed and incubated for 5 minutes at room temperature, and pulse spun in a microcentrifuge. The tube was then placed on a magnetic particle concentrator (MPC) for five minutes to allow the beads to concentrate. After the supernatant was discarded, the beads were washed twice with fresh 70% ethanol. All the ethanol was carefully removed, and the tube taken off the MPC. The beads were resuspended in 10 µl of Qiagen CD6 buffer and incubated at room temperature for 2 minutes. Next, the tube was pulse spun in a microfuge and placed in an MPC to allow the beads to concentrate. The supernatant (approx. 10 µl) containing the DNA was then transferred to a fresh 1.5 ml Lobind Eppendorf tube and a Qubit fluorometer using the dsDNA BR assay kit was used to assess the concentration of eluted DNA (Life Technologies, Loughborough, UK).

4.3.4 Library Preparation and sequencing

Libraries were generated with the SQK-LSK109 kit and Native Barcoding Expansion kit (EXP-NBD196), which allows PCR-free multiplexing of up to 96 samples. The New England Biolabs manufacturer's protocol was followed (version NBE_9121_v109_revE_19Jan2021). Sequencing was performed on a GridION using nine FLO-MIN106D flow cells and the MinKNOW (version 9.4.1) software's standard 72 hour run script. Reads were base-called by

Guppy (version 4.2.2) using the high accuracy model and a minimum pass read Q score of 9. A negative control was included in each of the sequencing runs, which contained ultrapure water in place of DNA elute, and was processed following the same protocol as above.

4.3.4 RevMet pipeline

A plant reference dataset was created from the flowering plant species encountered on floral transects during 2019 at two farms and 2021, at two additional farms. The reference dataset was comprised of 75 plant genome reference skims created in Chapter Two and 49 skims created by Peel et al. (2019) (see Chapter Two methods for further details). The RevMet pipeline from Peel et al. (2019) was adapted and used to assign plant taxa to the pollen loads. In brief, Illumina reads from the 124 plant genome skims were mapped against every long read MinION sequence in each of the samples. Unmapped reads and those under 15% coverage were classified as “unassigned”. The plant taxa with the highest percent coverage of the MinION sequence was assigned to that read.

4.3.5 Statistical analysis

The raw RevMet reads assigned to each taxon were used as the response variable, and unless specified these were converted to relative abundance. A minimum abundance filter removed any plant taxa that were present in a sample at less than 1% because they were unlikely to be an important pollen resource or may have resulted from contamination in the field or lab, and likely represented false positive assignments. Species accumulation curves were also run for each sample to check for sequencing biases, and any samples that did not reach asymptote were also removed from analysis (Supplementary Figure S4.4). All statistical analysis was performed in R v4.0.2 (R Core Team, 2020).

4.3.5.1 Aim 1: Comparison of bee-collected pollen communities among colonies and sites

We expect pollen loads to change depending on the floral diversity in the landscape, which is known to vary spatially and temporally (Lowe et al., 2022). We explore this at the colony and individual levels.

To test for changes in pollen composition for individuals across farms and sampling periods, we used a series of multivariate generalised linear models (MGLMs). Model-based community-analyses like these are being adopted over distance-based methods (e.g., Principal Coordinate Analysis) due to their improved power (Warton et al., 2012). MGLMs allow users to specify the mean-variance relationship of the dataset, and are less affected by taxa with the highest variation in abundance. We constructed MGLMs to test for changes in pollen taxa composition across farms and sampling periods using the mvabund package, which fits an individual GLM for each taxon (Wang et al., 2012). The plant taxa included in the model were those present at >1% in individual pollen loads in more than five samples. We used an offset of the total number of assigned reads to control for the differences in assigned read number per pollen load. The data were zero-inflated, as is common with multivariate abundance data, and a negative binomial distribution best fit the residuals (Supplementary Figure 4.1). Post-hoc analyses with adjusted p-values (bootstrap resampling) were used to find plant taxa that were driving seasonal differences (Warton et al., 2017).

For colonies, we used permutational multivariate analysis of variance (PERMANOVA) to investigate differences in pollen collection between the colonies on the same farm across the sampling periods. We chose this method, because our colonies samples contained different numbers of bees and MGLMs are unsuitable for unbalanced designs such as we have. We calculated Bray-Curtis dissimilarity matrices based on the proportional data from individual pollen loads using the vegan package. Separate PERMANOVA analyses were carried out for each farm, with 10,000 permutations. Nonmetric multidimensional scaling (NMDS) was used to visualise the beta diversity of the pollen loads and plotted using ggplot2 (Wickham, 2011). Only colonies containing five or more bee samples were included in the analyses.

4.3.5.2 Aim 2: Pollen richness

Pollen Richness within Individual Pollen Loads

The pollen richness of pollen loads provides information about floral constancy and the effects of increasing levels of landscape floral richness. We recorded the number of plant taxa per individual bee and tested whether sampling occasion had an effect on the pollen richness using a generalised linear model.

Pollen Richness at the Colony Scale

To test the effects of landscape floral richness, *Vaccinium* flower cover, sampling period and farm on the colony pollen richness we applied a generalised linear mixed model using the glmmTMB package (Brookes et al., 2017). Colony ID nested in farm was included as a random effect, and the model was fitted with a Poisson distribution, as is typical for count data (Sellers et al., 2012). The diagnostic plots of the residuals were tested using the DHARMa package (Hartig, 2020). We used a multimodel inference approach on the models using the 'dredge' function of the MuMIn package (Burnham & Anderson, 2002; Barton 2022) which lists candidate models from all possible combinations of explanatory variables. We then performed model selection based on AICc (Wagenmakers & Farrell, 2004) and report the models within Δ2 AICc.

4.3.5.3 Aim 3: Comparison of pollen communities to landscape floral resources

First, to measure the frequency of plant visits in a single foraging trip, we looked at the proportion of plant taxa in individual pollen loads (Jones et al., 2021). We assigned abundance classes based on the percentage of assigned reads for each taxon per pollen sample. Plants that were represented over 45% were classed as major, taxa between 10 and 45% were intermediate, and if present in a pollen load at less than 10% the taxa was classed as minor.

Second, to test whether the bees collected pollen from plants more or less than expected by chance based on their availability in the landscape, we took a null modelling approach using

the R package econullnetr (Vaughan et al., 2018). This modelling approach generates a null prediction, with a 95% confidence interval, for the relative amount of each pollen taxon that should be collected by individual bumblebees based on the relative abundance of flower units in the landscape. The null predictions assume the bumblebees collect resources in the same proportions as their availability, with no preference or aversion. We used quantitative counts in the model based on the number of reads allocated to each plant taxa by RevMet rather than a presence-absence model, which overinflates the importance of rare taxa (Deagle et al., 2019). We included plant taxa that were present at over 1% abundance in any single pollen sample or over 1% in the floral surveys from a single sampling period, because taxa present in low abundances are likely to be overestimated in importance (Warton et al., 2012). Since sampling period has an effect on pollen taxon composition, we modelled the four sampling periods separately.

4.3.5.4 Aim 4: Proportion of Assigned *Vaccinium* reads

We expect the proportion of *Vaccinium* pollen collected by foraging bees to be affected by the flower cover of the *Vaccinium* crop and also the flower richness of the surrounding landscape which can draw bees away from the crop. A generalised linear mixed model was created with proportion of assigned *Vaccinium* reads per sample as the response variable, and species richness of flowering taxa in the landscape, *Vaccinium* flower cover (estimated from flower transect surveys), sampling period and farm as explanatory variables, and colony ID nested in farm as a random effect using the glmmTMB package (Brookes et al., 2017). We fit the model with a beta distribution, which is most appropriate for proportional data (Cribari-Neto & Zeileis, 2010) and assessed the residuals with diagnostic plots (Hartig, 2020). Model selection was the same as above, listing all candidates from a global model and selecting the model with the lowest Δ AICc. We tested for effects of farm and sampling period using the car package and a type II Wald Chi-square test and applied post-hoc comparisons using the package emmeans to test for differences between farms and months, with an alpha of 0.05 (Lenth et al., 2018).

4.4 Results

4.4.1 Flower surveys

Over the course of the *Vaccinium* flowering period, a total of 390 floral transects were carried out over 15 visits across the four farms. Winterwood Farm only received three visits because on the final sampling period there were no foraging workers returning with pollen loads. Flowering plant species richness ranged from 5 to 25 species per farm, and the number of flower units per site ranged from 980 to 16669 (Supplementary Figure 4.2; Supplementary Figure 4.3). A complete species list per site is given in the Supplementary Table 4.2.

4.4.2 Sequencing results

In total, the pollen loads of 840 bumblebees were collected. Nanopore sequencing of the corbicular loads produced 27,675,120 reads in total with an average of 33,709 reads per sample and a mean read length of 2,583 bp.

Of the 840 pollen samples sequenced, 53 (6%) produced under 1000 reads and were removed from the dataset (Supplementary Figure 4.4). A further 38 samples had under 20% assigned reads and were also removed. I have applied a threshold of 20% assigned reads, because below this number we cannot be sure that the low proportion of assigned reads is not due to an incomplete reference database. Including these samples would overinflate the importance of the composition of taxa present, therefore they have been excluded from statistical analyses. Therefore, 749 pollen samples were included for statistical analysis.

Commercial bees collected pollen from 61 taxa**,** 52 genera and 21 families during the *Vaccinium* flowering period. There were 368 pollen samples (49.0%) that contained a single species of pollen, whilst 208 samples (27.8%) contained two species. The majority of pollen loads were dominated by a small number of taxa. When reads were calculated over all sampling periods and farms, the two most abundant taxa accounted for 76.1% of total

assigned reads; *Vaccinium* was assigned 45.5% of reads and *Salix caprea* 30.6% (a full list is provided in the Supplementary Table 4.3).

4.4.3 Aim 1: Comparison of bee-collected pollen communities among individuals, colonies and sites

Both sampling period ($LR_{3,776}$ = 468.0, p < 0.001) and farm ($LR_{3,779}$ = 375.3, p < 0.001) were good predictors of individual bees' pollen compositions (Figure 4.2). Post-hoc comparison of the sampling periods showed pollen collection varied significantly across all four time periods between all pairs (p = 0.002; Supplementary Table 4.4). These results suggest there is a phenological shift throughout the crop flowering period, as floral resource availability changes with the pollen diet of the bees. The farms also showed different patterns of pollen foraging by the workers with consistent differences across pairs of farms (Supplementary Table 4.4).

After adjusting P-values for multiple tests, univariate analyses for each individual plant taxon showed an increase in abundance of reads across the sampling periods for *Crataegus monogyna* (LR = 137.1, p = 0.002), *Leontodon saxatilis* (LR = 72.1, p = 0.004), *Pyrus communis* (LR = 151.9, p = 0.002), *Ranunculus acris* (LR = 47.8, p = 0.002), *Ranunculus repens* (LR = 57.4, p = 0.014), *Rubus fruticosus* (LR = 54.3, p = 0.018). Conversely, *Salix caprea* (LR = 120.2, p = 0.002) decreased in read abundance across the season, and two species, *Fragaria* (LR = 63.7, p = 0.012) and *Brassica rapa* (LR=64.0, p = 0.012) showed peaks during sampling period 3 (Supplementary Figure 4.5).

Communities of pollen taxa collected by bumblebee colonies were significantly different among sampling rounds, based on the results of the PERMANOVA (p < 0.002) (Supplementary Table 4.5)**.** However, there were no significant between-colony differences in pollen communities located on the same farm ($p > 0.05$). NMDS ordination of Bray-Curtis dissimilarities displayed no distinct separation between colonies in the pollen communities being collected, but there was evidence of clustering by sampling period (Figure 4.3).

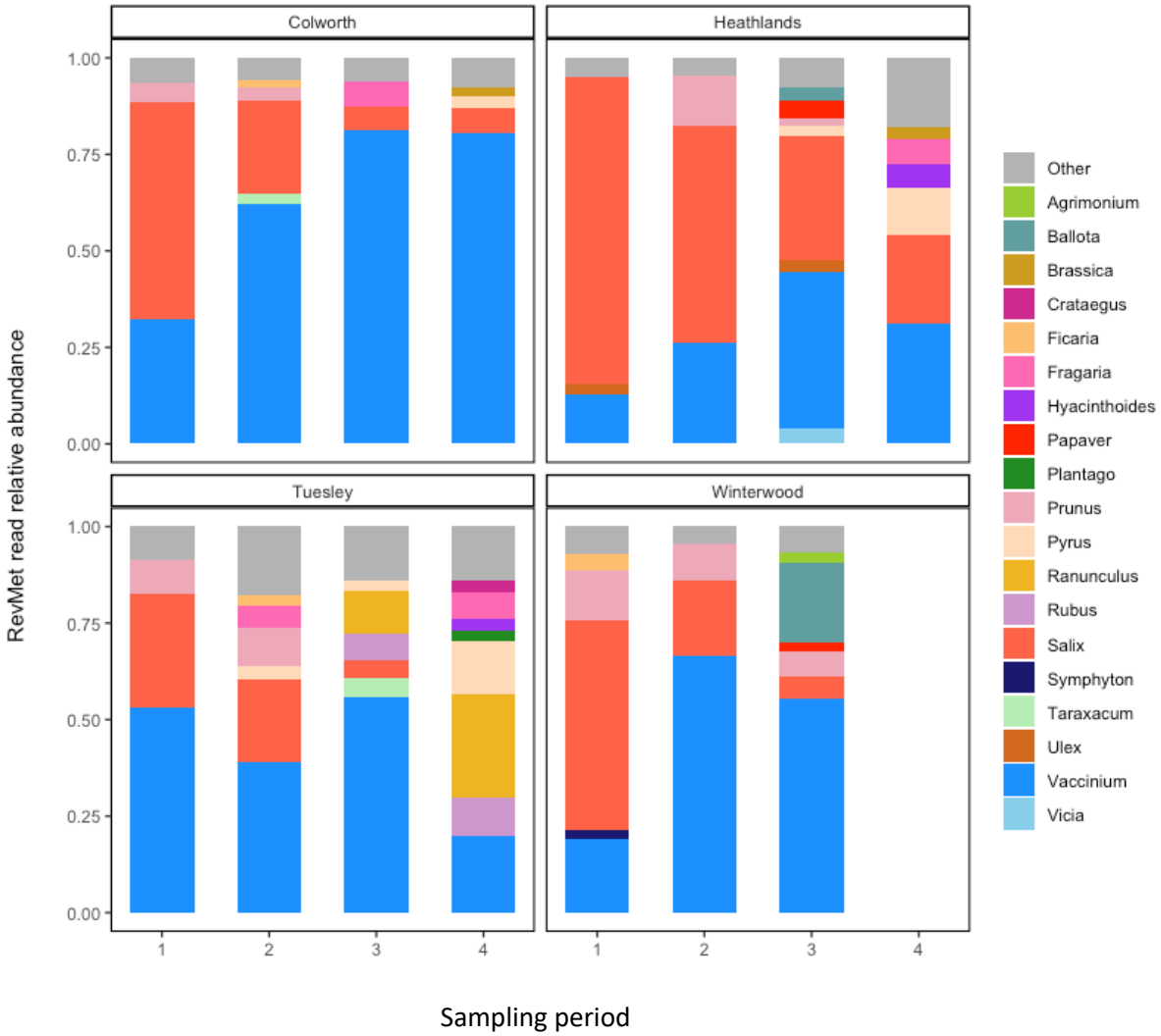


Figure 4.2. The origin of pollen from each sample period and farm. The sampling periods started in late March and were repeated approximately every 2 weeks until the end of May. The "Other" category includes taxa present at < 5% abundance at the colony

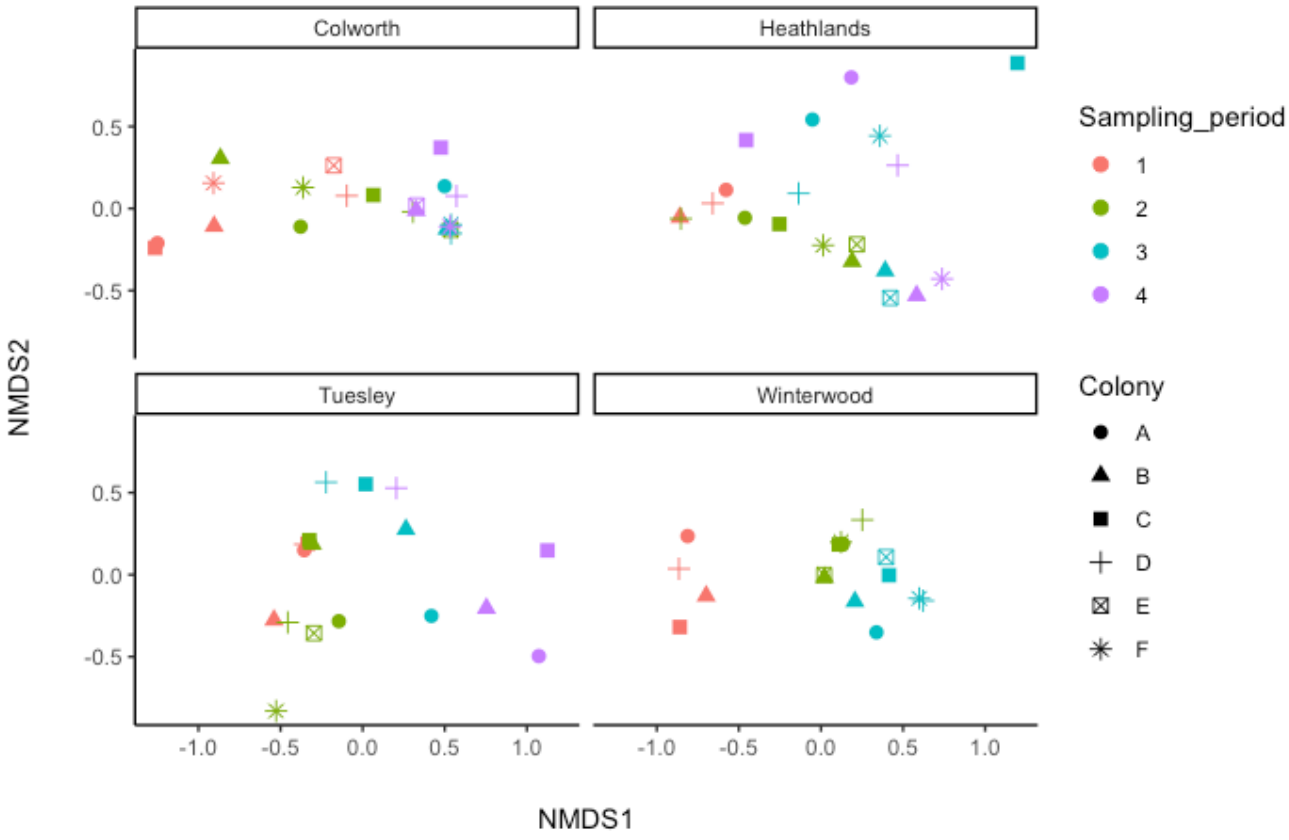


Figure 4.3 NMDS plots displaying the variation in colony-level foraging on each farm. Each point represents the mixed pollen diet of a colony. Colonies were included if they contained five or more bee samples. The stress for the plots are Colworth = 0.04, Heathlands = 0.09, Tuesley = 0.12, Winterwood = 0.07

4.4.4 Aim 2: Pollen richness

The average number of plant species per individual pollen load was 2.34 (SE ± 0.07), which did not vary significantly between the sampling periods (df =3, F = 1.84, p = 0.16). However, there was a significant effect of Farm (GLM t = 15.2, p < 0.001). On Colworth Farm the number of species per pollen load was 1.68 (SE ± 0.07), which was lower than Heathlands (2.25 SE ± 0.13), Tuesley (2.38 SE ± 0.14) and Winterwood (2.08 SE ± 0.12).

The average species richness of pollen brought in by the colonies over the crop flowering our sampling period was 4.28 (SE ± 0.25), higher than the number of species carried by each individual bumblebee. The effects of crop flower cover, flower richness in the landscape and sampling period on the richness of pollen collected by colonies were explained included in the final model (dAICc < 2) (Supplementary Table 4.6, Supplementary Table 4.7). We used the sum of Akaike weights (Σwi) to assess the relative importance of each explanatory variable, and found the *Vaccinium* flower cover to have the greatest importance (Σwi = 0.93) followed by landscape flower richness (Σwi = 0.80). Pollen richness increased with landscape floral richness (z = 2.05, p = 0.04; Figure 4.4) but decreased with *Vaccinium* flower cover (z = -2.06, p = 0.04). Sampling period (Σwi = 0.65) and the interaction between sampling period and crop flower cover (Σwi = 0.61) were also included in the best fitting model (dAICc = 0), but were assigned lesser importance (Supplementary Table 4.8).

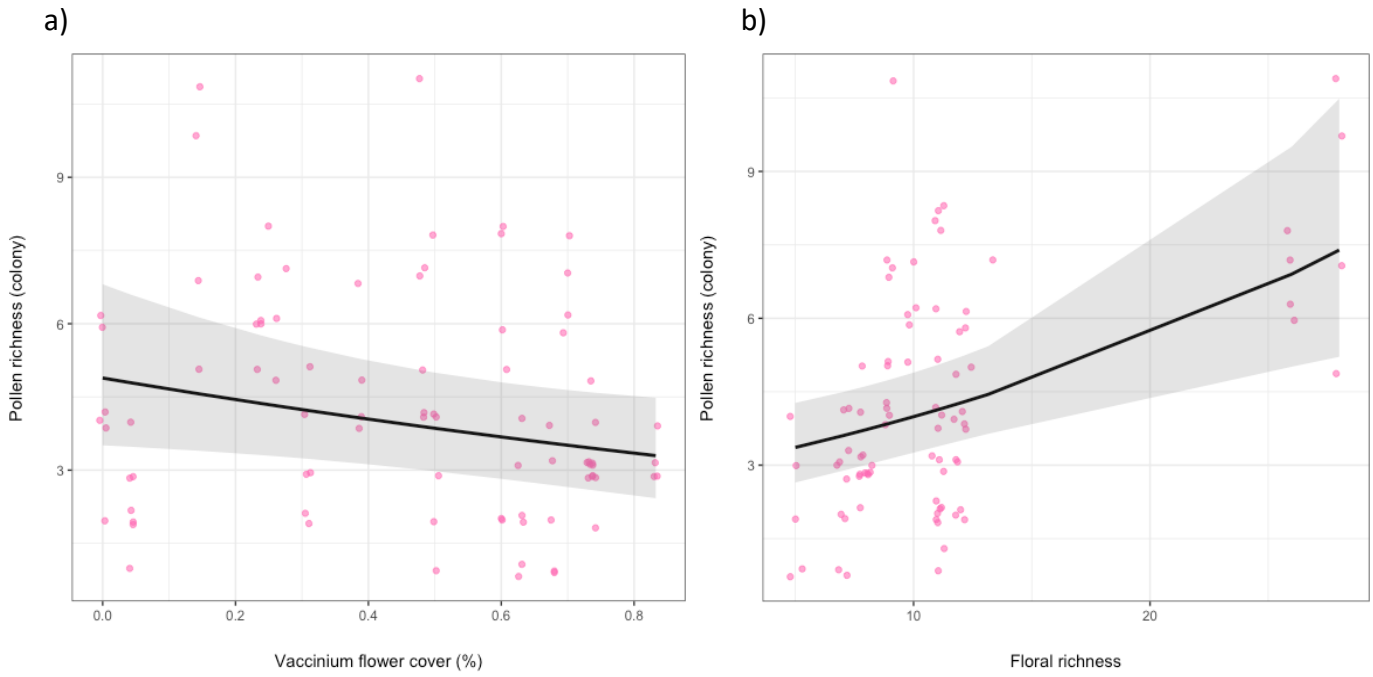


Figure 4.4 The effects of a) Vaccinium flower cover and b) landscape floral richness on the richness of pollen collected by the bumblebee colonies (n = 85). The regression line is obtained from mixed effect model estimates and the points are jittered for visual purposes.

4.4.5 Aim 3: Comparison of pollen communities to landscape floral resources

Over the whole *Vaccinium* flowering period 16 taxa were classified as a major source of pollen for at least one pollen load (>45% assigned reads): *Vaccinium*, *Salix caprea*, *Prunus spinosa*, *Pyrus communis, Rubus fruticosus, Fragaria, Ficaria verna, Ranunculus repens, Anthriscus sylvestris, Vibernum lantana, Ulex europaeus, Taraxacum agg., Hyacinthoides non-scripta, Agrimonium eupatoria, Ballota nigra* and *Vicia sativa* (Figure 4.5).

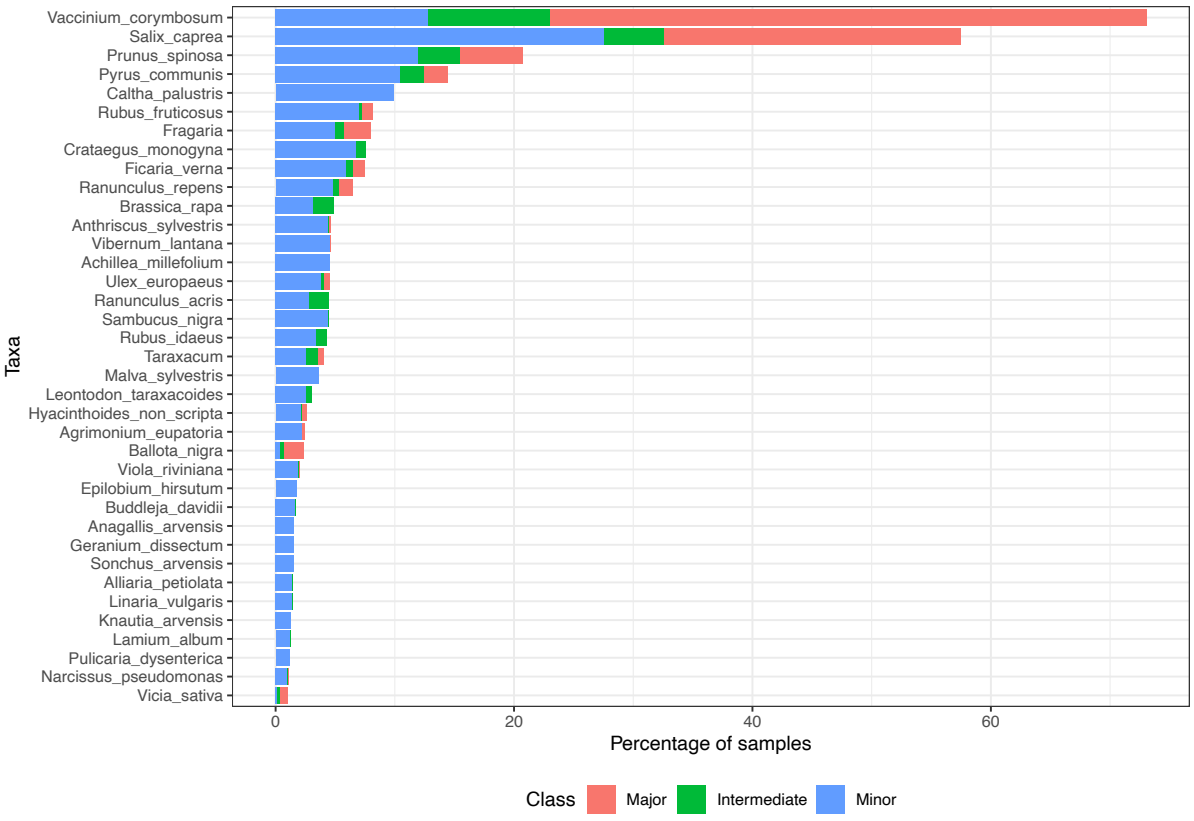


Figure 4.5 The frequency class of each plant taxa in the individual pollen loads. Classes represent the proportion of reads; major (> 45%), intermediate (10-45%) and minor (< 10%).

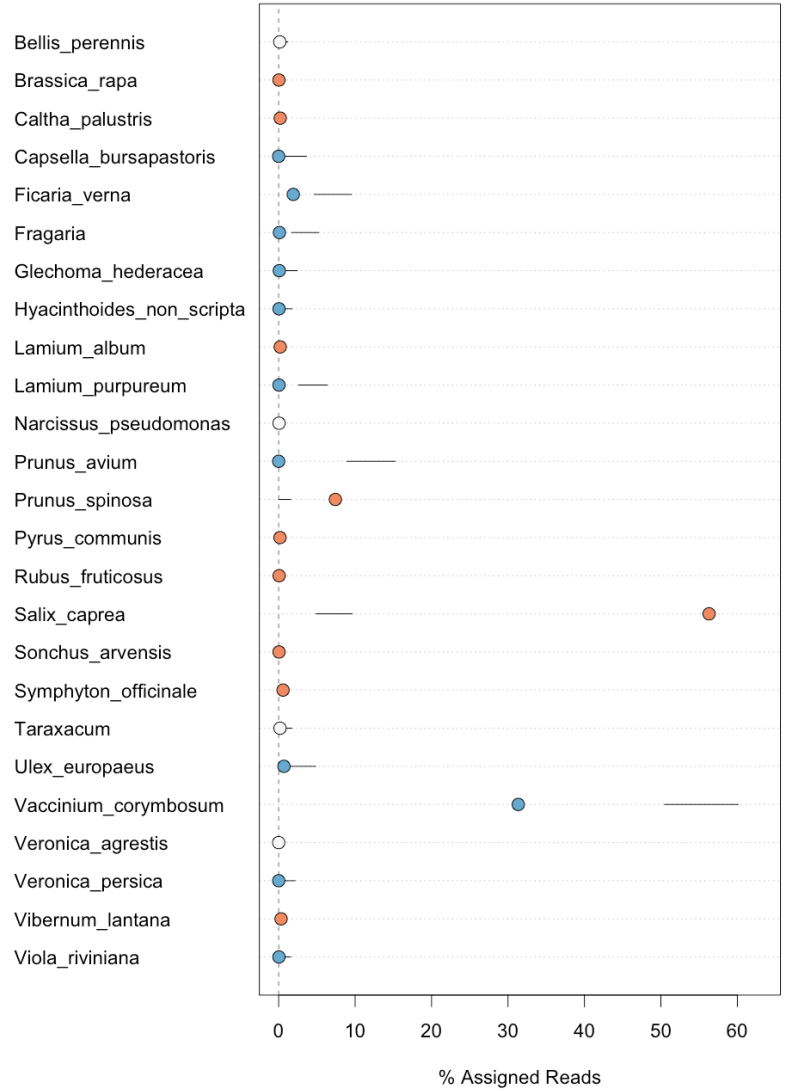
The colonies did not visit the full availability of floral resources in the landscape, as a relatively low proportion of taxa were identified in the pollen (Table 4.2). Comparing the observed pollen diet breadth from the RevMet data with the abundance of floral resources in the landscape for each sampling period revealed few taxa were collected by bumblebees relative to the floral abundance of that taxon (Figure 4.6). There were 68 plant taxa found at >1% abundance in the sequence reads or the landscape and therefore included in the model.

Over all four sampling periods, *Salix caprea* (genome size 392 Mbp) and *Prunus spinosa* (637 Mbp) were collected significantly more than predicted based on their abundance of floral units in the landscape. Other taxa that were collected more than predicted included *Pyrus communis* (591.7 Mbp)*, Ulex europaeus* (3822 Mbp) and *Ballota nigra* (1420 Mbp). The preference of certain taxa was found to change over the sampling time, including *Vaccinium* (651.7 Mbp), which was collected less than predicted in the first two sampling periods, but more than predicted in the latter two periods. The same was found for the other flowering crop, *Fragria x ananassa* (324.6 Mbp), which was avoided in all sampling periods except the second where it was collected as predicted based on its' abundance. *Brassica napus*, a mass-flowering crop, was avoided in the last sampling round, when it was at its highest abundance in the landscape, signifying a level of avoidance. The same was true of *Heracleum sphondylium* (2141.3 Mbp)*, Crataegus monogyna* (744.8 Mbp) and *Hyacinthoides non-scripta* (20776 Mbp) which were collected less than predicted based on landscape values. A high number of taxa were found at low abundance in either the RevMet or the floral survey data, and some taxa located in the RevMet results were not identified in the landscape surveys. For example, *Salix caprea* and *Prunus spinosa* were two of the most frequently collected pollen sources in the latter two sampling periods, but neither were recorded in the landscape.

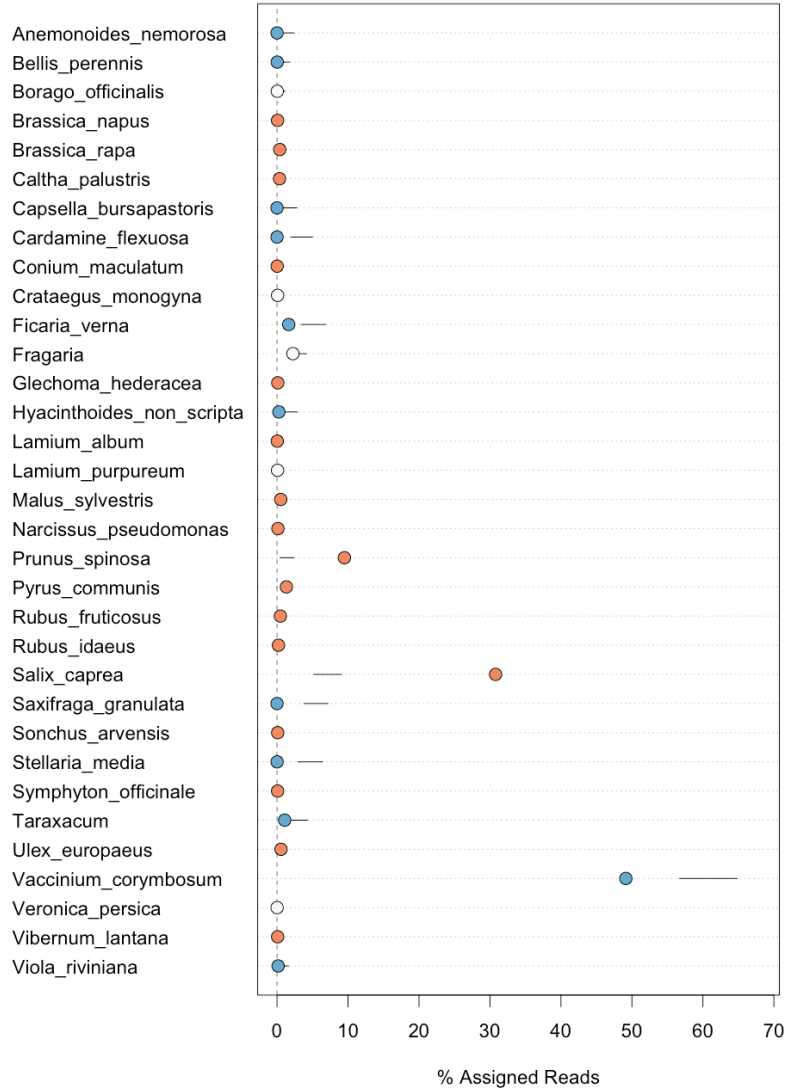
Sampling Period	Colworth				Heathlands				Tuesley				Winterwood				Total unique pollen taxa
	Pollen load richness	Pollen taxa represented in landscape	Landscape floral richness	Landscape taxa represented in pollen	Pollen load richness	Pollen taxa represented in landscape	Landscape floral richness	Landscape taxa represented in pollen	Pollen load richness	Pollen taxa represented in landscape	Landscape floral richness	Landscape taxa represented in pollen	Pollen load richness	Pollen taxa represented in landscape	Landscape floral richness	Landscape taxa represented in pollen	
1	5	80%	11	36%	4	75%	5	60%	3	67%	7	29%	6	33%	9	22%	8
2	6	67%	8	50%	4	75%	8	38%	9	44%	8	50%	3	67%	8	25%	11
3	4	50%	6	33%	11	36%	10	40%	10	20%	23	9%	7	14%	10	10%	17
4	*6*	*50%*	*10*	*30%*	8	75%	8	75%	13	62%	25	32%					20

Table 4.2 The plant taxa richness of the pollen from all colonies grouped by farm and sampling period. The plant taxa present in the pollen that were also counted in the floral surveys is calculated, with that number as a percentage of all the flower taxa recorded in the landscape.

a)


Bellis_perennis
Brassica_rapa
Caltha_palustris
Capsella_bursapastoris
Ficaria_verna
Fragaria
Glechoma_hederacea
Hyacinthoides_non_scripta
Lamium_album
Lamium_purpureum
Narcissus_pseudomonas
Prunus_avium
Prunus_spinosa
Pyrus_communis
Rubus_fruticosus
Salix_caprea
Sonchus_arvensis
Symphyton_officinale
Taraxacum
Ulex_europaeus
Vaccinium_corymbosum
Veronica_agrestis
Veronica_persica
Vibernum_lantana
Viola_riviniana
0
10
20
30
40
50
60
% Assigned Reads

b)


Anemonoides_nemorosa
Bellis_perennis
Borago_officinalis
Brassica_napus
Brassica_rapa
Caltha_palustris
Capsella_bursapastoris
Cardamine_flexuosa
Conium_maculatum
Crataegus_monogyna
Ficaria_verna
Fragaria
Glechoma_hederacea
Hyacinthoides_non_scripta
Lamium_album
Lamium_purpureum
Malus_sylvestris
Narcissus_pseudomonas
Prunus_spinosa
Pyrus_communis
Rubus_fruticosus
Rubus_idaeus
Salix_caprea
Saxifraga_granulata
Sonchus_arvensis
Stellaria_media
Symphyton_officinale
Taraxacum
Ulex_europaeus
Vaccinium_corymbosum
Veronica_persica
Vibernum_lantana
Viola_riviniana
0
10
20
30
40
50
60
70
% Assigned Reads

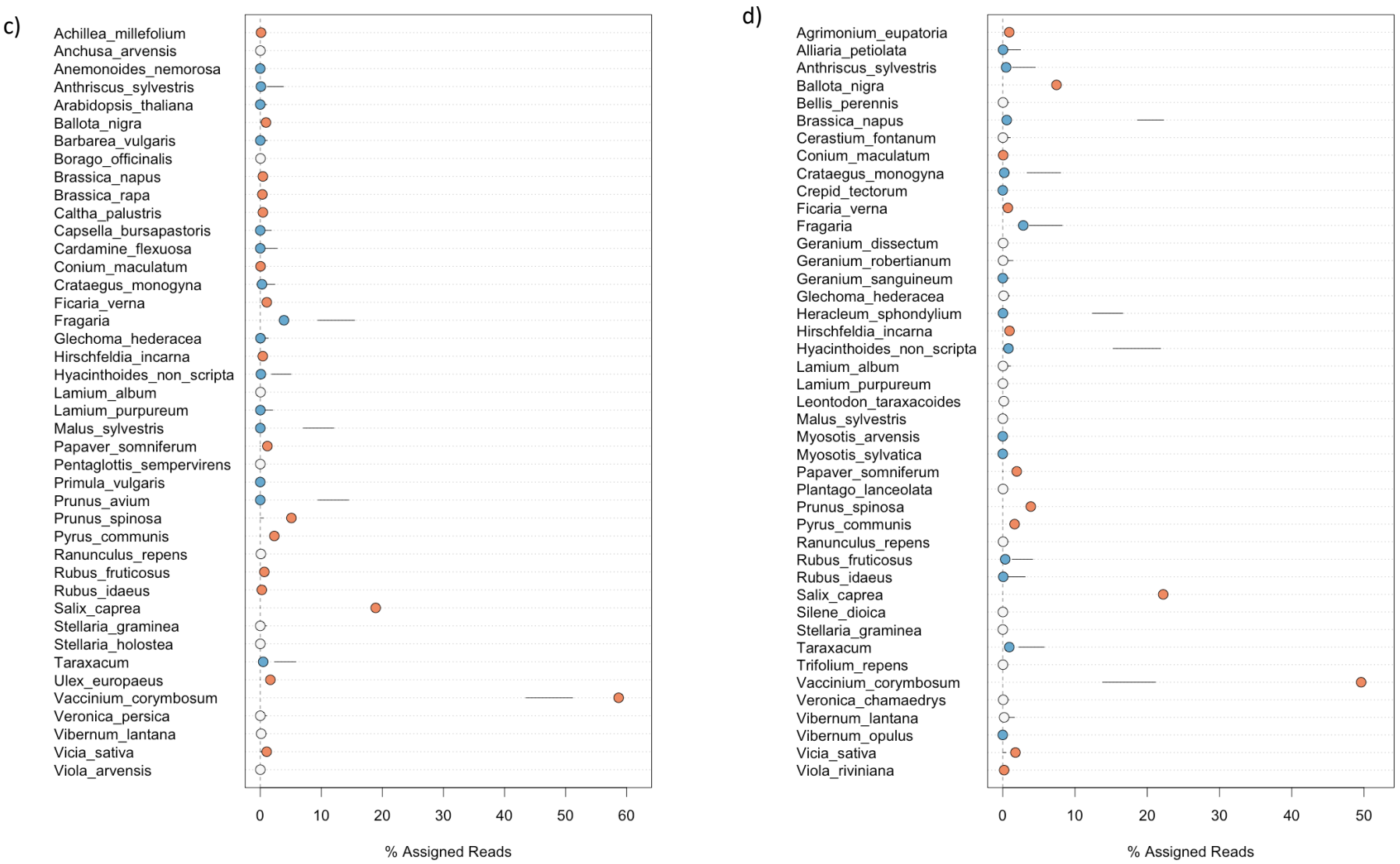


Figure 4.6 Expected and observed interactions between assigned RevMet sequences and the abundance of flowers in the landscape for Sampling periods 1 – 4 (a – d). Blue circles denote lower proportion of assigned reads than expected, white is as expected (proportion of reads is in proportion to relative abundance) and orange is a higher number of reads than predicted from relative abundance. Horizontal lines represent 95% confidence limits of the observed frequency of pollen reads.

4.4.6 Aim 4: Proportion of assigned Vaccinium reads

Two models had substantial empirical support (dAICc < 2), neither included landscape flower richness and this variable was therefore dropped (Supplementary Table 4.6). The best fitting model (dAICc = 0) included the effects of *Vaccinium* flower cover (Σwi = 0.98), sampling period (Σwi = 1) and the interaction between the two (Σwi = 0.51) (Supplementary Table 4.8; Supplementary Table 4.9). The proportion of assigned *Vaccinium* reads was higher when the *Vaccinium* flower cover was high (z = 3.37, $p = 0.04$; Figure 8). GLMM results showed significant effects of sample period ($\chi^2 = 16.9$, df = 3, $p < 0.001$) and farm ($\chi^2 = 20.1$, df = 3, $p < 0.001$) on the proportion of assigned *Vaccinium* reads (Figure 4.7). Pairwise comparison of estimated marginal means revealed a significantly lower proportion of assigned *Vaccinium* reads in the first sampling round (34.5% ± 5) than rounds three (58.3% ± 6) and four (56.3% ± 2). There were also varying proportions of *Vaccinium* pollen collected on the different farms; Colworth farm (63.5% ± 5) had a significantly higher proportions of *Vaccinium* reads than Heathlands (42.2% ± 4) and Tuesley (50.9% ± 6).

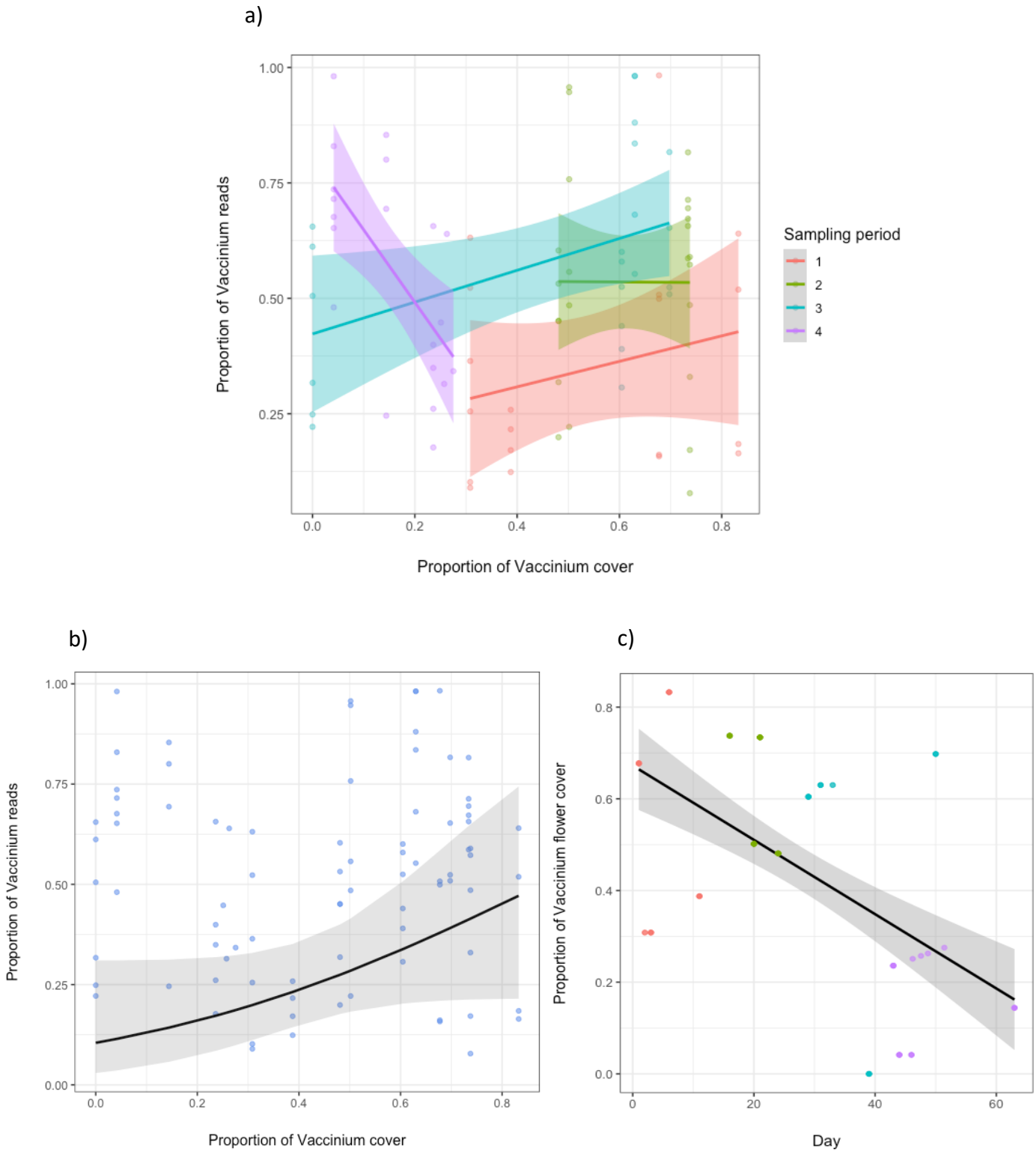


Figure 4.7 The effects of a) the interaction between sampling period and Vaccinium flower cover on the proportion of Vaccinium reads in the pollen, b) the effect of Vaccinium flower cover on the proportion of Vaccinium reads in the pollen loads, and c) the change in *Vaccinium* flower cover over the sampling period, where Day = 0 was the first day of sampling. Regression lines are obtained from mixed effect model estimates and colours in a) and c) relate to same sampling periods.

4.5 Discussion

In this study we investigated the pollen foraging of commercial bumblebees (*Bombus terrestris*) on highbush blueberry farms (*Vaccinium corymbosum*) over the crop flowering period using a nanopore sequencing approach. We found commercial bumblebees located within *Vaccinium* crops forage from a range of flowering taxa in the landscape, and not only from the crop itself. Sample period and farm site were significant predictors of the pollen community composition, likely driven by the phenology of flowering plants in the wider landscape and the different landscape configurations in the 1 km radius around the farms. We used a combination of pollen analysis and floral surveys to compare the proportions of taxa in bee-collected pollen to the relative abundances of floral taxa available in the landscape. When pollen diet composition was compared to floral availability, most flowering taxa were not visited in similar proportions to their abundance in the landscape, so we suggest quality may be more important than quantity for bumblebees foraging in landscapes around *Vaccinium* crops.

4.5.1 Pollen preferences and temporal shifts

The sampling period had a significant effect on the pollen taxon composition, indicating a shift in diet throughout the crop flowering period. Phenological progression, the succession in which flowers bloom, produces a shift in the resource landscape that reflects pollinator diets (Timberlake et al., 2019; Wood et al., 2018a). *Vaccinium* was a major component of the pollen diet throughout the sampling period, but less than half of all pollen reads were assigned to the taxon. *Salix caprea* dominated the pollen diet at all four farms in the early sample collection period, when floral richness in the landscape was at its lowest. *S. caprea* was the only plant species other than *Vaccinium* that comprised a major component of the pollen diet (>45% of assigned reads) in over 10% of the pollen samples. Floral transects in the final sampling period did not identify any *S. caprea* plants, indicating that foraging bees were collecting pollen from *S. caprea* that were scarce in the landscape, further than the 1 km radius from the centre of the farm, or located on private land that we were not able to survey. *Salix* sp. have been identified as important early spring pollen resources for multiple

bee species, indicating its attractiveness to pollinators at a time of year when there may be few other floral resources (Bertrand et al., 2019; Kämper et al., 2016; Wood et al., 2018).

There is evidence of woody plants and trees dominating the pollen diet of bumblebees in early spring, which supports our findings of *Salix caprea*, *Prunus spinosa*, *Ulex europaeus* and *Pyrus communis* being some of the most frequently collected pollen in March and April (Bertrand et al., 2019; Kämper et al., 2016). Similar patterns have been found for other bee species at this time of year, with *Salix, Ulex* and *Prunus* dominating the pollen diet of honey bees (De Vere et al., 2017; Lowe et al., 2022) and solitary bees (Wood et al., 2018). In the later spring, there were more herbaceous plants in the diet, such as *Ballota nigra*, which has been previously identified as attractive forage (Carvell, Roy, et al., 2006), as well as *Hyacinthoides non-scripta, Ficaria verna* and *Ranunculus repens*.

Bees are known to display floral constancy, whereby one plant taxa is predominantly visited on a single foraging trip (Heinrich, 1976, Lazaro & Totland, 2010). Our results of few taxa making up the major components (>45%) of the pollen diet agree with the foraging constancy theory (Figure 4.5). Floral constancy has been found to occur in areas of high floral abundance, high floral richness or short distances to and among patches of floral resources (Chittka et al., 1997; Grüter & Ratnieks, 2011). There was a higher diversity of taxa in the minor and intermediate categories, suggesting bees supplement their diets with small amounts of secondary taxa (Jones et al., 2021; Kleijn & Raemakers, 2008; Leonhardt & Blüthgen, 2012). By collecting additional pollen resources at low quantities bumblebees are able to track available resources in the landscape (Heinrich, 1976). A varied diet can serve to balancing the nutrition of the colony, which has been known to increase the immunocompetence and fitness of the brood (Vaudo et al., 2015).

Few plant taxa identified in the pollen loads were collected in proportion to their floral abundance in the landscape. Although *Vaccinium* was continuously the most abundant flower in the floral surveys, the pollen reads did not consistently reflect this pattern. In the two early collection periods, *Vaccinium* pollen was collected less than predicted based on its abundance in the landscape, while in the latter two sampling periods it was collected more than predicted. *Salix caprea* was the only taxon to be collected significantly more than

predicted based on its abundance in the landscape throughout the sampling period. Other taxa were preferred to a lesser degree or not consistently over the crop flowering, including *Prunus spinosa, Pyrus communis* and *Ballota nigra*. Some taxa were collected less than predicted by the model, including *Fragaria, Taraxacum officinale*, *Hyacinthoides non-scripta, Heracleum sphondylium, Ficaria verna* and *Prunus avium*. *Taraxacum* is a herbaceous flower often found in flower margins, and is highly rewarding in nectar production, but produces low quality pollen (Génissel et al., 2002; Hicks et al., 2016). Bumblebees visit flowers based on nectar or pollen rewards, so it is possible pollen was picked up during a nectar foraging bout, which is why lower proportions of assigned reads were recorded than the model predicted. Other studies show strong preferences of pollinators and selective use of available resources in urban ecosystems, which corroborates our suggestions that a few resources are used to a high degree (de Vere et al., 2017; Lowenstein et al., 2019). *Crataegus monogyna* (744.8 Mbp) was collected less than was predicted using the model. *C. monogyna* produces a low number of pollen grains per flower (Fowler et al., 2016), and a high density of mass-blooming flowers, such as *Vaccinium* or *Salix* would have been an attractive forage source close to the colonies, which could in part explain the low levels of *C. monogyna* RevMet reads in the results. The genome sizes may have had an impact on the RevMet results, as discussed in Chapter 3. We might expect the plants with small genome sizes, such as *Salix* (392 Mbp), to be under-reperesented and plants with large genomes, such as *Hyacinthoides* (20776 Mbp) to be over-represented in the RevMet results. However, in comparison to the landscape level surveys neither of these patterns were observed (figure 4.6).

Strawberry (*Fragaria* x *ananassa*) was the only other commercial soft fruit crop co-flowering on the same farms as *Vaccinium*. Of the total proportion of assigned reads, only 2% were *Fragaria*, which is surprising given the locality of the *Fragaria* crops. A low collection of *Fragaria* pollen by bumblebees in agro-ecosystems has been noted elsewhere, which suggests they are not an attractive resource to this genus (Bänsch et al., 2020; Bontšutšnaja et al., 2021; Foulis & Goulson, 2014). Interestingly, nocturnal pollination has been found to be important in increasing *Fragaria* yield quality, which suggests bumblebees may not be the most effective pollinator group (Fijen et al., 2023).

Pollen foraging by bumblebees is known to be affected by several factors, including the nutritional value of pollen grains. Bumblebees have been found to balance the nutritional needs of the colony by selecting pollen with a high protein content when given the choice (Vaudo, Patch, et al., 2016), which has also been shown in honeybees (Requier et al., 2015). The content of protein in *Vaccinium* pollen (13.9%) is low relative to other flower pollen, which could explain the high proportion of non-crop pollen in the diet (Somerville, 2001). The crude protein content of *Salix* (~20%) and members of the Rosaceae family (24.4%) are significantly higher than *Vaccinium*, which might have driven the high utilisation of these taxa (Roulston & Cane, 2000). Other nutritional qualities have also been found to affect pollen foraging and bumblebee health, including lipid concentration, amino acid ratios and sterol content (Archer et al., 2021; Vanderplanck et al., 2020; Vaudo, Patch, et al., 2016; Vaudo, Stabler, et al., 2016).

4.5.2 Pollen richness

The pollen richness of individual pollen loads did not vary over time, with bees visiting on average between one and three plants on a single foraging trip. However, there were observed differences at the colony-level, in that the pollen richness decreased when there was higher *Vaccinium* flower cover. Locally abundant *Vaccinium* flowers were likely driving the lower colony richness when there was a higher crop cover, which has also been observed in *Fragaria* crops (Bänsch et al., 2020). Colony-level pollen richness increased with landscape floral richness and over the *Vaccinium* flowering period. This result suggests that in landscapes of higher floral richness, bumblebees have a wider diet breadth, providing evidence of their generalism as foragers (Leonhardt & Blüthgen, 2012).

4.5.3 Vaccinium foraging

We would expect foraging bees to visit the most abundant plant in close proximity to reduce energy costs (Lihoreau et al., 2017; Westphal et al., 2006). The number of *Vaccinium* pollen reads collected was positively correlated to the *Vaccinium* flower cover, which suggests a larger bloom of flowers is more likely to attract bumblebee visits. The interaction of

Vaccinium cover with sampling period showed a decrease in *Vaccinium* pollen read proportion in the last sampling period, which was likely due to it coinciding with the end of crop bloom and therefore a lower abundance of flowers and the increase in richness of alternative floral resources in the landscape. There were later flowering *Vaccinium* crops present at Tuesley farm and Heathlands farm in open fields (as opposed to those grown in polytunnels) which started flowering from mid-May, not reaching peak bloom until approximately one month later, which were recorded in the flower transect survey data for this data collection period.

Mass-flowering crops (MFCs) represent a huge burst of floral resource, which bumblebees are likely to exploit, as has been found in other agro-ecosystems (Hemberger & Gratton, 2018; Rollin et al., 2013). Foraging efficiency is important in bumblebees, and they have been known to maximise pollen intake by collecting pollen from the most abundant resource, as handling time reduces with experience (Kämper et al., 2016). There is likely a trade-off between pollen quality and quantity, and the bumblebees in our study could be maximising pollen quantity over quality.

Vaccinium pollen was collected by bumblebees more than any other plant taxa over the sampling period. However, the overall proportion was relatively low at 45.6%. Our results show similar proportions of *Vaccinium* pollen collected by bumblebees to Toshack & Elle (2019), but we found a higher proportion of *Vaccinium* pollen than other studies that found less than 1% collected by honey bees (Colwell et al., 2017) or a range of 15 – 52% by three species of managed bee (Bobiwash et al., 2018). Although *Vaccinium* crops represent a large foraging resource, the poor quality of pollen probably explains why bees supplement their diet from other floral taxa.

The quantity of *Vaccinium* pollen collected by bees provides information about the attractiveness of the flowers to managed pollinators, which is important in pollinator-dependent agro-ecosystems. Kendall et al., (2020) found *Vaccinium* to display a high level of self-compatibility, and >90% probability of fruit set was achieved in three to five pollinator visits. In our system, the high proportion of *Vaccinium* pollen in our samples infers a high level of floral constancy exhibited by the workers. In a crop with high self-compatibility these

results could suggest adequate pollination services are being provided by the commercial bees, although we have not set out to explore that question here.

4.5.4 RevMet

Here, we demonstrate for the first time RevMet's suitability as a method of characterising a large number of pollen loads from individual bumblebees in an agro-ecological landscape. One advantage of WGS approaches such as RevMet over metabarcoding is its lack of PCR amplification and associated biases in quantification of sequence read abundance. Quantification of plant taxa is a crucial step in characterising the diet, because not all plants are visited in equal abundance on a single foraging trip, as shown by our results. Metabarcoding studies using presence-absence analyses alone to overcome the unreliable quantification problem, risk over-representing plant taxa that make up a minor component of the diet (Bell et al., 2019; Deagle et al., 2019).

The presence of *Salix* in the pollen collected in sampling periods where none was identified in the landscape provides evidence of how pollen studies reveal networks that might have been missed using standard observational techniques. Woody and tree taxa can be overlooked in importance due to the difficulty in surveying these habitats (Allen & Davies, 2023), but by using a DNA approach we are provided with a more complete picture of the foraging resources (Arstingstall et al., 2021; Pornon et al., 2017).

We suggest that RevMet and nanopore sequencing provide a low-cost and portable alternative to metabarcoding studies using the Illumina sequencing platform. Recently released kits from ONT can multiplex 96 samples (in our study, individual bee pollen loads) on a single flow cell, at a cost of ~£20 per sample, including DNA extraction, reagents and ONT MinION flow cells. An additional advantage of ONT's MinION devices is their portability, as they can be used outside of the lab and in remote fieldwork locations (Castro-Wallace et al., 2016; Pomerantz et al., 2018). In our study we created our own genome reference skims because there is a lack of publicly available WGS references for plant taxa, although there are continued efforts to create a global reference library (Lewin et al., 2018).

4.5.5 Limitations

There were several limitations to this study. First, our reference database contained 108 species of flowering plant taxa (details on plant reference database method in Chapter Two) which does not include all possible flowering taxa the bees may have visited. This means there may have been a degree of false negative identifications, as plant taxa could have been missed from analyses. At the time of data collection (March 2021) there were COVID-19 restrictions in place that meant we were unable to survey private gardens, therefore garden plants that were in the 1 km radius of the colonies could not be included in our reference database. There is evidence of garden plants providing important floral resources, so the richness of the pollen diets may have been underestimated in our results (Baldock et al., 2019; Bertrand et al., 2019; Hicks et al., 2016).

Second, the null model assumes all flowers are equally attractive to foraging bees – i.e. pollen resource quality is not taken into account. The number of flower units is the resource measurement used in the null model, but other factors could better explain the foraging behaviour, such as pollen quantity (Hicks et al., 2016) or nutritional value of the pollen (Vaudo et al., 2020). This information is not currently available for the range of flowering taxa recorded in this study, but would be an interesting next step in this research.

Third, the preference analysis also assumed the floral resources were equally distributed in the 1 km radius, and therefore had an equal chance of being visited. However, there's evidence that flower patches and mass-flowering crops closer to the colony have a higher likelihood of being exploited (Westphal et al., 2006). Therefore, it is likely that the model over-estimated the diet breadth of workers, because they were likely to restrict their foraging efforts in the presence of flowering *Vaccinium* crops.

Fourth, the RevMet results had a 1% minimum abundance filter applied throughout this thesis in order to remove plant taxa present at low abundances. This percentage has been used in eDNA and pollen DNA sequencing studies (Deagle et al., 2019; Peel et al., 2019).

Plant taxa present at <1% in the pollen loads was unlikely to be an important source of forage in a bee's diet, since it is estimated that bees visit 2-3 flower taxa on a single foraging trip (Peel et al., 2019; Yourstone et al., 2023). Without the 1% minimum abundance filter, the species richness of the pollen loads would have been abnormally high, and included plant taxa in the reference library that were not flowering at the time of sample collection. However, it would be preferable to calculate a minimum abundance threshold based on mock communities and negative controls (Drake et al., 2021).

Finally, we calculated the quantitative relationship of RevMet reads to the relative abundance of pollen grains in Chapter Three and found there to be an effect of genome size on the quantitative abilities of RevMet. This relationship became stronger once a genome-correction factor had been applied to the RevMet relative abundances. In Chapter Four we used the raw RevMet reads without the application of a genome correction factor because we feel this step needs further testing. The over- or underestimation in relative abundance occurs on a per sample basis because we sequenced individual bee corbicular loads rather than pooling samples together. Therefore, the probability of quantification bias depends on the taxa present in a single sample. We examined the abundance status of different taxa and categorised them into major, intermediate or minor pollen categories (Figure 6). If genome size was distorting our results we might expect *Ficaria verna* and *Hyacinthoides non-scripta,* two taxa in the top ten most utilised species which have large genomes relative to the other taxa, to frequently occur in the "major" category (>45% of assigned reads). This effect was not observed in our results as these taxa were predominantly in the "minor" category (<10% of assigned reads) although it is possible that their importance has still been overestimated. The results in our null model also predict *F. verna* and *H. non-scripta* to have been collected less than predicted based on their floral relative abundance in the landscape in some of the sample periods. If there was significant overestimation of these taxa, we might expect to have results that suggest bees were preferentially visiting these taxa in the landscape. Conversely, *Salix caprea* was underestimated in Chapter Three mock mixtures, so it could be even more prevalent than our results suggest. Regardless of quantitative bias, we still feel that the quantitative data used here is more informative than a binary presence/absence approach of taxa present in the pollen loads, which would inflate the

importance of all taxa present at low proportions. We would recommend further investigation to measure the potential quantitative bias in these samples.

4.5.6 Management recommendations

Agricultural intensification and associated habitat loss are major concerns in pollinator conservation, and creating habitats on farms that provide foraging and nesting resources for insects has the potential to improve the pollination services for a range of crops. Although we have not included analysis on the quality of pollen sources, we have information on the most commonly used plant taxa, which infers a level of attractiveness to bumblebees.

A high proportion of the taxa used in early spring were woody and hedgerow species, in particular *Salix caprea* and *Prunus spinosa*. Managing the landscape for wild pollinators should involve ensuring hedgerow species (*S. caprea* in particular) are present in every 1 km radius, and that they are managed to allow for maximum flowering potential, e.g. infrequent cutting. Herbaceous flowers were also used, such as *Ballota nigra, Ficaria verna, Ranunculus* sp. and *Hyacinthoides non-scripta*. These taxa can be found in flower margins and areas of semi-natural habitats on the farms, so we recommend allowing margins to grow and reducing the mowing regimes (Whittington et al., 2003; Brittain et al., 2022).

Here, we used commercial colonies to investigate *Bombus terrestris* foraging in early spring, a time of year when wild queen bumblebees have emerged from hibernation and are foraging and nest building. The queens need access to a supply of pollen and nectar during this time, otherwise resource scarcity has a negative impact on colony growth and development (Rotheray et al., 2017). March has also been identified as a gap in available resources in farmland, so providing high quality pollen sources in the landscape at this time of the year is an important step in pollinator conservation (Timberlake et al., 2019).

This knowledge builds on studies that have identified important sources of pollen in non-agricultural landscapes (Lowe et al., 2022) and nectar resources in farm landscapes (Timberlake et al., 2019) and gardens (Tew et al., 2021). Timing of resources is an important

consideration, and careful planning should be included in agri-environment schemes to recommend plant taxa that include woody flowering species.

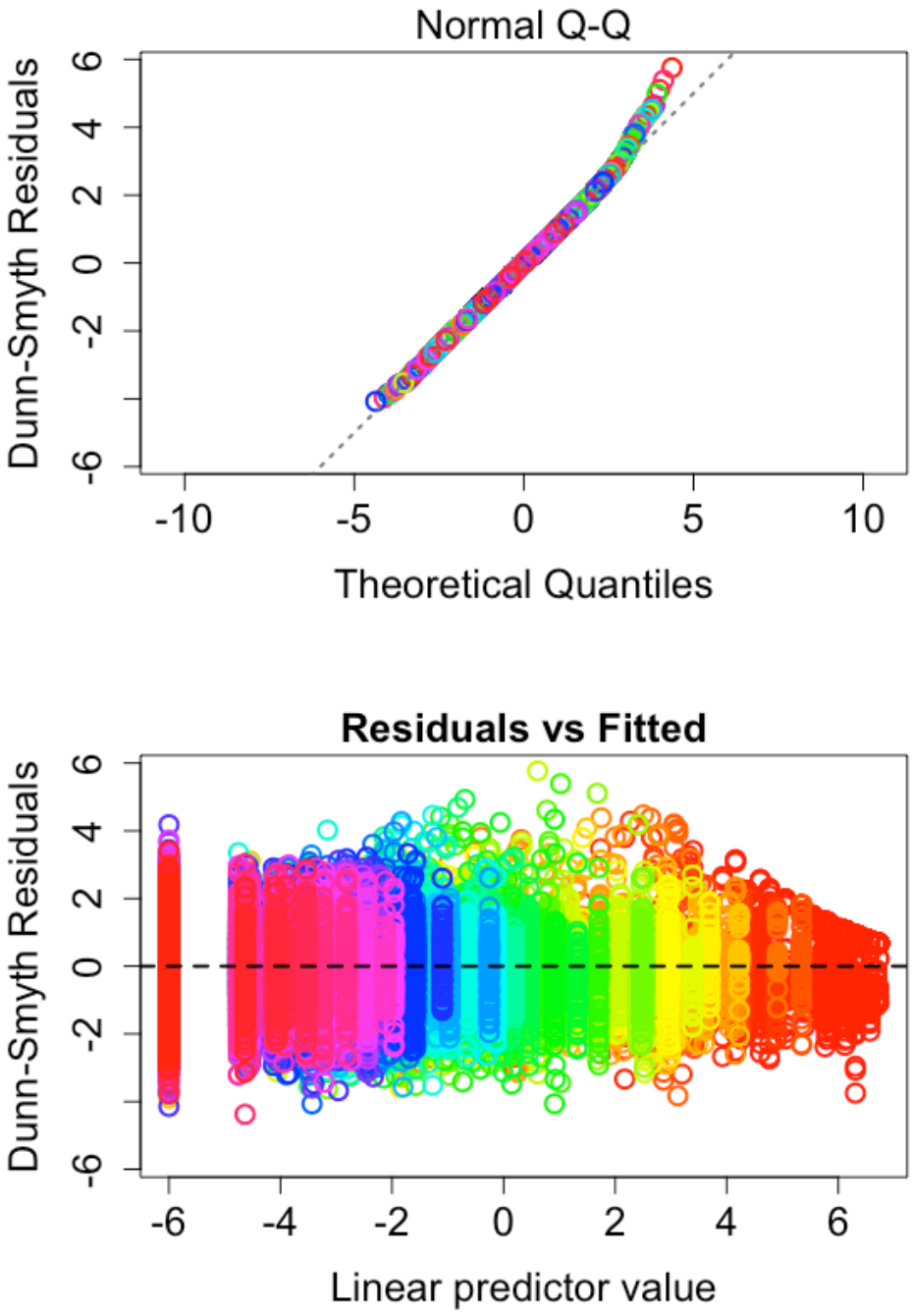
4.6 Conclusions

Our study suggests that commercial bees situated within mass-flowering *Vaccinium* crops collect a large proportion of pollen from the crop itself, but also forage from a range of flowering plants in the landscape, with a particular preference for woody and hedgerow species. The seasonal pattern of the diet depends on flowering phenology, as the taxa present in the pollen diets and the flowering community of plants, shift together over the crop flowering period. Our results show that even when located in the centre of a mass-flowering crop, bumblebee workers will forage off the crop, likely due to their requirement for a diverse pollen diet. Therefore, we suggest farms growing soft-fruit crops maintain areas of semi-natural habitat to support wild and commercial pollinators that will fulfil the pollination requirements of the crop.

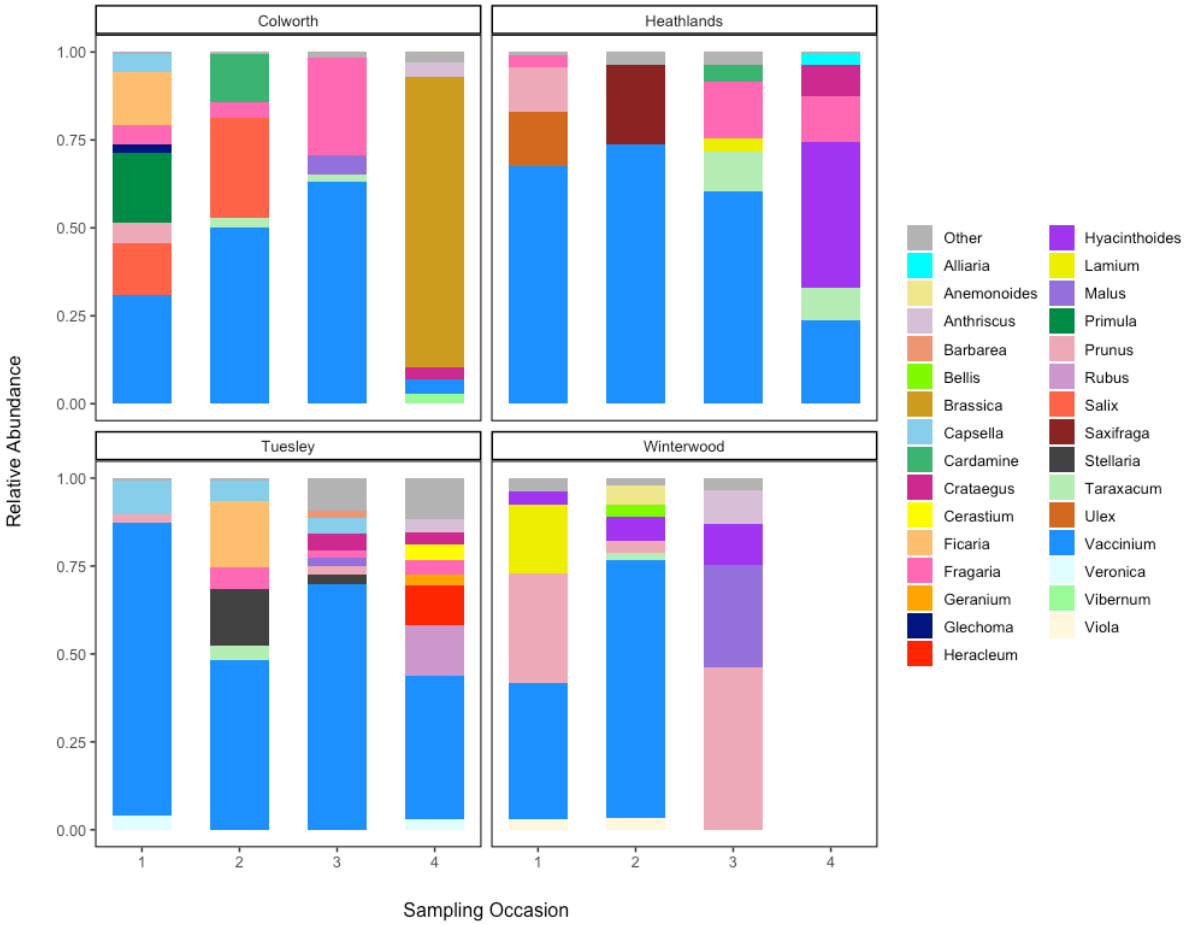
This study is the first to use a nanopore sequencing approach to characterise the pollen diets of a large sample of commercial bees. One strength of our approach is our self-curated database of WGS skims, but for wider application this needs development as it currently contains a limited number of plant taxa genomes. We advocate the use of RevMet as a new molecular approach in characterising pollen diets, which can be used to aid and inform pollinator conservation.

4.7 Supplementary Material

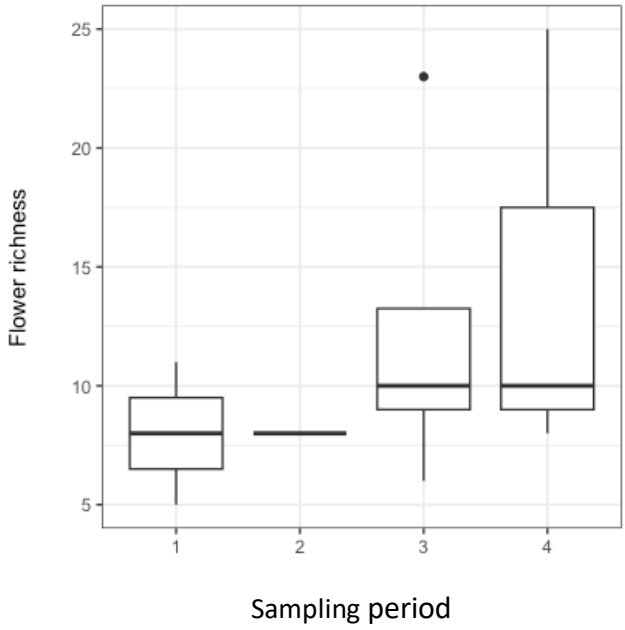
Supplementary Figure 4.1 A QQ plot (top) and residuals plot (bottom) showing a strong relationship between the proportion of RevMet assigned reads and the variance. Family used is negative binomial. Created using the meanvar.plot function from the mvabund package in R.



Supplementary Figure 4.2 Stacked bar charts showing proportional abundance of floral taxa at each survey period and farm. The sample periods started in late March and were repeated approximately every 2 weeks until the end of May. The "other" category includes taxa present at < 2% abundance. Note that Winterwood Farm had only three sampling periods.

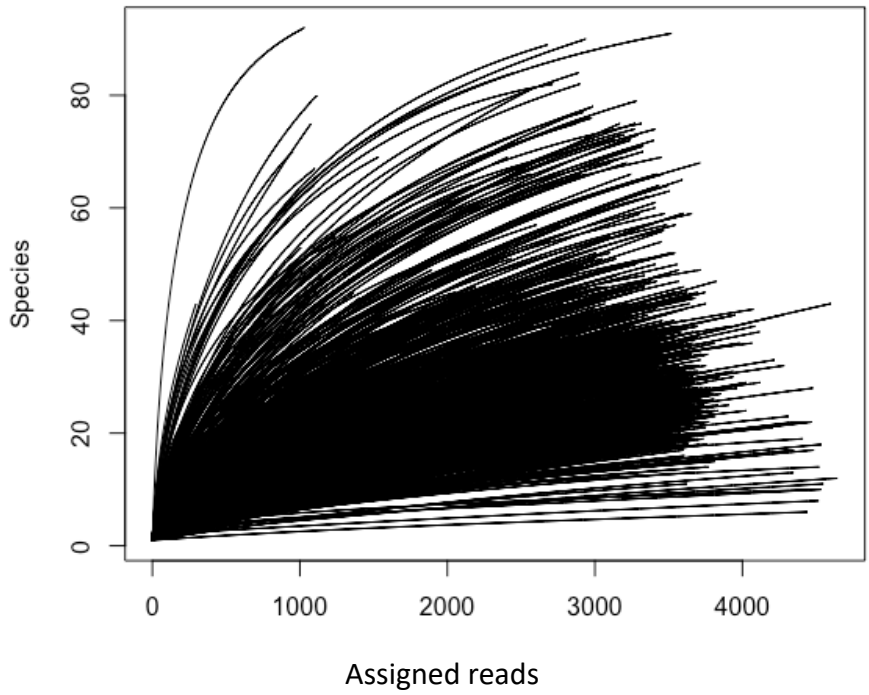


Supplementary Figure 4.3 Floral species richness across the four sampling occasions (S1 – S4). Data was collected in flower transect surveys conducted approximately every 2 weeks at each farm, which have been grouped into sampling periods. Sampling period two had a floral richness of eight taxa, which was the same on all farms.

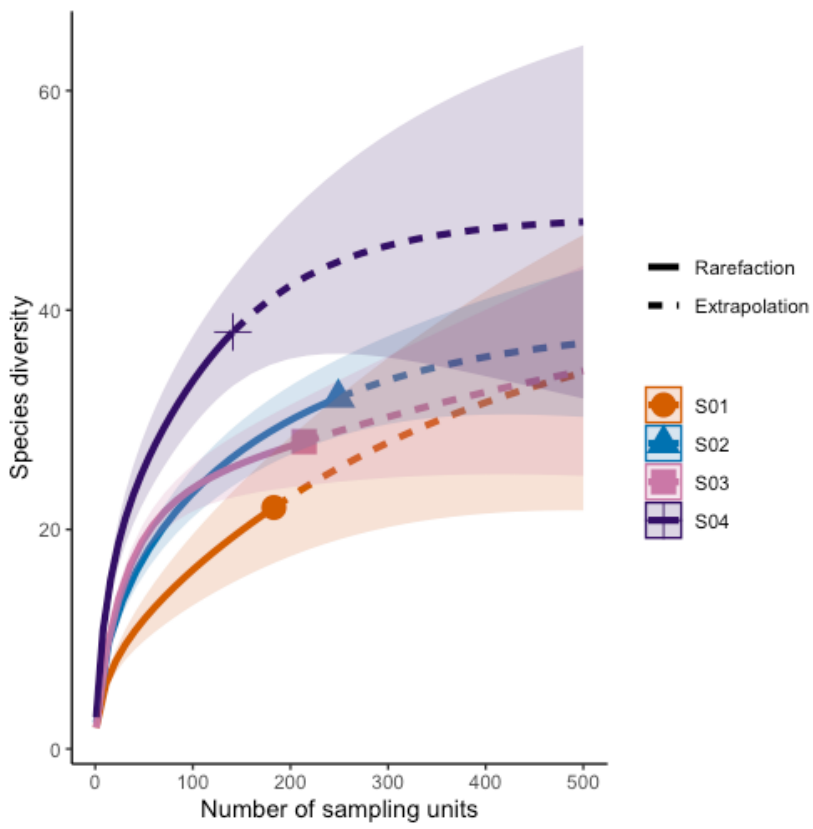


Supplementary Figure 4.4 a) rarefaction curve for all samples. Samples that did not reach asymptote were removed from further analyses (n=53), b) species richness curves for the four sampling occasions (SO1 – SO4), calculated using the iNEXT package.

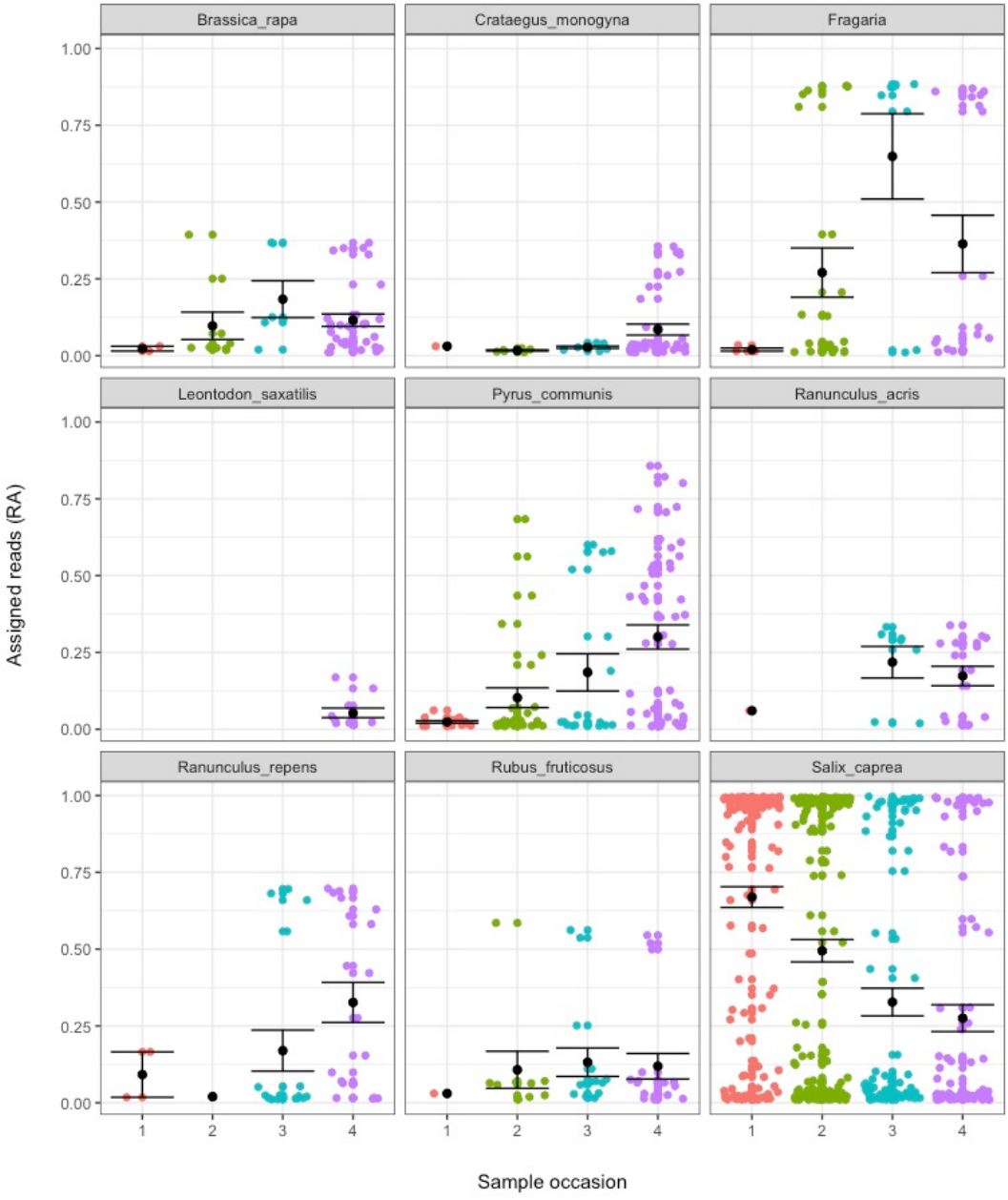
a)



b)



Supplementary Figure 4.5 Plant taxa identified as having different abundances over the *Vaccinium* crop flowering periods (Sampling periods 1-4). Univariate analyses calculated using *mvabund.*



Supplementary Table 4.1 The sampling periods and dates on which floral transects and pollen collection took place

Farm	Sampling period	Flower transect date	Pollen collection date
Heathlands	1	29/03/2021	30/03/2021
Heathlands	2	13/04/2021	14/04/2021
Heathlands	3	30/04/2021	27/04/2021
Heathlands	4	10/05/2021	11/05/2021
Tuesley	1	06/04/2021	04/04/2021
Tuesley	2	21/04/2021	22/04/2021
Tuesley	3	18/05/2021	18/05/2021
Tuesley	4	01/06/2021	31/05/2021
Colworth	1	01/04/2021	01/04/2021
Colworth	2	16/04/2021	18/04/2021
Colworth	3	29/04/2021	29/04/2021
Colworth	4	12/05/2021	12/05/2021
Winterwood	1	09/04/2021	09/04/2021
Winterwood	2	20/04/2021	19/04/2021
Winterwood	3	06/05/2021	07/05/2021

Supplementary Table 4.2 A list of the flowering plant taxa found at each farm and sampling period.

Farm	Family	Species	S1	S2	S3	S4
Colworth	Adoxaceae	Vibernum lantana	N	N	Y	Y
	Amaryllidaceae	Narcissus pseudomonas	Y	N	N	N
	Apiaceae	Anthriscus sylvestris	N	N	N	Y
	Asteraceae	Bellis perennis	N	N	N	Y
		Taraxacum	Y	Y	Y	Y
	Brassicaceae	Brassica napus	N	N	N	Y
		Capsella bursapastoris	Y	N	N	N
		Cardamine flexuosa	N	Y	N	N
	Ericaceae	Vaccinium corymbosum	Y	Y	Y	Y
	Lamiaceae	Glechoma hederacea	Y	N	N	Y
	Plantaginaceae	Veronica persica	Y	Y	N	N
	Primulaceae	Primula vulgaris	Y	N	N	N
	Ranunculaceae	Ficaria verna	Y	Y	N	N
	Rosaceae	Crataegus monogyna	N	N	Y	Y
		Fragaria	Y	Y	Y	Y
		Malus sylvestris	N	N	Y	Y
		Prunus spinosa	Y	Y	N	N
	Salicaceae	Salix caprea	Y	Y	N	N
Heathlands	Adoxaceae	Vibernum lantana	N	N	Y	N
	Apiaceae	Anthriscus sylvestris	N	N	N	N
	Asparagaceae	Hyacinthoides non scripta	N	N	N	Y
	Asteraceae	Taraxacum	Y	Y	Y	Y
	Boraginaceae	Myosotis sylvatica	N	N	N	N
	Brassicaceae	Alliaria petiolata	N	N	N	Y
		Cardamine impatiens	N	N	Y	N
	Caryophyllaceae	Stellaria media	N	Y	N	N
	Ericaceae	Vaccinium corymbosum	Y	Y	Y	Y
	Fabaceae	Trifolium repens	N	N	N	N
		Ulex europaeus	Y	N	N	N
	Geraniaceae	Geranium robertianum	N	N	N	N
		Geranium sanguineum	N	N	N	N
	Lamiaceae	Glechoma hederacea	N	N	Y	N
		Lamium album	N	N	N	Y
		Lamium purpureum	N	Y	Y	N
	Plantaginaceae	Veronica persica	N	N	Y	N
	Ranunculaceae	Ranunculus repens	N	N	N	N
	Rosaceae	Crataegus monogyna	N	Y	Y	Y
		Fragaria	Y	N	Y	Y
		Malus sylvestris	N	N	Y	Y
		Prunus spinosa	Y	Y	N	N
		Rubus fruticosus	N	N	N	N
		Rubus idaeus	N	N	N	N
	Salicaceae	Salix caprea	N	Y	N	N
	Saxifragaceae	Saxifraga granulata	N	Y	N	N

Farm	Family	Species	S1	S2	S3	S4
Tuesley	Adoxaceae	Vibernum opulus	N	N	N	Y
	Apiaceae	Anthriscus sylvestris	N	N	N	Y
		Heracleum sphondylium	N	N	N	Y
	Asparagaceae	Hyacinthoides non scripta	N	Y	Y	Y
	Asteraceae	Cerastium fontanum	N	N	N	Y
		Crepis tectorum	N	N	N	Y
		Leontodon taraxacoides	N	N	N	Y
		Taraxacum	Y	Y	Y	Y
	Boraginaceae	Anchusa arvensis	N	N	Y	N
		Borago officinalis	N	N	Y	N
	Brassicaceae	Arabidopsis thaliana	N	N	Y	N
		Barbarea vulgaris	N	N	Y	N
		Capsella bursapastoris	Y	Y	Y	N
		Cardamine flexuosa	N	N	Y	N
	Caryophyllaceae	Silene dioica	N	N	N	Y
		Stellaria graminea	N	N	Y	Y
		Stellaria holostea	N	N	Y	N
		Stellaria media	N	Y	N	N
	Ericaceae	Vaccinium corymbosum	Y	Y	Y	Y
	Fabaceae	Trifolium repens	N	N	N	Y
		Vicia sativa	N	N	Y	Y
	Geraniaceae	Geranium dissectum	N	N	N	Y
		Geranium robertianum	N	N	N	Y
	Lamiaceae	Glechoma hederacea	N	N	Y	Y
		Lamium album	N	N	Y	Y
	Papaveraceae	Papaver rhoeas	N	N	Y	Y
	Plantaginaceae	Plantago lanceolata	N	N	N	Y
		Veronica chamaedrys	N	N	N	Y
		Veronica persica	Y	N	Y	N
	Ranunculaceae	Caltha palustris	Y	N	Y	N
		Ficaria verna	N	Y	N	N
		Ranunculus repens	N	N	Y	Y
	Rosaceae	Crataegus monogyna	N	N	Y	Y
		Fragaria	N	Y	Y	Y
		Malus sylvestris	N	N	Y	N
		Myosotis arvensis	N	N	N	Y
		Prunus avium	N	N	Y	N
		Prunus spinosa	Y	Y	N	N
		Rubus idaeus	N	N	N	Y
	Salicaceae	Salix caprea	Y	N	N	N
	Violaceae	Viola arvensis	N	N	Y	N
Winterwood	Apiaceae	Anthriscus sylvestris	N	N	Y	
	Asparagaceae	Hyacinthoides non scripta	Y	Y	Y	
	Asteraceae	Bellis perennis	Y	Y	N	
		Taraxacum	Y	Y	Y	
	Boraginaceae	Borago officinalis	N	Y	N	
		Pentaglottis sempervirens	N	N	Y	
	Ericaceae	Vaccinium corymbosum	Y	Y	N	
	Lamiaceae	Lamium purpureum	Y	N	Y	
	Plantaginaceae	Veronica agrestis	Y	N	N	
	Primulaceae	Primula vulgaris	N	N	Y	
	Ranunculaceae	Anemonoides nemorosa	N	Y	Y	
	Rosaceae	Malus sylvestris	N	N	Y	
		Prunus avium	Y	N	Y	
		Prunus spinosa	Y	Y	Y	
	Violaceae	Viola riviniana	Y	Y	N	

Supplementary Table 4.3 Proportion of plant taxa in pollen samples, assigned using RevMet. Proportions are assigned by sampling period and farm and ordered by highest number of assigned reads.

Family	Taxa	Colworth				Heathlands				Tuesley				Winterwood			Total no. assigned reads	Total %
		S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3		
Ericaceae	Vaccinium_corymbosum	32.86	62.28	81.32	80.48	12.87	26.21	39.52	31.74	53.53	39.55	56.09	20.33	19.89	66.88	54.70	1087227	45.5
Salicaceae	Salix_caprea	56.66	24.52	6.70	7.76	80.02	55.74	31.52	23.59	29.91	22.18	5.86	2.12	54.85	19.92	6.14	731913	30.6
Rosaceae	Prunus_spinosa	4.96	3.79	0.79	0.16	2.10	12.85	2.34	0.88	9.18	10.45	0.47	0.44	13.18	10.01	6.55	124119	5.2
Rosaceae	Pyrus_communis	0.12	0.10	0.15	3.28	0.34	0.30	2.54	12.82	0.56	4.22	3.35	14.19	0.16	0.19	0.56	54718	2.3
Rosaceae	Fragaria_ananassa	0.01	0.68	6.91	0.04	0.11	1.37	1.57	6.71	0.29	6.23	0.10	7.14	0.00	0.02	0.06	47620	2.0
Lamiaceae	Ballota_nigra	0.00	0.00	0.00	0.00	0.01	0.00	3.41	1.62	0.04	0.04	0.01	0.02	0.00	0.01	20.32	47178	2.0
Ranunculaceae	Ficaria_verna	2.11	2.19	1.52	0.07	0.15	0.15	0.08	0.16	0.57	3.37	2.25	0.35	4.33	0.62	0.07	38005	1.6
Ranunculaceae	Ranunculus_repens	0.01	0.07	0.01	0.01	0.01	0.00	0.01	1.10	0.54	0.05	7.96	18.58	0.01	0.01	0.02	32621	1.4
Fabaceae	Ulex_europus	0.10	0.05	0.05	0.05	2.77	0.10	3.20	1.65	0.31	1.88	1.21	0.08	0.04	0.06	0.15	23294	1.0
Asparagaceae	Hyacinthoides_non_scripta	0.02	0.01	0.02	0.03	0.07	0.02	0.03	6.15	0.16	0.89	0.09	3.27	0.04	0.03	0.05	18695	0.8
Rosaceae	Rubus_fruticosus	0.03	1.03	0.60	0.06	0.03	0.15	0.16	0.58	0.20	0.60	4.32	6.49	0.04	0.05	0.13	16274	0.7
Ranunculaceae	Ranunculus_acris	0.00	0.03	0.01	0.01	0.01	0.00	0.00	0.47	0.20	0.02	4.43	8.70	0.01	0.00	0.01	16116	0.7
Asteraceae	Taraxacum_officinale	0.03	2.94	0.14	0.11	0.16	0.24	0.05	0.29	0.04	0.71	5.18	0.10	0.46	0.29	0.31	16036	0.7
Brassicaceae	Brassica_rapa	0.02	0.02	0.03	0.99	0.10	0.07	1.27	3.45	0.12	1.26	0.64	2.39	0.01	0.06	0.27	15880	0.7
Papaveraceae	Papaver_somniferum	0.01	0.00	0.01	0.00	0.00	0.00	4.28	0.01	0.02	0.02	0.00	0.00	0.01	0.01	2.05	12501	0.5
Fabaceae	Vicia_sativa	0.02	0.00	0.01	0.01	0.00	0.00	3.88	0.01	0.01	0.02	0.01	0.00	0.00	0.00	1.90	11373	0.5
Rosaceae	Rubus_ideaus	0.00	0.62	0.11	0.02	0.00	0.03	0.03	0.09	0.14	0.12	3.12	4.39	0.00	0.01	0.01	9076	0.4
Rosaceae	Crataegus_monogyna	0.02	0.01	0.03	1.57	0.10	0.03	0.11	1.88	0.05	0.20	0.46	4.08	0.01	0.04	0.05	8598	0.4
Rosaceae	Agrimonium_eupatoria	0.02	0.04	0.13	0.10	0.01	0.04	0.07	0.15	0.04	0.13	0.09	0.12	0.03	0.04	2.74	5886	0.2
Brassicaceae	Hirschfeldia_incarna			0.00	0.27	0.00	0.00	1.47	1.32	0.01	0.07		0.00	0.00	0.00	0.03	5415	0.2
Lamiaceae	Lamium_album	0.17	0.05	0.03	0.01	0.04	0.01	0.02	1.20	0.06	0.06	0.27	0.40	0.86	0.01	0.02	5182	0.2
Ranunculaceae	Caltha_palustris	0.07	0.14	0.12	0.16	0.11	0.10	0.08	0.24	0.74	0.97	0.56	0.28	0.12	0.08	0.20	4902	0.2
Boraginaceae	Symphyton_officinale	0.02	0.02	0.03	0.01	0.02	0.02	0.01	0.09	0.10	0.14	0.02	0.03	2.46	0.14	0.03	4260	0.2
Lamiaceae	Lamium_purpuruem	1.21	0.24	0.01	0.01	0.01	0.00	0.00	0.24	0.02	0.02	0.01	0.00	0.16	0.02	0.01	3144	0.1
Adoxaceae	Vibernum_lantana	0.02	0.04	0.06	0.08	0.01	0.03	0.03	0.10	0.16	0.20	0.24	0.06	1.27	0.04	0.10	2784	0.1
Rosaceae	Malus_sylvestris	0.01	0.00	0.00	0.05	0.00	0.00	0.00	0.01	0.02	1.83	0.00	0.01	0.02	0.00	0.01	2757	0.1
Asteraceae	Achillea_millefolium	0.06	0.09	0.16	0.37	0.06	0.06	0.07	0.14	0.10	0.16	0.37	0.16	0.08	0.09	0.07	2489	0.1
Fabaceae	Trifolium_repens	0.02	0.00	0.02	1.56	0.05	0.00	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	2180	0.1
Apiaceae	Anthriscus_sylvestris	0.03	0.04	0.03	0.06	0.05	0.06	0.04	0.33	0.15	0.19	0.43	0.11	0.08	0.07	0.06	1889	0.1
Brassicaceae	Barbaria_vulgaris	0.00	0.00	0.00	0.00	0.01	0.00	0.00	1.25	0.01	0.02	0.00	0.00	0.02	0.00	0.01	1775	0.1
Violaceae	Viola_riviniana	0.04	0.03	0.03	0.02	0.07	0.06	0.03	0.05	0.10	0.43	0.04	0.05	0.05	0.03	0.04	1420	0.1
Asteraceae	Sonchus_arvensis	0.02	0.09	0.05	0.04	0.02	0.03	0.02	0.03	0.17	0.21	0.20	0.02	0.02	0.02	0.07	1386	0.1
Malvaceae	Malva_sylvestris	0.04	0.06	0.04	0.05	0.05	0.04	0.04	0.07	0.14	0.16	0.14	0.14	0.04	0.03	0.08	1348	0.1

Adoxaceae	Sambucus_nigra	0.02	0.02	0.04	0.05	0.01	0.03	0.03	0.07	0.15	0.19	0.16	0.06	0.05	0.04	0.11	1216	0.1
Asteraceae	Leontodon_saxatilis	0.03	0.00	0.01	0.03	0.00	0.01	0.00	0.03	0.02	0.03	0.05	1.82	0.01	0.01	0.01	1114	0.0
Plantaginaceae	Plantago_lanceolata	0.01	0.01	0.04	0.05	0.00	0.01	0.01	0.02	0.02	0.02	0.03	2.71	0.01	0.01	0.01	1099	0.0
Lamiaceae	Glechoma_hederacea	0.01	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.04	0.33	0.02	0.01	0.29	0.02	0.03	1075	0.0
Amaryllidaceae	Narcissus_pseudomonas	0.02	0.01	0.01	0.01	0.06	0.02	0.02	0.04	0.07	0.08	0.04	0.01	0.03	0.40	0.03	959	0.0
Brassicaceae	Brassica_napus	0.00	0.00	0.00	1.24	0.00	0.00	0.00	0.05	0.02	0.25	0.00	0.02	0.01	0.00	0.00	799	0.0
Asteraceae	Crepis_capilaris	0.02	0.06	0.03	0.16	0.01	0.01	0.01	0.07	0.02	0.03	0.10	0.04	0.02	0.03	0.02	796	0.0
Geraniaceae	Geranium_dissectum	0.03	0.03	0.03	0.05	0.05	0.03	0.03	0.04	0.06	0.03	0.10	0.08	0.02	0.02	0.02	721	0.0
Fabaceae	Lotus_corniculatus	0.38	0.01	0.03	0.02	0.01	0.01	0.01	0.02	0.05	0.05	0.02	0.01	0.02	0.01	0.03	711	0.0
Primulaceae	Anagallis_arvensis	0.00	0.02	0.03	0.04	0.00	0.01	0.02	0.03	0.08	0.11	0.07	0.02	0.01	0.02	0.04	710	0.0
Asteraceae	Picris_echioides	0.02	0.04	0.01	0.04	0.01	0.01	0.01	0.02	0.06	0.09	0.08	0.04	0.03	0.03	0.03	683	0.0
Plantaginaceae	Linaria_vulgaris	0.02	0.02	0.04	0.09	0.01	0.01	0.02	0.04	0.05	0.05	0.09	0.05	0.02	0.03	0.06	661	0.0
Plantaginaceae	Veronica_chamaedrys	0.13	0.01	0.01	0.03	0.02	0.01	0.01	0.01	0.04	0.03	0.01	0.03	0.02	0.01	0.02	646	0.0
Caprifoliaceae	Succisa_pratensis	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.03	0.10	0.12	0.05	0.03	0.01	0.02	0.05	645	0.0
Asteraceae	Bellis_perennis	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.02	0.02	0.03	0.02	0.21	0.58	0.08	0.02	640	0.0
Boraginaceae	Borago_officinalis	0.02	0.02	0.02	0.02	0.04	0.02	0.02	0.03	0.05	0.07	0.04	0.02	0.02	0.02	0.03	640	0.0
Caprifoliaceae	Dipsacus_fullonum	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.03	0.06	0.07	0.03	0.02	0.02	0.02	0.03	614	0.0
Apiaceae	Conium_maculatum	0.01	0.01	0.01	0.01	0.00	0.01	0.08	0.08	0.05	0.06	0.02	0.02	0.01	0.01	0.02	606	0.0
Brassicaceae	Alliaria_petiolata	0.01	0.01	0.00	0.02	0.01	0.01	0.00	0.11	0.05	0.15	0.02	0.01	0.04	0.01	0.02	605	0.0
Geraniaceae	Geranium_robertianum	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.04	0.04	0.03	0.11	0.02	0.02	0.03	586	0.0
Caprifoliaceae	Knautia_arvensis	0.01	0.01	0.02	0.01	0.00	0.02	0.01	0.02	0.09	0.10	0.03	0.02	0.02	0.01	0.04	551	0.0
Balsaminaceae	Impatiens_glandulifera	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.06	0.08	0.02	0.02	0.01	0.01	0.03	550	0.0
Asteraceae	Tripleurospermum_inodorum	0.02	0.03	0.03	0.07	0.01	0.01	0.02	0.02	0.02	0.04	0.03	0.01	0.03	0.03	0.02	543	0.0
Asteraceae	Cirsium_arvense	0.01	0.02	0.01	0.01	0.05	0.02	0.01	0.01	0.08	0.08	0.05	0.02	0.02	0.01	0.03	537	0.0
Violaceae	Viola_arvensis	0.02	0.02	0.01	0.01	0.04	0.03	0.02	0.02	0.03	0.07	0.02	0.02	0.03	0.01	0.01	524	0.0
Asteraceae	Pulicaria_dysenterica	0.01	0.03	0.03	0.06	0.01	0.01	0.01	0.02	0.04	0.05	0.06	0.02	0.01	0.02	0.02	491	0.0
Apiaceae	Heracleum_sphondylium	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.02	0.03	0.04	0.04	0.04	0.02	0.02	0.01	477	0.0
Hypericaceae	Hypericum_perforatum	0.01	0.02	0.02	0.04	0.01	0.01	0.02	0.02	0.03	0.05	0.06	0.03	0.01	0.02	0.02	458	0.0
Onagraceae	Epilobium_hirsutum	0.01	0.01	0.01	0.02	0.00	0.01	0.01	0.03	0.05	0.08	0.06	0.03	0.01	0.01	0.04	436	0.0
Asteraceae	Senecio_jacobea	0.01	0.02	0.01	0.03	0.00	0.00	0.02	0.02	0.03	0.04	0.05	0.02	0.01	0.01	0.02	354	0.0
Convolvulaceae	Calystegia_sepium	0.01	0.01	0.02	0.03	0.01	0.01	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.03	333	0.0
Plantaginaceae	Veronica_agrestis	0.09	0.00	0.00	0.00		0.00	0.02	0.01	0.00	0.01	0.01		0.00	0.00	0.01	324	0.0
Asteraceae	Arctium_lappa	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.07	0.02	0.04	0.02	0.01	0.01	0.00	0.01	318	0.0
Boraginaceae	Anchusa_arvensis	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.05	0.02	0.02	0.03	0.00	0.02	310	0.0
Caryophyllaceae	Cerastium_arvense	0.01	0.00	0.01	0.01	0.02	0.02	0.01	0.01	0.04	0.03	0.02	0.01	0.01	0.02	0.01	299	0.0
Convolvulaceae	Calystegia_sylvetica	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.03	0.03	0.04	0.02	0.00	0.01	0.01	294	0.0
Convolvulaceae	Convulvulus_arvensis	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.01	0.00	0.01	0.02	0.02	279	0.0
Geraniaceae	Geranium_molle	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.08	0.01	0.01	0.01	270	0.0
Malvaceae	Malva_moschata	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.02	0.04	0.05	0.03	0.01	0.01	0.01	0.02	269	0.0
Primulaceae	Primula_veris	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.03	0.03	0.03	0.01	0.00	0.01	0.01	252	0.0

Polygonaceae	Persicaria_maculosa	0.00	0.01	0.01	0.02	0.00	0.01	0.01	0.02	0.02	0.03	0.02	0.00	0.01	0.01	0.01	241	0.0
Plantaginaceae	Digitalis_purpurea	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.03	0.04	0.02	0.01	0.00	0.01	0.02	228	0.0
Apiaceae	Chaerophyllum_temulum	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.02	0.03	0.02	0.01	0.01	0.01	0.01	225	0.0
Caryophyllaceae	Stellaria_holostea	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.03	0.04	0.00	0.01	0.00	0.01	0.01	207	0.0
Asteraceae	Leucanthemum_vulgare	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.03	0.04	0.01	0.02	0.01	0.00	0.01	204	0.0
Cucurbitaceae	Bryonia_dioica	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.01	0.00	0.00	0.00	0.02	202	0.0
Poaceae	Bromus_commutatus	0.00		0.00	0.00		0.00	0.01	0.01	0.03	0.03	0.00	0.01	0.00	0.00	0.02	200	0.0
Lamiaceae	Stachys_sylvatica	0.02	0.00	0.01	0.00	0.00	0.00	0.01	0.02	0.01	0.02	0.00	0.00	0.01	0.00	0.01	194	0.0
Boraginaceae	Pentaglottis_sempervirens	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.02	0.03	0.01	0.00	0.02	0.00	0.01	177	0.0
Urticaceae	Urtica_dioica	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.03	0.03	0.01	0.00	0.01	0.01	0.00	175	0.0
Brassicaceae	Cardamine_flexuosa	0.00	0.00	0.00		0.00	0.00	0.00	0.09	0.00	0.01	0.00	0.01	0.00	0.00	0.00	173	0.0
Asteraceae	Tripleurospermum_maritimum	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.03	0.00	0.01	0.01	0.00	0.01	160	0.0
Plantaginaceae	Veronica_persica	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.01	0.01	0.00	0.00	0.01	153	0.0
Lamiaceae	Clinopodium_vulgare	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.03	0.00	0.01	0.01	0.01	143	0.0
Caryophyllaceae	Silene_dioica	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.01	140	0.0
Polygonaceae	Rumex_obtusifolius	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.01	0.01	0.00	0.01	0.01	132	0.0
Poaceae	Elymus_caninus	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.02	0.00	0.01	0.02	129	0.0
Caryophyllaceae	Stellaria_media	0.00	0.00	0.00	0.01	0.00	0.00	0.00		0.02	0.02	0.00	0.00	0.00	0.00	0.00	119	0.0
Fabaceae	Trifolium_campestre	0.01	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00		0.00	0.01	117	0.0
Poaceae	Lolium_perenne	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.02	0.01	0.00	0.01	116	0.0
Resedaceae	Reseda_luteola	0.00	0.00	0.00	0.01		0.00	0.00	0.02	0.01	0.02	0.02	0.00	0.00	0.00	0.02	107	0.0
Asteraceae	Hypochaeris_radicata	0.00	0.01	0.00	0.00		0.00	0.00	0.01	0.00	0.01	0.01	0.01		0.00	0.01	102	0.0
Phrymaceae	Mimulus_guttatus	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.00	0.01	100	0.0
Rubiaceae	Galium_verum	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01		0.00	0.00	0.01	99	0.0
Asteraceae	Centaurea_nigra	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00	95	0.0
Papaveraceae	Papaver_rhoes	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.00	0.00	85	0.0
Caryophyllaceae	Stellaria_graminea	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00		0.00	0.00	74	0.0
Caryophyllaceae	Silene_vulgaris	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.01	0.01		0.01	0.00	0.00	0.00	73	0.0
Lamiaceae	Prunella_vulgaris						0.00	0.01	0.01							0.01	60	0.0
Poaceae	Phleum_pratense	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	58	0.0
Poaceae	Holcus_lanatus	0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00	0.01	0.01	0.00		0.00	53	0.0
Asteraceae	Matricaria_discoidea		0.00	0.00	0.00		0.00	0.00	0.00	0.01	0.00	0.00		0.00	0.00	0.01	47	0.0
Fabaceae	Medicago_lupulina	0.00	0.00		0.00		0.00	0.00		0.00	0.00			0.00		0.01	20	0.0

Supplementary Table 4.4 Pairwise analyses of sampling period and farm on the pollen communities of the pollen loads, using *mvabund*.

Sampling period pair	Observed statistic	p-value
1 vs 4	654.0	**0.002**
2 vs 4	542.9	**0.002**
1 vs 3	373.4	**0.002**
3 vs 4	296.8	**0.002**
2 vs 3	287.8	**0.002**
1 vs 2	220.9	**0.002**
Tuesley vs Winterwood	232.5	**0.001**
Colworth vs Tuesley	198.6	**0.001**
Heathlands vs Tuesley	107.4	**0.001**
Heathlands vs Winterwood	96.9	**0.001**
Colworth vs Heathlands	83.2	**0.001**
Colworth vs Winterwood	81.0	**0.001**

Supplementary Table 4.5 results from PERMANOVA analyses testing the effects of sampling period and colony on the pollen communities.

Farm	Predictor	R^2	F	p-value
Heathlands	Sampling.period	0.24	6.5	**0.002**
	Colony	0.29	1.6	0.102
Tuesley	Sampling.period	0.37	11.7	**0.001**
	Colony	0.28	1.7	0.075
Colworth	Sampling.period	0.50	21.0	**0.001**
	Colony	0.10	0.8	0.424
Winterwood	Sampling.period	0.75	16.2	**0.001**
	Colony	0.06	0.5	0.86
All farms	Sampling.period	0.15	21.9	**0.001**
	Farm	0.23	10.8	**0.001**
	Colony	0.36	1.2	0.094

Supplementary Table 4.6 Summary of best fitting models with a dAICc < 2, the global model and the null model, with the corresponding explanatory variables. Explanatory variables: Flr.rich = floral richness in the landscape, RA.vacc.fl = Proportion of *Vaccinium* flowers in the landscape, SO = sampling period.

Model	df	logLik	AICc	dAICc	Aikeke weight (*wi*)	Explanatory variable
Pollen richness						
PR1	5	-170.8	352.4	0.00	0.352	Flr.rich + RA.vacc.flr
PR2	11	-162.4	352.7	0.30	0.305	RA.vacc.flr + SO + SO: RA.vacc.flr
PR3	*6*	-170.2	353.5	1.07	0.207	Flr.rich + RA.vacc.flr + flr.rch:RA.vcc.flw
PR4	*4*	-172.9	354.3	1.88	0.137	RA.vacc.flr
Global	12	-162.4	353.2	2.72	0.099	Flr.rich + RA.vacc.flr + SO + RA.vacc.flr:SO + flr.rich:RA.vacc.flr
Null	3	-177.8	361.9	11.43	0.001	1
Vaccinium RA						
V1	11	20.8	-15.9	0.00	0.53	RA.vacc.flr + SO + RA.vacc.flr:SO
V2	8	16.8	-15.7	0.25	0.47	RA.vacc.flr + SO
Global	13	21.6	-12.0	3.91	0.05	Flr.rich + RA.vacc.flr + SO + RA.vacc.flr:SO + flr.rich:RA.vacc.flr
Null	4	6.18	-3.9	12.08	0.001	1

Supplementary Table 4.7 Effects of flower richness in the landscape and proportion of *Vaccinium corymbosum* flowers in the landscape on the pollen richness (PR) in the colony. The table shows the estimates of a poisson GLMM.

Model	Explanatory variables	Estimates	SE
PR1	(Intercept)	1.38	0.23
	Floral.richness	0.03	0.01
	RA.vacc.flr	-0.47	0.23
PR2	(Intercept)	2.06	0.41
	RA.vacc.flr	*-1.*59	0.70
	Factor(Sample.period)2	*0.40*	0.76
	Factor(Sample.period)3	*-0.69*	*0.42*
	Factor(Sample.period)4	*-0.76*	*0.46*
	RA.vacc.flr: Factor(Sample.period)2	*-0.24*	*1.30*
	RA.vacc.flr: Factor(Sample.period)3	*1.82*	*0.77*
	RA.vacc.flr: Factor(Sample.period)4	*3.90*	*1.59*

Supplementary Table 4.8 The relative importance of explanatory variables expressed by the Σwi for models to explain the effects of *Vaccinium* RA in the landscape, landscape floral richness, sampling period on the pollen richness of the colonies and proportion of *Vaccinium* pollen collected.

Response variable	RA.vacc.flr	Flr.rich	SO	SO:RA.vacc.flr	Flr.rich:RA.vacc.flr
Pollen richness	0.86	0.61	0.32	0.25	0.17
Vaccinium RA	0.98	0.12	1.0	0.46	< 0.01

Supplementary Table 4.9 Effects of flower richness in the landscape and proportion of *Vaccinium corymbosum* flowers in the landscape on the proportion of *Vaccinium* reads (V) in the pollen loads. The table shows the estimates of a GLMM with a beta distribution.

Model	Explanatory variables	Estimates	SE
V1	(Intercept)	-2.15	0.69
	RA.vacc.flr	2.44	1.39
	Factor(Sample.period)2	1.69	1.29
	Factor(Sample.period)3	1.65	0.69
	Factor(Sample.period)4	3.65	0.76
	RA.vacc.flr: Factor(Sample.period)2	-1.50	2.32
	RA.vacc.flr: Factor(Sample.period)3	0.59	1.44
	RA.vacc.flr: Factor(Sample.period)4	-8.50	3.33
V2	(Intercept)	-1.68	0.46
	RA.vacc.flr	1.85	0.58
	Factor(Sample.period)2	0.73	0.30
	Factor(Sample.period)3	1.25	0.31
	Factor(Sample.period)4	1.89	0.46

Chapter Five

Discussion

5.1 Overall aim

The overall aims of this thesis have been to test the ability of a new molecular technique to identify and quantify the pollen collected by foraging bumblebees. Agricultural landscapes are facing declines in insect diversity and abundance, which is of particular concern due to their importance as pollinators. The prominent floral resources in pollinator pollen diets can be identified to increase the plantings and improve the management of these taxa in the landscape, which serve as effective approaches in supporting communities of insects in farmland.

To achieve these aims we used a combination of microscopy, nanopore sequencing and landscape-level floral surveys to investigate the foraging of commercial bumblebees (*Bombus terrestris*) over the flowering period of highbush blueberry (*Vaccinium corymbosum*). RevMet is a recently developed approach, which was tested in this thesis for its ability to characterise and quantify the pollen grains of mixed species corbicular loads collected from foraging bees. Practically, the research has provided evidence and recommendations for the use of RevMet to analyse pollen loads that reflect real pollen grain relative abundance. From a theoretical perspective, the thesis adds to a growing literature on the foraging preferences of bumblebees in mass-flowering crops, with a focus on *V. corymbosum.* The effect of mass-flowering crops on pollinator foraging is not well studied in *V. corymbosum*, and this thesis has contributed to an understanding of how floral resource quality and quantity can affect the decisions of bumblebee workers in their pollen collection.

In this chapter the aims and findings of my three data chapters are summarised and synthesised. RevMet is then reviewed as a method to compare pollen identification and quantification, from which we outline future directions in research for this approach. We conclude with a list of considerations for supplementing floral resources in agricultural landscapes using the results from this thesis and outline the main limitations encountered in carrying out this research.

5.2 Summary of findings

5.2.1 Chapter Two

The aim of Chapter Two was to test the quantitative abilities of RevMet against microscopic grain counts. Mixed species pollen loads were collected from commercial *B. terrestris* workers and identified using light microscopy with comparisons to RevMet reads sequenced with the nanopore platform. Further, the test enabled an exploration of the foraging patterns of bumblebees including changes in diversity and composition between early and peak *V. corymbosum* crop flowering.

RevMet results revealed a higher taxa richness and level of taxonomic discrimination compared to the microscopy results. There was a positive relationship between the relative abundance of RevMet reads and the number of pollen grains identified using microscopy. In the early crop flowering period bees were foraging primarily from *Salix caprea* and in peak crop flowering the majority of pollen originated from *V. corymbosum* and also taxa in the Rosaceae family. In the RevMet results there was a significant proportion of reads classified as unassigned, which was fairly constant throughout the samples. We conclude the unassigned reads are an artifact of the RevMet process or originate from eDNA sources, with no significant effect on the plant taxa compositional results.

5.2.2 Chapter Three

The aim of Chapter Three was to compare the qualitative and quantitative abilities of different molecular approaches that use nanopore sequencing to characterise mock pollen mixtures of known proportions. The two approaches used nanopore sequencing, one using long reads (RevMet) and the other spanning the ITS2 barcode region. Whole genome sequencing approaches such as RevMet are more likely to be affected by factors such as genome size, so we created a correction factor based on C-values in order to improve the quantitative results produced by RevMet.

We found that ITS2 produced strong correlations against relative abundances of pollen estimated using microscopy at species-, genus- and family-levels of taxonomy. Whilst

RevMet provided species-level identifications and low false negative detection rates, the quantitative results were less accurate. The applications of a genome correction factor improved the quantitative relationship of RevMet and mock mixture pollen grain proportions, so we recommend further analysis with a greater number of plant taxa. The results of this chapter suggest that the biases affecting whole genome sequencing approaches should be explored.

5.2.3 Chapter Four

Chapter Four investigated the foraging of commercial bumblebees at four farms during the *V. corymbosum* flowering period. We used landscape-level floral surveys that encompassed the approximate foraging range of bumblebees to measure the availability of resources in the landscape. We used the floral survey results to identify the preferences of foraging workers based on the proportion of sequencing reads allocated to each plant taxon's relative abundance in the landscape. We also analysed the effects of *Vaccinium* flower cover and landscape floral richness on the number of plant taxa collected and the proportion of *V. corymbosum* reads in the pollen loads.

The findings demonstrated that although approximately half of the sequencing reads originated from *Vaccinium corymbosum*, the workers foraged extensively off-crop, as in Chapter Two. Foraging was affected by the proportion of *Vaccinium* crop and the richness of floral resources in the landscape, in part explained by the timing of flowering, i.e. phenology. Woody and shrubby plant taxa were the most utilised by bumblebee workers in early spring. This information can be used to create plant lists in order to recommend pollen resources that best support bumblebees at this time of year. We concluded that a higher cover of *V. corymbosum* crop attracted foraging bees and that flower relative abundance in the landscape did not reflect the relative abundance of plant taxa in the pollen diet.

5.3 Testing the RevMet approach

The thesis tested the RevMet approach and applied it in different contexts across the three data chapters. RevMet was initially developed as a low-cost method of characterising mixed pollen loads of pollinators, which it successfully achieved (Peel et al., 2019). Peel et al. (2019) also tested RevMet's quantitative abilities, using extracted DNA as the mock mixture inputs. The results showed that RevMet was able to predict the relative abundances of plant taxa in a "semi-quantitative" way, i.e. pollen present in low or high abundances were correctly estimated. What was missing from this study was the comparison between pollen grain proportions and sequence read proportions allocated by RevMet, which had not yet been explored but is an important step in the ecological application of this method. We set out to answer this question in Chapter Two, using pollen loads collected from foraging bees. Pollen grains were counted and identified using light microscopy, calculating the relative abundances of pollen taxa in order to compare the proportions to the RevMet results. We found that there was a positive relationship between the relative abundances predicted between the two methods, although this was complicated by the bumblebees' preference for a single plant taxon on a foraging trip. The flower constancy exhibited by *Bombus terrestris* workers meant proportions of plants taxa were either in very high (> 80%) or low (< 20%) abundance, making the predictive ability of the method to detect different quantities of pollen hard to discern.

These results prompted further questions about the quantitative ability of RevMet that we set out to answer in Chapter Three. Constructed mock mixtures of pollen grains were created from corbicular loads collected during the 2019 and 2021 field seasons. The mock mixtures contained taxa with a range of genome sizes, which were found to affect the quantitative ability of RevMet when results were compared to the input proportions of mock mixture pollen grain taxa. The taxon with the largest genome, *Ficaria verna,* was found to be consistently over-represented in the results, whilst the taxon with the smallest genome size, *Salix caprea*, was found to be underestimated. Pollen grains that belong to a plant taxon with a large genome size will release a higher yield of DNA, and will therefore produce a larger amount of sequencing data. RevMet was accurate in estimating the true positives of

the mock mixtures, so we suggest a solution to the genome-bias is to assess the genome sizes of the positively identified plant taxa and, if there is a large degree of variation, to apply a correction factor. In this chapter we created a correction factor based on the genome sizes of the mock mixture species, which was effective in producing a strong positive relationship between the adjusted RevMet proportions and the proportions of mock mixture pollen grains. The correction factor was only tested on a small group of taxa, so we recommend further investigation into the impact of genome size variation and the application of correction factors with which to reduce the quantitative bias.

Finally, the RevMet approach was applied to a landscape-scale study of foraging *B. terrestris* workers in UK farms growing highbush blueberry (*Vaccinium corymbosum*). The approach was used to identify the proportions of different taxa collected by commercial bees with a particular focus on the amount of *V. corymbosum* proportion brought back to the colonies, which can provide information about crop visitation and therefore pollination services. The nanopore sequencing for RevMet utilised the available 96-barcode kits, which made processing a larger number of samples more efficient and cost effective.

5.4 The suitability of RevMet in pollen identification

The most utilised method in pollen identification is DNA metabarcoding, which uses short regions of the genome that contain high interspecific variation and low intraspecific variation to identify taxa (Bell et al., 2022). DNA metabarcoding has been found to be good at revealing hidden networks and identifying rare plant taxa that could be missed by microscopic identification (Pornon et al., 2017; Smart, 2017). RevMet performs a comparable role in identifying low levels of plant taxa in pollen loads and to a high level of taxonomic discrimination.

We found RevMet to be an effective approach in identifying the species present in the mock mixtures, which contained similar or higher levels of species richness to those we can expect in individual bumblebee pollen loads. If this technique was expanded to honey bee hives, where the samples contain corbicular loads from a large number of bees and therefore

higher taxa richness, RevMet would need to be tested to ensure the full diversity of pollen community has been captured. In Chapters Two, Three and Four we applied a minimum abundance sequence threshold of 1%, because that was effective in reducing the number of false positive identifications. However, as demonstrated in Chapter 4 for *Prunus spinosa*, this is likely to remove taxa that are true positives present at low abundances, so care should be taken when using the filtering step depending on the pollen collection method and source. In future studies we would recommend modelling a minimum abundance filter based on the sequencing results from mock communities and negative controls, to create a more accurate and reasonable cut off for reads that might represent false positives in the data (Drake et al., 2021).

The RevMet approach uses whole genome sequencing (WGS) rather than short barcode regions, which are used in DNA metabarcoding. WGS techniques such as RevMet have shown potential in providing strong quantitative abilities in comparison to pollen grain counts (Bell et al., 2021; Lang et al., 2019; Peel et al., 2019). The WGS approaches used varied in approach; Lang et al. (2019) used plastid DNA to create reference genomes, while Bell et al. (2021) used publicly available genomes to align short pollen reads generated from whole genome sequencing. These amplification-free approaches remove the possibility of PCR-biases, with the potential to improve quantitative abilities. When Bell et al. (2021) compared WGS results to barcode results from the same samples, WGS provided a stronger correlation between sequence proportions and pollen grain proportions. The quantitative results we found using ITS2 may have had a stronger correlation to pollen grain counts than those reported in the literature because the reference database used in Chapter Three was not a full plant database for the UK, because we wanted to provide a comparison to the RevMet approach. With a larger database, there is likely to be higher levels of misidentification in amplicon sequencing due to similarities between barcodes of closely related taxa. Overall, this thesis adds to a growing body of research that suggests WGS approaches could be more commonly adopted in the future, based on RevMet's strong qualitative and semi-quantitative results, but requires further testing.

5.5 Future uses and developments for RevMet

RevMet has the potential to be used in different applications. In this thesis we used RevMet to characterise bumblebee collected pollen loads, but it could be applied to plant-pollinator studies on solitary bees, honey bees and other groups of insects. The amount of DNA required for library preparation can be lower than the 400ng ONT recommends (as has been shown in this thesis), and depending on the expected dietary breadth of the samples there might not be a requirement for large quantities of pollen, such as, pollen present on the bodies of solitary bees. There could be opportunities to explore the sensitivity of RevMet using samples containing eDNA or microbial communities, for example soil samples or air samples. These samples contain a high diversity of taxa at low quantities, and so would provide an interesting test of RevMet's qualitative and quantitative abilities (Kraaijeveld et al., 2015; Stevens et al., 2023).

A second potential application for RevMet is airborne pollen sampling, where pollen is collected and assessed for its presence of allergenic pollens. This information could be used to better predict spikes in hay fever and identify the plant taxa that are the main contributors to the condition (Brennan et al., 2019; Polling et al., 2022).

Outside of pollen studies, RevMet could be applied to other agricultural questions. Rapid detection of crop pathogens is an important area of research into maximising crop production. Nanopore technology has been used to successfully detect bacterial, fungal and viral pathogen RNA or DNA in crops (Chalupowicz et al., 2019). A major advantage of using RevMet to detect pathogen sequences is in the availability of assembled genomes for these organisms, which is not true for plant taxa. This enables the RevMet pipeline to be applied without the necessity to assemble genome skims to align against, based on shotgun Illumina sequencing, rendering the whole process significantly less costly.

In studies where the focal taxa are organisms that do not have a readily available library of assembled reference genomes, such as plants, WGS approaches can be limited in their application (Bell et al., 2021). RevMet provides a simple and low-cost solution by creating a set of low-coverage genome skims in the place of fully constructed genomes. Although this

is an added expense, Illumina sequencing is becoming more cost-efficient with new platforms being capable of higher sequencing outputs. We aimed for a 1x coverage, in order to capture the variation in loci between closely related plants, but we recommend further investigation into the ability of RevMet to distinguish between closely related species. The advantage of creating a reference library is that the plants used are relevant to the study and capture the foraging landscape of the study. There are projects in place that aim to sequence all eukaryote species for public use, therefore genomes will be available for future studies which will make WGS approaches a much more promising alternative (Lewin et al., 2018). The low coverage of the RevMet genome skims means that closely related species might be harder to discriminate between (Peel et al., 2019), so using assembled genomes for the RevMet approach will likely improve taxonomic resolution.

5.6 The future of nanopore sequencing

In this thesis we provide evidence of the capabilities of nanopore sequencing for pollen taxa characterisation. Nanopore sequencing was used for both long read sequencing (Chapters Two, Three and Four) and ITS2 barcode sequencing (Chapter Three).

Nanopore has been used as an alternative to the Illumina platform in pollen metabarcoding studies, where the compositional results were found to be comparable between the two methods (Leidenfrost et al., 2020). The advantages of nanopore over Illumina lie in its long read capabilities, whereas Illumina is restricted to 150 – 250 bp paired end reads, which often does not span the full barcode region. Nanopore suffers from lower read accuracy, although these have increased from approximately 60% on first release (Rang et al., 2018), to current estimations of 90% (Delahaye & Nicolas, 2021), although ONT claim higher. These accuracies are significantly lower than the Illumina platform, which demonstrates 99% read accuracy (Rang et al., 2018). We used high accuracy basecalling, which filters sequences based on a read quality score and was applied to reduce the possibility of false positives.

New models, chemistry and basecalling updates are continuously improving the accuracy and efficiency of nanopore sequencing, which will improve applications in metabarcoding

and WGS approaches (Ferguson et al., 2022). Upcoming releases include ONT's Plongle, which includes adapters that allow for low sequencing depth on a larger number of samples, thereby reducing the per-sample cost. SmidgeION is a more compact, smartphone attachment that will allow for in-field sequencing of samples such as real-time species identification and rapid diagnosis of infectious disease. We can also anticipate improvements in the membrane of the nanopores, with more robust materials likely to be used over the current protein nanopores. Currently, flow cells have a shelf life of approximately 3 months in a refrigerator or 30 days in ambient temperatures. With the development of more solid-state pores made from materials such as graphene, it is likely the use of the flow cells will be prolonged, and they will withstand higher rates of re-use (He et al., 2021).

5.7 Factors affecting bumblebee foraging preferences

We found several factors to have significant effects on foraging, which provide information on the foraging decisions made by bumblebees. First, phenology was a significant contributor to the foraging patterns displayed by the workers. The availability of resources changes throughout the year and this affected the foraging patterns of the bumblebees in Chapter Four. This change in floral resource availability was mirrored in the bumblebee pollen diets, which highlights the importance of a diverse landscape of flowering plants. This has been found in previous studies that examine resource use in bees, and highlights the need to fill gaps in resource provisioning (Timberlake et al., 2019).

Second, the quantity of a resource in the landscape affects bumblebee foraging. Our results from Chapter Four suggest a higher cover of *V. corymbosum* increased the proportions of crop pollen that were brought back to the colony. The presence of mass-flowering crops in the landscape has been found to increase the pollinator abundance at the local scale, depending on the landscape configuration (Kovács-Hostyánszki et al., 2013). In agricultural landscapes with few areas of semi-natural habitat, there was a larger effect of a mass-blooming resource on pollinator diversity and abundance on the crop (Jauker et al., 2012; Riedinger et al., 2014). On our farms there were differences in the surrounding landscape,

which is likely to affect the reliance on crop flowers for pollen rewards. If there are no co-flowering crops in the surrounding area, and fewer areas of semi-natural habitat or gardens, we would expect a higher proportion of crop pollen in the diet. We might expect this to have a detrimental effect on the colony health, because bees require a balanced and nutritious diet for colony health and development, and there is potential for chemical exposure originating from pesticides and fungicides that are sprayed on the crop, but we are unsure of these effects in our study (Gill & Raine, 2014; Roger et al., 2017).

Bumblebees have been found to forage based on quality rather than quantity, but our results suggest that in the presence of a mass-flowering crop the high abundance and density of floral resource was attractive to workers. However, bees were still found to collect the majority of the pollen from non-*Vaccinium* sources, preferring woody and shrub plant taxa such as *Salix caprea* and *Prunus spinosa* as pollen sources. These plants could produce higher quality pollen, in terms of protein content and other nutritional parameters, than *Vaccinium,* and were likely found in high abundance in the pollen as a way for the bumblebees to balance the nutritional intake of the colony (Roulston et al., 2000; Somerville, 2001). The results of Chapter Four provide recommendations for wild flower plantings and floral resource management in early spring to increase important foraging resources for bumblebees in agricultural landscapes, which can be resource-poor (Carvell et al., 2006; Samuelson et al., 2018).

5.8 Considerations for *Vaccinium corymbosum* growers

Using the findings from this thesis, there are some considerations when designing pollinator-friendly landscapes in agricultural ecosystems. Mass-flowering crops (MFCs) provide a large volume of floral resources that are utilised by wild and commercial pollinators. There is a degree of competition between MFCs and other co-flowering crops or wildflowers in the landscape, which depends on the attractiveness and quality of the competing resources (Grab et al., 2017). We observed that although *V. corymbosum* was the most collected pollen taxon, bumblebees would often forage elsewhere. There is a balance between quantity and quality of pollen, and we found that in landscapes of low-quality pollen (such as *V.*

corymbosum), regardless of the high crop flower density, workers were likely to forage a significant amount of pollen from higher quality sources. In Chapters Two and Four we identified *Salix caprea* to be a major contributor to the pollen diet, as well as other woody taxa such as *Prunus spinosa,* both of which have higher protein contents than *V. corymbosum* (Roulston et al., 2000).

If we were to recommend flowering taxa to plant, or encourage through management, in order to support local populations of bumblebees, our results suggest that *S. caprea* and *P. spinosa* should be included as excellent early spring foraging resources. However, in the context of pollinator-dependent crops the recommendations are not so simple. We found that commercial bees preferentially foraged from these taxa over the crop flower, which at first might seem problematic to growers. However, a varied pollen diet has been found to improve the health of the colony which might increase foraging abilities on the crop. We did not measure this outcome, nor the effect of more attractive resources in the landscape on the yield and quality of the blueberry fruits, but previous studies have shown that *V. corymbosum* has a relatively low pollinator dependence (Kendall et al., 2020). This means the flowers need a low number of pollinator visits to achieve full fruit set so it may not be of consequence if a significant proportion of the workers forage elsewhere. This finding requires further investigation into the effect of high densities of bumblebee colonies on the flower visitation rates, resource use, and fruit yield of the crop.

5.9 Limitations

As with all research, there are some limitations, and our findings should be interpreted with caution in the light of them. We see five main caveats:

First, the RevMet approach uses a genome skim reference library that we created from samples of plant taxa that were recorded as being present in the approximate foraging range of bumblebees (1 km). The sampling strategy used meant that only a small proportion of the 1 km radius around the centre of the farm could be explored, which was mostly due to time and the sampling effort required. As a result, we may have missed a large number of

flowering plant taxa that were present but unrecorded. This remains an unknown in our results, as we cannot accurately predict the number of false negatives in the data. We attempted to resolve this problem by creating a 20% threshold on the number of assigned reads ins Chapters Two and Four, i.e. if fewer than 20% of the total number of sample reads were assigned, we would exclude the sample from analysis. This threshold level was supported by rarefaction curves of the sequencing depth. We made an assumption that a low number of unassigned reads signified the presence of a plant taxon that was missing from our reference database. Using this assumption, we could make inferences about the number of samples that had missing taxa (or to be precise, taxa missing at > 80% proportions). Approximately 6% of the samples in Chapters Two and Four were given this unassigned status, which means 6% of the bees' foraging resources were unaccounted for. As plant genome reference databases become more complete, the percentage of reads that are unaccounted for will likely decrease. It is also possible that a proportion of the unassigned reads encountered in the RevMet originated from microbial genomes and other sources of environmental DNA. Pollen grains are collected from the flower anther by the bee and moistened using saliva and nectar, before being gathered into corbicular loads and transported back to the colony. There is evidence of communities of microorganisms including bacteria and fungi that are present in fresh bee pollen and likely to have been included in the collected samples (Mauriello et al., 2017; Pelka et al., 2021). In NGS studies, all the DNA is sequenced in a non-targeted approach, which makes it likely that microorganismal DNA and eDNA will be sequenced alongside plant DNA. In our study it's likely that these external sources of DNA present in the pollen loads were sequenced with the pollen DNA, and categorised as "unassigned" in the RevMet alignment process because their genomes were not present in the reference library. It would be an interesting next step to sequence these reads and investigate their origin, to better understand the communities of microorganisms and eDNA sources that are carried in corbicular loads.

Second, the COVID-19 pandemic impacted the fieldwork that was planned during the four years of research. By the spring of 2021 there were still significant restrictions in place, which meant that certain areas around the farms could not be accessed, for example, private and residential gardens. Gardens are known to be important sources of floral

diversity, particularly in comparison to agricultural landscapes (Samuelson et al., 2018; Tew et al., 2021). The proportion of gardens has been shown to be an important predictor of *Bombus terrestris* colony density in southwest England, relatively close our sites (Timberlake et al., 2021). This suggests gardens provide important pollen resources for this species, while recognising from a study in a botanical garden, that garden bees and hoverflies have a preference for native and near-native plant taxa over horticultural taxa (Lowe et al., 2022). We were unable to include garden plants in the Chapter Four analyses, so the importance of urban flower plantings in our results cannot be evaluated.

Third, the limited number of farms in Chapter Four means we were unable to make larger claims on the impact of crop flower cover on the foraging preferences of commercial bumblebees. The inclusion of four farms meant we were able to survey the flowering landscape in more detail, but at the expense of being able to make conclusions on the landscape effects of foraging on *V. corymbosum* farms. Other studies that have measured the pollen diets of pollinators in agriculture have measured resource use at a landscape level, including percentage cover of crop or semi-natural habitat in analyses (Bänsch et al., 2020). Few have conducted floral surveys of the landscape around the crop that have quantified the abundance of flowering taxa so we can argue that our analyses provide a more detailed prospective of foraging (Jha et al., 2013),

Fourth, the thesis was limited to studying a single species of bumblebee, *Bombus terrestris,* which had been deployed for commercial reasons. Results from the foraging preference studies should not be used to infer preferences for a wider pollinator community, because groups of insect taxa have varied dietary needs and preferences (Nichols et al., 2019). *B. terrestris* is a generalist species, known to be adaptable in its pollen diet and with a large foraging range relative to other species (Dicks et al 2015). Other pollinators, such as solitary bees, are under-studied and might be more vulnerable to environmental changes than generalist bumblebee species (LeBuhn & Vargas Luna, 2021). Therefore, we can recommend plant taxa that were used to a high degree in this research to attract *B. terrestris*, but further plant lists are required to support a full range of insect species.

Finally, we calculated the quantitative relationship of RevMet reads to the relative abundance of pollen grains in Chapter Three and found there to be an effect of genome size on the quantitative abilities of RevMet. This relationship became stronger once a genome-correction factor had been applied to the RevMet relative abundances. In Chapter Four we used the raw RevMet reads without the application of a genome correction factor because we feel this step needs further testing. The over- or underestimation in relative abundance occurs on a per sample basis because we sequenced individual bee corbicular loads rather than pooling samples together. Therefore, the probability of quantification bias depends on the taxa present in a single sample. We examined the abundance status of different taxa and categorised them into major, intermediate or minor pollen categories (Figure 6). If genome size was distorting our results we might expect *Ficaria verna* and *Hyacinthoides non-scripta,* two taxa in the top ten most utilised species which have large genomes relative to the other taxa, to frequently occur in the "major" category (>45% of assigned reads). This effect was not observed in our results as these taxa were predominantly in the "minor" category (<10% of assigned reads) although it is possible that their importance has still been overestimated. The results in our null model also predict *F. Verna* and *H. non-scripta* to have been collected less than predicted based on their floral relative abundance in the landscape in some of the sample periods. If there was significant overestimation of these taxa, we might expect to have results that suggest bees were preferentially visiting these taxa in the landscape. Similarly, *Salix caprea* was underestimated in Chapter Three mock mixtures, so it could be even more prevalent than our results suggest. Regardless of quantitative bias, we still feel that the quantitative data used here is more informative than a binary presence/absence approach of taxa present in the pollen loads, which would inflate the importance of all taxa present at low proportions. There may also be an effect of bicellular or tricellular pollen on the amount of DNA released by the pollen grain, potentially impacting the RevMet results. Some pollen cells have one generative cell on release from the anther whereas others have two, which will produce a confounding effect of cell number on the quantity of DNA produced by the pollen grain at the DNA extraction stage, regardless of genome size (Thomas et al., 2003). We recommend further investigation to measure the potential quantitative bias in these samples.

5.10 Future directions

This thesis has highlighted how nanopore sequencing can be used to provide information on the foraging patterns of bumblebees in agricultural landscapes. We have demonstrated good qualitative and quantitative abilities of RevMet when compared to proportions of pollen grains estimated using microscopy, but there is a requirement for further research and improvements.

The high number of unassigned reads could be resolved by automatically reassigning those reads until there are only "unassignable" reads left, i.e. reads of low quality and short length or belonging to taxa missing from the database. We investigated the unassigned reads in Chapter Two, and found a consistent level of ~30% unassigned reads per sample. These reads could be reassigned when re-run through the pipeline, and didn't have an effect on the compositional results. From this we can conclude that for the majority of samples (those not excluded from analysis due to low assigned read percentage) the taxa were unassigned at the same rate and so we didn't find a requirement to assign 100% of the taxa. However, the effects of this development could be the focus of other projects using future versions of RevMet.

This thesis applied RevMet as a method to mixed samples of pollen loads and bee-collected pollen samples to test its quantitative ability. The results were mixed. In Chapter Two we found a stronger quantitative ability when compared to estimated microscopic relative abundances than when we created mock mixtures in Chapter Three. There is little research on whole genome sequencing of mixed pollen loads with which to compare our results to, so valuable further research could be undertaken to test RevMet on samples of varying degrees of taxon relatedness, higher species diversity, and lower quantities of pollen grains.

We used DNA sequencing to answer questions about bumblebee foraging in an agricultural landscape. Previous research has been conducted into the effect of mass-flowering crops on foraging preferences by pollinators, but none on UK blueberry farms. We found a strong effect of season and farm location, so we suggest floral plantings that are appropriate for the

time of year, the target pollinator groups and the existing landscape-level plant taxa. Future studies should further investigate the impact of mass-flowering crops on pollen foraging in bees, with a particular focus on the equilibrium between foraging for quality or quantity. Crop flowers that produce low-quality pollen might cause pollinators to forage off-crop, possibly moving to a more attractive co-flowering crop (Grab et al., 2017). These are management implications that should be considered when growing multiple flowering crops on one farm, especially if their blooming periods overlap.

The question of pollen quality is a key one that might provide answers to how pollinators make their foraging decisions. Pollen ratios and concentrations of protein, lipids, sterols and amino acids have all been found to have an effect on bee foraging and colony health (Vaudo et al., 2020; Vaudo, Patch, et al., 2016; Vaudo, Stabler, et al., 2016). There is evidence to suggest bumblebees can discriminate between high- and low-quality pollen, which affects their foraging patterns (Kämper et al., 2016; Patrizia et al., 2016; Ruedenauer et al., 2016). This means floral resource quality plays an important role in bee foraging preferences and should be further explored to understand its place in the interplay between foraging decisions and consequences on the health of the colony (Brodschneider & Crailsheim, 2010; Di Pasquale et al., 2013; Roger et al., 2017). Currently, the main drawback is that this information is only available for a small selection of plant taxa, so there should be increased efforts to record pollen nutritional quality for common plant taxa.

5.11 Concluding remarks

This thesis set out to test a new bioinformatic approach, RevMet, and its ability to characterise and quantify the pollen load compositions of foraging bees. There are improvements and developments to be made with the technique, but the work presented in this thesis shows its potential to answer ecological questions. In particular, this work adds to a growing body of knowledge that bumblebee foraging is a complex system that is influenced by multiple factors, including the quantity, quality and spatial distribution of

resources. This research demonstrates a significant and urgent need for further research on plant-pollinator interactions and the implementation of appropriate conservation strategies in the face of rapid and unsustainable environmental decline.

References

Adamowicz, S. J., & Scoles, G. J. (2015). International Barcode of Life: Evolution of a global research community. *Genome*, *58*(5), 151–162. https://doi.org/10.1139/gen-2015-0094

Aizen, M. A., & Harder, L. D. (2009). Report The Global Stock of Domesticated Honey Bees Is Growing Slower Than Agricultural Demand for Pollination. In *Current Biology* (Vol. 19, pp. 915–918). https://doi.org/10.1016/j.cub.2009.03.071

Alaux, C., Ducloz, F., Crauser, D., & le Conte, Y. (2010). Diet effects on honeybee immunocompetence. *Biology Letters*, *6*(4), 562–565. https://doi.org/10.1098/rsbl.2009.0986

Albrecht, M., Kleijn, D., Williams, N. M., Tschumi, M., Blaauw, B. R., Bommarco, R., Campbell, A. J., Dainese, M., Drummond, F. A., Entling, M. H., Ganser, D., Arjen de Groot, G., Goulson, D., Grab, H., Hamilton, H., Herzog, F., Isaacs, R., Jacot, K., Jeanneret, P., … Sutter, L. (2020). The effectiveness of flower strips and hedgerows on pest control, pollination services and crop yield: a quantitative synthesis. *Ecology Letters*, *23*(10), 1488–1498. https://doi.org/10.1111/ELE.13576

Allen, G., & Davies, R. G. (2023). Canopy sampling reveals hidden potential value of woodland trees for wild bee assemblages. *Insect Conservation and Diversity*, *16*(1), 33–46. https://doi.org/10.1111/ICAD.12606

Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*, 1–15. https://doi.org/10.1002/9781118445112.STAT07841

Archer, C. R., Fähnle, J., Pretzner, M., Üstüner, C., Weber, N., Sutter, A., Doublet, V., & Wilfert, L. (2021). Complex relationship between amino acids, fitness and food intake in Bombus terrestris. *Amino Acids*, *53*(10), 1545–1558. https://doi.org/10.1007/S00726-021-03075-8/FIGURES/7

Arstingstall, K. A., DeBano, S. J., Li, X., Wooster, D. E., Rowland, M. M., Burrows, S., & Frost, K. (2021). Capabilities and limitations of using DNA metabarcoding to study plant–pollinator interactions. *Molecular Ecology*, *30*(20), 5266–5297. https://doi.org/10.1111/MEC.16112

Baksay, S., Andalo, C., Galop, D., Burrus, M., Escaravage, N., & Pornon, A. (2022). Using Metabarcoding to Investigate the Strength of Plant-Pollinator Interactions From Surveys

of Visits to DNA Sequences. *Frontiers in Ecology and Evolution*, *10*(March). https://doi.org/10.3389/fevo.2022.735588

Baksay, S., Pornon, A., Burrus, M., Mariette, J., Andalo, C., & Escaravage, N. (2020). Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-61198-6

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 and Evolution*, *3*(3), 363–373. https://doi.org/10.1038/S41559-018-0769-Y

Baloglu, G. H., & Gurel, F. (2015). The effects of pollen protein content on colony development of the bumblebee, bombus terrestris l. *Journal of Apicultural Science*, *59*(1), 83–88. https://doi.org/10.1515/JAS-2015-0009

Bänsch, S., Tscharntke, T., Wünschiers, R., Netter, L., Brenig, B., Gabriel, D., & Westphal, C. (2020). Using ITS2 metabarcoding and microscopy to analyze shifts in pollen diets of honey bees and bumble bees along a mass-flowering crop gradient. *Molecular Ecology*, *September 2019*, 1–16. https://doi.org/10.1111/mec.15675

Barlow, S. E., O'Neill, M. A., & Pavlik, B. M. (2019). A prototype RFID tag for detecting bumblebee visitations within fragmented landscapes. *Journal of Biological Engineering*, *13*(1), 1–6. https://doi.org/10.1186/S13036-019-0143-X/FIGURES/2

Bartomeus, I., Potts, S. G., Steffan-Dewenter, I., Vaissière, B. E., Woyciechowski, M., Krewenka, K. M., Tscheulin, T., Roberts, S. P. M., Szentgyörgyi, H., Westphal, C., & Bommarco, R. (2014). Contribution of insect pollinators to crop yield and quality varies with agricultural intensification. *PeerJ*, *2014*(1), e328. https://doi.org/10.7717/PEERJ.328/SUPP-1

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*(4), 1006–1016. https://doi.org/10.1111/cobi.12536

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*(7588), 85–88. https://doi.org/10.1038/NATURE16532

Beduschi, T., Kormann, U. G., Tscharntke, T., & Scherber, C. (2018). Spatial community turnover of pollinators is relaxed by semi-natural habitats, but not by mass-flowering crops in agricultural landscapes. *Biological Conservation*, *221*, 59–66. https://doi.org/10.1016/j.biocon.2018.01.016

Bell, K. L., Burgess, K. S., Botsch, J. C., Dobbs, E. K., Read, T. D., & Brosi, B. J. (2018). Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology*, *July*. https://doi.org/10.1111/mec.14840

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*(2), 431–455. https://doi.org/10.1111/MEC.14840

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*(9), 629–640. https://doi.org/10.1139/gen-2015-0200

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, K., Gruenewald, D., Lawley, B., Morozumi, C., & Brosi, B. J. (2017). Applying pollen DNA metabarcoding to the study of plant-pollinator interactions. *Applications in Plant Sciences*, *5*(6), 1600124. https://doi.org/10.3732/apps.1600124

Bell, K. L., Petit, R. A., Cutler, A., Dobbs, E. K., Macpherson, J. M., Read, T. D., Burgess, K. S., & Brosi, B. J. (2021). Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. *Ecology and Evolution*, *11*(22), 16082–16098. https://doi.org/10.1002/ECE3.8281

Bell, K. L., Turo, K. J., Lowe, A., Nota, K., Keller, A., Encinas-Viso, F., Parducci, L., Richardson, R. T., Leggett, R. M., Brosi, B. J., Burgess, K. S., Suyama, Y., & de Vere, N. (2022). Plants, pollinators and their interactions under global ecological change: The role of pollen DNA metabarcoding. *Molecular Ecology*, *00*, 1–18. https://doi.org/10.1111/MEC.16689

Benjamin, F. E., & Winfree, R. (2014). Lack of Pollinators Limits Fruit Production in Commercial Blueberry (Vaccinium corymbosum). *Environmental Entomology*, *43*(6), 1574–1583. https://doi.org/10.1603/EN13314

Berger, B. A., Han, J., Sessa, E. B., Gardner, A. G., Shepherd, K. A., Ricigliano, V. A., Jabaily, R. S., & Howarth, D. G. (2017). The Unexpected Depths of Genome-Skimming Data: A Case Study Examining Goodeniaceae Floral Symmetry Genes. *Applications in Plant Sciences*. https://doi.org/10.3732/apps.1700042

Bertrand, C., Eckerter, P. W., Ammann, L., Entling, M. H., Gobet, E., Herzog, F., Mestre, L., Tinner, W., & Albrecht, M. (2019). Seasonal shifts and complementary use of pollen sources by two bees, a lacewing and a ladybeetle species in European agricultural landscapes. *Journal of Applied Ecology*, *56*(11), 2431–2442. https://doi.org/10.1111/1365-2664.13483

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., & Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, *18*(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888

Blaauw, B. R., & Isaacs, R. (2014). Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *Journal of Applied Ecology*, *51*(4), 890–898. https://doi.org/10.1111/1365-2664.12257

Bobiwash, K., Uriel, Y., & Elle, E. (2018). Pollen Foraging Differences Among Three Managed Pollinators in the Highbush Blueberry (*Vaccinium* corymbosum) Agroecosystem. *Journal of Economic Entomology*, *111*(1), 26–32. https://doi.org/10.1093/jee/tox344

Bontšutšnaja, A., Karise, R., Mänd, M., & Smagghe, G. (2021). Bumble bee foraged pollen analyses in spring time in southern estonia shows abundant food sources. *Insects*, *12*(10). https://doi.org/10.3390/INSECTS12100922/S1

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 and Evolution*, *3*(5), 750–754. https://doi.org/10.1038/S41559-019-0849-7

Brittain, C., Kremen, C., & Klein, A.-M. (2013). Biodiversity buffers pollination from changes in environmental conditions. *Global Change Biology*, *19*(2), 540–547. https://doi.org/10.1111/gcb.12043

Brodschneider, R., & Crailsheim, K. (2010). Nutrition and health in honey bees. *Apidologie*, *41*(3), 278–294. https://doi.org/10.1051/apido/2010012

Brown, M. J. F., Schmid-Hempel, R., & Schmid-Hempel, P. (2003). Strong context-dependent virulence in a host–parasite system: reconciling genetic evidence with theory. *Journal of Animal Ecology*, *72*(6), 994–1002. https://doi.org/10.1046/J.1365-2656.2003.00770.X

Buchmann, S. (1983). Buzz Pollination in Angiosperms. *Floral Cues and Animal Behavior*. https://digitalcommons.usu.edu/bee_lab_bo/266

Button, L., & Elle, E. (2014). Wild bumble bees reduce pollination deficits in a crop mostly visited by managed honey bees. *Agriculture, Ecosystems and Environment*, *197*, 255–263. https://doi.org/10.1016/j.agee.2014.08.004

Campbell, A. J., Wilby, A., Sutton, P., & Wäckers, F. L. (2017). Do sown flower strips boost wild pollinator abundance and pollination services in a spring-flowering crop? A case study from UK cider apple orchards. *Agriculture, Ecosystems and Environment*, *239*, 20–29. https://doi.org/10.1016/j.agee.2017.01.005

Carvalheiro, L. G., Seymour, C. L., Nicolson, S. W., & Veldtman, R. (2012). Creating patches of native flowers facilitates crop pollination in large agricultural fields: mango as a case study. *Journal of Applied Ecology*, *49*(6), 1373–1383. https://doi.org/10.1111/j.1365-2664.2012.02217.x

Carvalheiro, L. G., Veldtman, R., Shenkute, A. G., Tesfay, G. B., Pirk, C. W. W., Donaldson, J. S., & Nicolson, S. W. (2011). Natural and within-farmland biodiversity enhances crop productivity. *Ecology Letters*, *14*(3), 251–259. https://doi.org/10.1111/j.1461-0248.2010.01579.x

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*(4), 481–489. https://doi.org/10.1016/J.BIOCON.2006.05.008

Carvell, C., Westrich, P., Meek, W. R., Pywell, R. F., & Nowakowski, M. (2006). Assessing the value of annual and perennial forage mixtures for bumblebees by direct observation and pollen analysis 1. *Apidologie*, *37*, 326–340. https://doi.org/10.1051/apido:2006002

Carvell, C., Bourke, A. F. G., Dreier, S., Freeman, S. N., Hulmes, S., Jordan, W. C., Redhead, J. W., Sumner, S., Wang, J., & Heard, M. S. (2017). Bumblebee family lineage survival is enhanced in high-quality landscapes. *Nature*, *543*(7646), 547–549. https://doi.org/10.1038/NATURE21709

Carvell, C., Meek, W. R., Pywell, R. F., Goulson, D., & Nowakowski, M. (2007). Comparing the efficacy of agri-environment schemes to enhance bumble bee abundance and diversity on arable field margins. *Journal of Applied Ecology*, *44*(1), 29–40. https://doi.org/10.1111/J.1365-2664.2006.01249.X

Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., Dworkin, J. P., Lupisella, M. L., Smith, D. J., Botkin, D. J., Stephenson, T. A., Juul, S., Turner, D. J., Izquierdo, F., Federman, S., Stryke, D., Somasekar, S., Alexander, N., Yu, G., … Burton, A. S. (2016). Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *BioRxiv*, 077651. https://doi.org/10.1101/077651

Chalupowicz, L., Dombrovsky, A., Gaba, V., Luria, N., Reuven, M., Beerman, A., Lachman, O., Dror, O., Nissan, G., & Manulis-Sasson, S. (2019). Diagnosis of plant diseases using the Nanopore sequencing platform. *Plant Pathology*, *68*(2), 229–238. https://doi.org/10.1111/ppa.12957

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). https://doi.org/10.1371/JOURNAL.PONE.0008613

Chittka, L., Gumbert, A., & Kunze, J. (1997). Foraging dynamics of bumble bees: correlates of movements within and between plant species. *Behavioral Ecology*, *8*(3), 239–249. https://doi.org/10.1093/BEHECO/8.3.239

Colwell, M. J., Williams, G. R., Evans, R. C., & Shutler, D. (2017). Honey bee-collected pollen in agro-ecosystems reveals diet diversity, diet quality, and pesticide exposure. *Ecology and Evolution*, *7*. https://doi.org/10.1002/ece3.3178

Corbet, S. A., Fussell, M., Ake, R., Fraser, A., Gunson, C., Savage, A., & Smith, K. (1993). Temperature and the pollinating activity of social bees. *Ecological Entomology*, *18*(1), 17–30. https://doi.org/10.1111/J.1365-2311.1993.TB01075.X

Cornman, R. S., Otto, C. R. v., Iwanowicz, D., & Pettis, J. S. (2015). Taxonomic Characterization of Honey Bee (Apis mellifera) Pollen Foraging Based on Non-Overlapping Paired-End Sequencing of Nuclear Ribosomal Loci. *PLOS ONE*, *10*(12), e0145365. https://doi.org/10.1371/journal.pone.0145365

Cresswell, J. E., Osborne, J. L., & Goulson, D. (2000). An economic model of the limits to foraging range in central place foragers with numerical solutions for bumblebees.

Ecological Entomology, *25*(3), 249–255. https://doi.org/10.1046/J.1365-2311.2000.00264.X

De Luca, P. A., & Vallejo-Marín, M. (2013). What's the 'buzz' about? The ecology and evolutionary significance of buzz-pollination. *Current Opinion in Plant Biology*, *16*(4), 429–435. https://doi.org/10.1016/J.PBI.2013.05.002

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*. https://doi.org/10.1038/SREP42838

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? *Mol. Ecol.*, *28*(2), 391–406. https://doi.org/10.1111/mec.14734

Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLOS ONE*, *16*(10), e0257521. https://doi.org/10.1371/JOURNAL.PONE.0257521

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*(8), e72016. https://doi.org/10.1371/journal.pone.0072016

Dicks, L. v., Baude, M., Roberts, S. P. M., Phillips, J., Green, M., & Carvell, C. (2015). How much flower-rich habitat is enough for wild pollinators? Answering a key policy question with incomplete knowledge. *Ecological Entomology*, *40*(S1), 22–35. https://doi.org/10.1111/een.12226

Dicks, L. V., Breeze, T. D., Ngo, H. T., Senapathi, D., An, J., Aizen, M. A., Basu, P., Buchori, D., Galetto, L., Garibaldi, L. A., Gemmill-Herren, B., Howlett, B. G., Imperatriz-Fonseca, V. L., Johnson, S. D., Kovács-Hostyánszki, A., Kwon, Y. J., Lattorff, H. M. G., Lungharwo, T., Seymour, C. L., … Potts, S. G. (2021). A global-scale expert assessment of drivers and risks associated with pollinator decline. *Nature Ecology & Evolution 2021 5:10*, *5*(10), 1453–1461. https://doi.org/10.1038/s41559-021-01534-9

Dicks, L. V, Viana, B., Bommarco, R., Brosi, B., Del Coro Arizmendi, M., Cunningham, S. A., Galetto, L., Hill, R., Lopes, A. V, Pires, C., Taki, H., & Potts, S. G. (2016). Ten policies for pollinators. *Science*, *354*(6315), 975–976. https://doi.org/10.1126/science.aai9226

Diekötter, T., Peter, F., Jauker, B., Wolters, V., & Jauker, F. (2014). Mass-flowering crops increase richness of cavity-nesting bees and wasps in modern agro-ecosystems. *GCB Bioenergy*, *6*(3), 219–226. https://doi.org/10.1111/gcbb.12080

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*(6), 927–930. https://doi.org/10.1111/J.1654-1103.2003.TB02228.X

Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*, *20*(9), 525–527. https://doi.org/10.1016/j.tplants.2015.06.012

Donkersley, P. (2019). Trees for bees. *Agriculture, Ecosystems and Environment*, *270–271*, 79–83. https://doi.org/10.1016/J.AGEE.2018.10.024

Drake, L. E., Cuff, J. P., Young, R. E., Marchbank, A., Chadwick, E. A., & Symondson, W. O. C. (2022). An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data. *Methods in Ecology and Evolution*, 13, 694–710. https://doi.org/10.1111/2041-210X.13780

Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, *7*(17), 6918–6926. https://doi.org/10.1002/ECE3.3192

Estravis-Barcala, M. C., Palottini, F., Macri, I., Nery, D., & Farina, W. M. (2021). Managed honeybees and South American bumblebees exhibit complementary foraging patterns in highbush blueberry. *Scientific Reports 2021 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-87729-3

FAO. (2009). *Food Insecurity in the World Economic crises-impacts and lessons learned*. http://www.fao.org/3/i0876e/i0876e.pdf

Ferguson, S., McLay, T., Andrew, R. L., Bruhl, J. J., Schwessinger, B., Borevitz, J., & Jones, A. (2022). Species-specific basecallers improve actual accuracy of nanopore sequencing in plants. *Plant Methods*, *18*(1). https://doi.org/10.1186/s13007-022-00971-2

Fijen, T. P. M., & Kleijn, D. (2017). How to efficiently obtain accurate estimates of flower visitation rates by pollinators. *Basic and Applied Ecology*, *19*, 11–18. https://doi.org/10.1016/J.BAAE.2017.01.004

Fijen, T. P. M., Roovers, A., van Deijk, J., & van Grunsven, R. H. A. (2023). Nocturnal pollination is equally important as, and complementary to, diurnal pollination for

strawberry fruit production. *Agriculture, Ecosystems & Environment*, *350*, 108475. https://doi.org/10.1016/J.AGEE.2023.108475

Foulis, E. S. J., & Goulson, D. (2014). Commercial bumble bees on soft fruit farms collect pollen mainly from wildflowers rather than the target crops. *Journal of Apicultural Research*, *53*(3), 404–407. https://doi.org/10.3896/IBRA.1.53.3.08

Fowler, R.E., Rotheray, E.L. and Goulson, D. (2016), Floral abundance and resource quality influence pollinator choice. *Insect Conservation and Diversity*, 9: 481-494. https://doi.org/10.1111/icad.12197

Fox R, Dennis EB, Harrower CA, Blumgart D, Bell JR, Cook P, Davis AM, Evans-Hill LJ, Haynes F, Hill D, Isaac NJB, Parsons MS, Pocock MJO, Prescott T, Randle Z, Shortall CR, Tordoff GM, Tuson D & Bourn NAD (2021) The State of Britain's Larger Moths 2021. Butterfly Conservation, Rothamsted Research and UK Centre for Ecology & Hydrology, Wareham, Dorset, UK.

Fürst, M. A., McMahon, D. P., Osborne, J. L., Paxton, R. J., & Brown, M. J. F. (2014). Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature 2014 506:7488*, *506*(7488), 364–366. https://doi.org/10.1038/nature12977

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*(10), e109363. https://doi.org/10.1371/journal.pone.0109363

Gallai, N., Salles, J. M., Settele, J., & Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, *68*(3), 810–821. https://doi.org/10.1016/j.ecolecon.2008.06.014

Gardner, E., Breeze, T. D., Clough, Y., Smith, H. G., Baldock, K. C. R., Campbell, A., Garratt, M. P. D., Gillespie, M. A. K., Kunin, W. E., McKerchar, M., Potts, S. G., Senapathi, D., Stone, G. N., Wäckers, F., Westbury, D. B., Wilby, A., & Oliver, T. H. (2021). Field boundary features can stabilise bee populations and the pollination of mass-flowering crops in rotational systems. *Journal of Applied Ecology*, *58*(10), 2287–2304. https://doi.org/10.1111/1365-2664.13948

Garibaldi, L. A., Carvalheiro, L. G., Vaissière, B. E., Gemmill-Herren, B., Hipólito, J., Freitas, B. M., Ngo, H. T., Azzu, N., Sáez, A., Åström, J., An, J., Blochtein, B., Buchori, D., García, F. J. C., Silva, F. O. da, Devkota, K., Ribeiro, M. de F., Freitas, L., Gaglianone, M. C., … Zhang,

H. (2016). Mutually beneficial pollinator diversity and crop yield outcomes in small and large farms. *Science*, *351*(6271), 388–391. https://doi.org/10.1126/SCIENCE.AAC7287

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., … Klein, A. M. (2011). Stability of pollination services decreases with isolation from natural areas despite honey bee visits. *Ecology Letters*, *14*(10), 1062–1072. https://doi.org/10.1111/j.1461-0248.2011.01669.x

Garibaldi, L. A., Steffan-Dewenter, I., Winfree, R., Aizen, M. A., Bommarco, R., Cunningham, S. A., Kremen, C., Carvalheiro, L. G., Harder, L. D., Afik, O., Bartomeus, I., Benjamin, F., Boreux, V., Cariveau, D., Chacoff, N. P., Dudenhöffer, J. H., Freitas, B. M., Ghazoul, J., Greenleaf, S., … Klein, A. M. (2013). Wild Pollinators Enhance Fruit Set of Crops Regardless of Honey Bee Abundance. *Science*, *339*(6127), 1608–1611. https://doi.org/10.1126/SCIENCE.1230200

Garratt, M. P. D., Senapathi, D., Coston, D. J., Mortimer, S. R., & Potts, S. G. (2017). The benefits of hedgerows for pollinators and natural enemies depends on hedge quality and landscape context. *Agriculture, Ecosystems and Environment*, *247*, 363–370. https://doi.org/10.1016/J.AGEE.2017.06.048

Garrido-Sanz, L., Senar, M. À., & Piñol, J. (2022). Relative species abundance estimation in artificial mixtures of insects using mito-metagenomics and a correction factor for the mitochondrial DNA copy number. *Molecular Ecology Resources*, *22*(1), 153–167. https://doi.org/10.1111/1755-0998.13464

Gathmann, A., & Tscharntke, T. (2002). Foraging ranges of solitary bees. *Journal of Animal Ecology*, *71*(5), 757–764. https://doi.org/10.1046/J.1365-2656.2002.00641.X

Génissel, A., Aupinel, P., Bressac, C., Tasei, J. N., & Chevrier, C. (2002). Influence of pollen origin on performance of Bombus terrestris micro-colonies. *Entomologia Experimentalis et Applicata*, *104*(2–3), 329–336. https://doi.org/10.1046/J.1570-7458.2002.01019.X

Gibbs, J., Elle, E., Bobiwash, K., Haapalainen, T., & Isaacs, R. (2016). Contrasting pollinators and pollination in native and non-native regions of highbush blueberry production. *PLoS ONE*, *11*(7), 1–24. https://doi.org/10.1371/journal.pone.0158937

Gilbert, J. A., & Dupont, C. L. (2011). Microbial Metagenomics: Beyond the Genome. *Annual Review of Marine Science*, *3*(1), 347–371. https://doi.org/10.1146/annurev-marine-120709-142811

Gill, R. J., & Raine, N. E. (2014). Chronic impairment of bumblebee natural foraging behaviour induced by sublethal pesticide exposure. *Functional Ecology*, *28*(6), 1459–1471. https://doi.org/10.1111/1365-2435.12292

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*(NOV), 2224. https://doi.org/10.3389/FMICB.2017.02224/BIBTEX

Gonçalves, A. T., Collipal-Matamal, R., Valenzuela-Muñoz, V., Nuñez-Acuña, G., Valenzuela-Miranda, D., & Gallardo-Escárate, C. (2020). Nanopore sequencing of microbial communities reveals the potential role of sea lice as a reservoir for fish pathogens. *Scientific Reports 2020 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-59747-0

Goulson, D. (2003). Effects of Introduced Bees on Native Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 1–26. https://doi.org/10.1146/annurev.ecolsys.34.011802.132355

Goulson, D. (2010). *Bumblebees: Behaviour, Ecology, and Conservation*. Oxford University Press, Oxford, UK.

Goulson, D. (2013). An overview of the environmental risks posed by neonicotinoid insecticides. *Journal of Applied Ecology*, *50*(4), 977–987. https://doi.org/10.1111/1365-2664.12111

Goulson, D., Nicholls, E., Botías, C., & Rotheray, E. L. (2015). Bee declines driven by combined Stress from parasites, pesticides, and lack of flowers. *Science*, *347*(6229). https://doi.org/10.1126/SCIENCE.1255957

Grab, H., Blitzer, E. J., Danforth, B., Loeb, G., & Poveda, K. (2017). Temporally dependent pollinator competition and facilitation with mass flowering crops affects yield in co-blooming crops. *Scientific Reports 2017 7:1*, *7*(1), 1–9. https://doi.org/10.1038/srep45296

Greenleaf, S. S., & Kremen, C. (2006). Wild bees enhance honey bees' pollination of hybrid sunflower. *Proceedings of the National Academy of Sciences*, *103*(37), 13890–13895. https://doi.org/10.1073/PNAS.0600929103

Greenleaf, S. S., Williams, N. M., Winfree, R., & Kremen, C. (2007). Bee foraging ranges and their relationship to body size. *Oecologia*, *153*(3), 589–596. https://doi.org/10.1007/S00442-007-0752-9

Gresty, C. E. A., Clare, E., Devey, D. S., Cowan, R. S., Csiba, L., Malakasi, P., Lewis, O. T., & Willis, K. J. (2018). Flower preferences and pollen transport networks for cavity-nesting solitary bees: Implications for the design of agri-environment schemes. *Ecology and Evolution*, *8*(15), 7574–7587. https://doi.org/10.1002/ece3.4234

Grüter, C., & Ratnieks, F. L. W. (2011). Flower constancy in insect pollinators: Adaptive foraging behaviour or cognitive limitation? *Communicative & Integrative Biology*, *4*(6), 633. https://doi.org/10.4161/CIB.16972

Hagen, M., Wikelski, M., & Kissling, W. D. (2011). Space Use of Bumblebees (Bombus spp.) Revealed by Radio-Tracking. *PLOS ONE*, *6*(5), e19997. https://doi.org/10.1371/JOURNAL.PONE.0019997

Hanley, M. E., Franco, M., Pichon, S., Goulson, D., Franco1, M., Pich, S., N2, ?, Darvill3, B., & Goulson3, D. (2008). Breeding System, Pollinator Choice and Variation in Pollen Quality in British Herbaceous Plants. *Functional Ecology*, *22*(4), 592–598.

Harrison, R. E., Luby, J. J., & Ascher, P. D. (1993). Pollen source affects yield components and reproductive fertility of four half-high blueberry cultivars. *Journal of American Society for Horticultural Science*, *119*(1), 84–89.

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*(8), 1–20. https://doi.org/10.1371/journal.pone.0134735

He, Y., Tsutsui, M., Zhou, Y., & Miao, X. S. (2021). Solid-state nanopore systems: from materials to applications. In *NPG Asia Materials* (Vol. 13, Issue 1). Nature Research. https://doi.org/10.1038/s41427-021-00313-z

Heavens, D., Chooneea, D., Giolai, M., Cuber, P., Aanstad, P., Martin, S., Alston, M., Misra, R., Clark, M. D., & Leggett, R. M. (2021). How low can you go? Driving down the DNA input requirements for nanopore sequencing. *BioRxiv*. https://doi.org/10.1101/2021.10.15.464554

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Heinrich, B. (1976). The Foraging Specializations of Individual Bumblebees. *Ecological Monographs*, *46*(2), 105–128. https://doi.org/10.2307/1942246

Hemberger, J. A., Rosenberger, N. M., & Williams, N. M. (2023). Experimental heatwaves disrupt bumblebee foraging through direct heat effects and reduced nectar production. *Functional Ecology*, *37*(3), 591–601. https://doi.org/10.1111/1365-2435.14241

Hicks, D. M., Ouvrard, P., Baldock, K. C. R., Baude, M., Goddard, M. A., Kunin, W. E., Mitschunas, N., Memmott, J., Morse, H., Nikolitsi, M., Osgathorpe, L. M., Potts, S. G., Robertson, K. M., Scott, A. V., Sinclair, F., Westbury, D. B., & Stone, G. N. (2016). Food for pollinators: Quantifying the nectar and pollen resources of urban flower meadows. *PLoS ONE*, *11*(6). https://doi.org/10.1371/JOURNAL.PONE.0158117

Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *108*(49), 19451–19452. https://doi.org/10.1073/PNAS.1116812108

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019254

Hollingsworth, P. M., Li, D. Z., 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*(1702), 20150338. https://doi.org/10.1098/rstb.2015.0338

Holzschuh, A., Dormann, C. F., Tscharntke, T., & Steffan-Dewenter, I. (2011). Expansion of mass-flowering crops leads to transient pollinator dilution and reduced wild plant pollination. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1723), 3444–3451. https://doi.org/10.1098/RSPB.2011.0268

Holzschuh, A., Dormann, C. F., Tscharntke, T., & Steffan-Dewenter, I. (2013). Mass-flowering crops enhance wild bee abundance. *Oecologia*, *172*(2), 477–484. https://doi.org/10.1007/s00442-012-2515-5

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. https://doi.org/10.1111/2041-210X.12613

Hutchinson, L. A., Oliver, T. H., Breeze, T. D., Bailes, E. J., Brünjes, L., Campbell, A. J., Erhardt, A., de Groot, G. A., Földesi, R., García, D., Goulson, D., Hainaut, H., Hambäck, P. A., Holzschuh, A., Jauker, F., Klatt, B. K., Klein, A. M., Kleijn, D., Kovács-Hostyánszki, A., … Garratt, M. P. D. (2021). Using ecological and field survey data to establish a national list of the wild bee pollinators of crops. *Agriculture, Ecosystems and Environment*, *315*. https://doi.org/10.1016/j.agee.2021.107447

Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., Leggett, R. M., Eccles, D. A., Zalunin, V., Urban, J. M., Piazza, P., Bowden, R. J., Paten, B., Mwaigwisya, S., Batty, E. M., Simpson, J. T., Snutch, T. P., Birney, E., Buck, D., … Olsen, H. E. (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research 2015 4:1075*, *4*, 1075. https://doi.org/10.12688/f1000research.7201.1

IPBES. (2016). *The assessment report on pollinators, pollination and food production*. http://nora.nerc.ac.uk/id/eprint/514356/

Isaacs, R., & Kirk, A. K. (2010). Pollination services provided to small and large highbush blueberry fields by wild and managed bees. *Journal of Applied Ecology*, *47*(4), 841–849. https://doi.org/10.1111/J.1365-2664.2010.01823.X

Jain, M., Tyson, J. R., Loose, M., Ip, C. L. C., Eccles, D. A., O'Grady, J., Malla, S., Leggett, R. M., Wallerman, O., Jansen, H. J., Zalunin, V., Birney, E., Brown, B. L., Snutch, T. P., Olsen, H. E., & Consortium, M. A. and R. (2017). MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Research*, *6*. https://doi.org/10.12688/f1000research.11354.1

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., … Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345. https://doi.org/10.1038/nbt.4060

Jauker, F., Peter, F., Wolters, V., & Diekötter, T. (2012). Early reproductive benefits of mass-flowering crops to the solitary bee Osmia rufa outbalance post-flowering disadvantages. *Basic and Applied Ecology*, *13*, 268–276. https://doi.org/10.1016/j.baae.2012.03.010

Javorek, S. K., Mackenzie, K. E., & Kloet, S. P. Vander. (2002). Comparative Pollination Effectiveness Among Bees (Hymenoptera: Apoidea) on Lowbush Blueberry (Ericaceae:

Vaccinium angustifolium). *Annals of the Entomological Society of America*, *95*(3), 345–351. https://academic.oup.com/aesa/article/95/3/345/82511

Jha, S., Stefanovich, L., & Kremen, C. (2013). Bumble bee pollen use and preference across spatial scales in human-altered landscapes. *Ecological Entomology*, *38*(6), 570–579. https://doi.org/10.1111/een.12056

Ji, Y., Ashton, L., Pedley, S. M., David, ‡, Edwards, P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*, 1245–1257. https://doi.org/10.1111/ele.12162

Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., & Tighe, S. W. (2017). Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *Journal of Biomolecular Techniques*, *28*(1), 2–7. https://doi.org/10.7171/jbt.17-2801-009

Jones, L., Brennan, G. L., Lowe, A., Creer, S., Ford, C. R., & de Vere, N. (2021). Shifts in honeybee foraging reveal historical changes in floral resources. *Communications Biology*, *4*(1). https://doi.org/10.1038/S42003-020-01562-4

Jones, L., Twyford, A. D., Ford, C. R., Rich, T. C. G., Davies, H., Forrest, L. L., Hart, M. L., McHaffie, H., Brown, M. R., Hollingsworth, P. M., & de Vere, N. (2021). Barcode UK: A complete DNA barcoding resource for the flowering plants and conifers of the United Kingdom. *Molecular Ecology Resources*, *21*(6), 2050–2062. https://doi.org/10.1111/1755-0998.13388

Kamo, T., Kusumoto, Y., Tokuoka, Y., Okubo, S., Hayakawa, H., Yoshiyama, M., Kimura, K., & Konuma, A. (2018). A DNA barcoding method for identifying and quantifying the composition of pollen species collected by European honeybees, Apis mellifera (Hymenoptera: Apidae). *Applied Entomology and Zoology*, *53*(3), 353–361. https://doi.org/10.1007/s13355-018-0565-9

Kämper, W., Werner, P. K., Hilpert, A., Westphal, C., Blüthgen, N., Eltz, T., & Leonhardt, S. D. (2016). How landscape, pollen intake and pollen quality affect colony growth in Bombus terrestris. *Landscape Ecology*, *31*(10), 2245–2258. https://doi.org/10.1007/S10980-016-0395-5/FIGURES/4

Kearns, C. A., Inouye, D. W., & Waser, N. M. (1998). Endangered Mutualisms: The Conservation of Plant-Pollinator Interactions. *Annual Review of Ecology and Systematics*, *29*, 83–112. www.annualreviews.org

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., Rost, S., Härtel, S., & Steffan-Dewenter, I. (2015). Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology*, *17*(2), 558–566. https://doi.org/10.1111/PLB.12251

Kembel, S. W., Wu, M., Eisen, J. A., & Green, J. L. (2012). Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Computational Biology*, *8*(10). https://doi.org/10.1371/JOURNAL.PCBI.1002743

Kendall, L. K., Gagic, V., Evans, L. J., Cutting, B. T., Scalzo, J., Hanusch, Y., Jones, J., Rocchetti, M., Sonter, C., Keir, M., & Rader, R. (2020). Self-compatible blueberry cultivars require fewer floral visits to maximize fruit production than a partially self-incompatible cultivar. *Journal of Applied Ecology*, *57*(12), 2454–2462. https://doi.org/10.1111/1365-2664.13751

Kennedy, C. M., Lonsdorf, E., Neel, M. C., Williams, N. M., Ricketts, T. H., Winfree, R., Bommarco, R., Brittain, C., Burley, A. L., Cariveau, D., Carvalheiro, L. G., Chacoff, N. P., Cunningham, S. A., Danforth, B. N., Dudenhöffer, J. H., Elle, E., Gaines, H. R., Garibaldi, L. A., Gratton, C., … Kremen, C. (2013). A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. *Ecology Letters*, *16*(5), 584–599. https://doi.org/10.1016/j.joca.2015.07.015

Khansari, E., Zarre, S., Alizadeh, K., Attar, F., Aghabeigi, F., & Salmaki, Y. (2012a). Pollen morphology of Campanula (Campanulaceae) and allied genera in Iran with special focus on its systematic implication. *Flora: Morphology, Distribution, Functional Ecology of Plants*, *207*(3), 203–211. https://doi.org/10.1016/J.FLORA.2012.01.006

Khansari, E., Zarre, S., Alizadeh, K., Attar, F., Aghabeigi, F., & Salmaki, Y. (2012b). Pollen morphology of Campanula (Campanulaceae) and allied genera in Iran with special focus on its systematic implication. *Flora - Morphology, Distribution, Functional Ecology of Plants*, *207*(3), 203–211. https://doi.org/10.1016/J.FLORA.2012.01.006

Kim, H. S., Jeon, S., Kim, C., Kim, Y. K., Cho, Y. S., Kim, J., Blazyte, A., Manica, A., Lee, S., & Bhak, J. (2019). Chromosome-scale assembly comparison of the Korean Reference

Genome KOREF from PromethION and PacBio with Hi-C mapping information. *GigaScience*, *8*(12). https://doi.org/10.1093/gigascience/giz125

Kleijn, D., & Raemakers, I. (2008). A retrospective analysis of pollen host plant use by stable and declining bumble bee species. *Ecology*, *89*(7), 1811–1823. https://doi.org/10.1890/07-1275.1

Kleijn, D., Winfree, R., Bartomeus, I., Carvalheiro, L. G., Henry, M., Isaacs, R., Klein, A. M., Kremen, C., M'Gonigle, L. K., Rader, R., Ricketts, T. H., Williams, N. M., Lee Adamson, N., Ascher, J. S., Báldi, A., Batáry, P., Benjamin, F., Biesmeijer, J. C., Blitzer, E. J., … Potts, S. G. (2015). Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nature Communications*, *6*. https://doi.org/10.1038/NCOMMS8414

Klein, A. M., Steffan-Dewenter, I., & Tscharntke, T. (2003). Fruit set of highland coffee increases with the diversity of pollinating bees. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1518), 955–961. https://doi.org/10.1098/rspb.2002.2306

Klein, A. M., Vaissière, 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*(1608), 303–313. https://doi.org/10.1098/RSPB.2006.3721

Koh, I., Lonsdorf, E. v, Williams, N. M., Brittain, C., Isaacs, R., Gibbs, J., & Ricketts, T. H. (2016). Modeling the status, trends, and impacts of wild bee abundance in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 140–145. https://doi.org/10.1073/pnas.1517685113

Kovács-Hostyánszki, A., Haenke, S., Batáry, P., Jauker, B., Báldi, A., Tscharntke, T., & Holzschuh, A. (2013). Contrasting effects of mass-flowering crops on bee pollination of hedge plants at different spatial and temporal scales. *Ecological Applications*, *23*(8), 1938–1946. https://doi.org/10.1890/12-2012.1

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*(1), 8–16. https://doi.org/10.1111/1755-0998.12288

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod

metabarcoding. *Scientific Reports*, *7*(1), 17668. https://doi.org/10.1038/s41598-017-17333-x

Kremen, C., Williams, N. M., Aizen, M. A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., Packer, L., Potts, S. G., Roulston, T., Steffan-Dewenter, I., Vázquez, D. P., Winfree, R., Adams, L., Crone, E. E., Greenleaf, S. S., Keitt, T. H., Klein, A.-M., Regetz, J., & Ricketts, T. H. (2007). Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. *Ecology Letters*, *10*(4), 299–314. https://doi.org/10.1111/j.1461-0248.2007.01018.x

Kremen, C., Williams, N. M., & Thorp, R. W. (2002). Crop pollination from native bees at risk from agricultural intensification. *Proceedings of the National Academy of Sciences*, *99*(26), 16812–16816. https://doi.org/10.1073/pnas.262413599

Kress, W. J. (2017). Plant DNA barcodes: Applications today and in the future. *Journal of Systematics and Evolution*, *55*(4), 291–307. https://doi.org/10.1111/jse.12254

Kriesell, L., Hilpert, A., & Leonhardt, S. D. (2017). Different but the same: bumblebee species collect pollen of different plant sources but similar amino acid profiles. *Apidologie*, *48*(1), 102–116. https://doi.org/10.1007/S13592-016-0454-6

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2018). How quantitative is metabarcoding: a meta-analytical approach. *Molecular Ecology*, *May*, 1–11. https://doi.org/10.1111/mec.14920

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*(2), 420–430. https://doi.org/10.1111/MEC.14920

Lang, D., Tang, M., Hu, J., & Zhou, X. (2019). Genome-skimming provides accurate quantification for pollen mixtures. *Molecular Ecology Resources*, *19*(6), 1433–1446. https://doi.org/10.1111/1755-0998.13061

Lau, P., Bryant, V., & Rangel, J. (2017). *Determining the minimum number of pollen grains needed for accurate honey bee (Apis mellifera) colony pollen pellet analysis*. https://doi.org/10.1080/01916122.2017.1306810

Lautenbach, S., Seppelt, R., Liebscher, J., & Dormann, C. F. (2012). Spatial and Temporal Trends of Global Pollination Benefit. *PLoS ONE*, *7*(4), e35954. https://doi.org/10.1371/journal.pone.0035954

Laver, T., Harrison, J., O'neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification*, *3*, 1–8. https://doi.org/10.1016/j.bdq.2015.02.001

LeBuhn, G., & Vargas Luna, J. (2021). Pollinator decline: what do we know about the drivers of solitary bee declines? In *Current Opinion in Insect Science* (Vol. 46, pp. 106–111). Elsevier Inc. https://doi.org/10.1016/j.cois.2021.05.004

Leggett, R. M., & Clark, M. D. (2017). A world of opportunities with nanopore sequencing. *Journal of Experimental Botany*, *68*(20), 5419–5429. https://doi.org/10.1093/jxb/erx289

Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D., & Caccamo, M. (2014). *Sequence analysis NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries*. *30*(4), 566–568. https://doi.org/10.1093/bioinformatics/btt702

Leidenfrost, R. M., Bänsch, S., Prudnikow, L., Brenig, B., Westphal, C., & Wünschiers, R. (2020). Analyzing the Dietary Diary of Bumble Bee. *Frontiers in Plant Science*, *11*(March), 1–9. https://doi.org/10.3389/fpls.2020.00287

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Package 'emmeans' Type Package Title Estimated Marginal Means, aka Least-Squares Means*. https://doi.org/10.1080/00031305.1980.10483031

Leonhardt, S. D., & Blüthgen, N. (2012). The same, but different: Pollen foraging in honeybee and bumblebee colonies. *Apidologie*, *43*(4), 449–464. https://doi.org/10.1007/s13592-011-0112-y

Leontidou, K., Vernesi, C., De Groeve, J., Cristofolini, F., Vokou, D., & Cristofori, A. (2018). DNA metabarcoding of airborne pollen: new protocols for improved taxonomic identification of environmental samples. *Aerobiologia*, *34*. https://doi.org/10.1007/s10453-017-9497-z

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., … Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), 4325–4333. https://doi.org/10.1073/PNAS.1720115115

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Project, G., & Subgroup, D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lihoreau, M., Charleston, M. A., Senior, A. M., Clissold, F. J., Raubenheimer, D., Simpson, S. J., & Buhl, J. (2017). Collective foraging in spatially complex nutritional environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1727). https://doi.org/10.1098/RSTB.2016.0238

Lobaton, J., Andrew, R., Duitama, J., Kirkland, L., Macfadyen, S. & Rader, R. (2021) Using RNA-seq to characterize pollen–stigma interactions for pollination studies. *Scientific Reports, 11*(6635). https://doi.org/10.1038/s41598-021-85887-y

Lonsdorf, E., Kremen, C., Ricketts, T., Winfree, R., Williams, N., & Greenleaf, S. (2009). Modelling pollination services across agricultural landscapes. *Annals of Botany*, *103*(9), 1589–1600. https://doi.org/10.1093/aob/mcp069

Lowe, A., Jones, L., Brennan, G., Creer, S., & de Vere, N. (2022). Seasonal progression and differences in major floral resource use by bees and hoverflies in a diverse horticultural and agricultural landscape revealed by DNA metabarcoding. *Journal of Applied Ecology*, *59*(6), 1484–1495. https://doi.org/10.1111/1365-2664.14144

Lowenstein, D.M., Matteson, K.C. & Minor, E.S. Evaluating the dependence of urban pollinators on ornamental, non-native, and 'weedy' floral resources. (2019) *Urban Ecosystems. 22*, 293–302. https://doi.org/10.1007/s11252-018-0817-z

Lucas, A., Bodger, O., Brosi, B. J., Ford, C. R., Forman, D. W., Greig, C., Hegarty, M., Neyland, P. J., & de Vere, N. (2018). Generalisation and specialisation in hoverfly (Syrphidae) grassland pollen transport networks revealed by DNA metabarcoding. *Journal of Animal Ecology*, *87*(4), 1008–1021. https://doi.org/10.1111/1365-2656.12828

Macgregor, C. J., Kitson, J. J. N., Fox, R., Hahn, C., Lunt, D. H., Pocock, M. J. O., & Evans, D. M. (2018). Construction, validation, and application of nocturnal pollen transport networks in an agro-ecosystem: a comparison using light microscopy and DNA metabarcoding. *Ecological Entomology*, 17–29. https://doi.org/10.1111/een.12674

Mallinger, R. E., & Gratton, C. (2015). Species richness of wild bees, but not the use of managed honeybees, increases fruit set of a pollinator-dependent crop. *Journal of Applied Ecology*, *52*, 323–330. https://doi.org/10.1111/1365-2664.12377

Mallinger, R. E., Gaines-Day, H. R., & Gratton, C. (2017). Do managed bees have negative effects on wild bees?: A systematic review of the literature. *PLoS ONE*, *12*(12). https://doi.org/10.1371/JOURNAL.PONE.0189268

Mander, L., Baker, S. J., Belcher, C. M., Haselhorst, D. S., Rodriguez, J., Thorn, J. L., Tiwari, S., Urrego, D. H., Wesseln, C. J., & Punyasena, S. W. (2014). Accuracy and consistency of grass pollen identification by human analysts using electron micrographs of surface ornamentation. *Applications in Plant Sciences*, *2*(8). https://doi.org/10.3732/apps.1400031

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12. http://www-huber.embl.de/users/an-

Martinet, B., Lecocq, T., Smet, J., & Rasmont, P. (2015). A Protocol to Assess Insect Resistance to Heat Waves, Applied to Bumblebees (Bombus Latreille, 1802). *PLOS ONE*, *10*(3), e0118591. https://doi.org/10.1371/JOURNAL.PONE.0118591

Martinet, B., Zambra, E., Przybyla, K., Lecocq, T., Anselmo, A., Nonclercq, D., Rasmont, P., Michez, D., & Hennebert, E. (2021). Mating under climate change: Impact of simulated heatwaves on the reproduction of model pollinators. *Functional Ecology*, *35*(3), 739–752. https://doi.org/10.1111/1365-2435.13738/SUPPINFO

Marzinzig, B., Brünjes, L., Biagioni, S., Behling, H., Link, W., & Westphal, C. (2018). Bee pollinators of faba bean (Vicia faba L.) differ in their foraging behaviour and pollination efficiency. *Agriculture, Ecosystems & Environment*, *264*, 24–33. https://doi.org/10.1016/J.AGEE.2018.05.003

Mauriello, G., De Prisco, A., Di Prisco, G., La Storia, A., & Caprio, E. (2017) Microbial characterization of bee pollen from the Vesuvius area collected by using three different traps. *PLoS ONE 12*(9): e0183208. https://doi.org/10.1371/journal.pone.0183208

McGhie, T. K., Ainge, G. D., Barnett, L. E., Cooney, J. M., & Jensen, D. J. (2003). Anthocyanin Glycosides from Berry Fruit Are Absorbed and Excreted Unmetabolized by Both Humans and Rats. *Journal of Agricultural and Food Chemistry*, *51*, 4539–4548. https://doi.org/10.1021/jf026206w

Memmott, J. (1999). The structure of a plant-pollinator food web. *Ecology Letters*, *2*(5), 276–280. https://doi.org/10.1046/J.1461-0248.1999.00087.X

Memmott, J., Craze, P. G., Waser, N. M., & Price, M. v. (2007). Global warming and the disruption of plant–pollinator interactions. *Ecology Letters*, *10*(8), 710–717. https://doi.org/10.1111/J.1461-0248.2007.01061.X

Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., Bernardi, M., Xumerle, L., Loader, S., & Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS ONE*, *12*(10).

M'Gonigle, L. K., Ponisio, L. C., Cutler, K., & Kremen, C. (2015). Habitat restoration promotes pollinator persistence and colonization in intensively managed agriculture. *Ecological Applications*, *25*(6), 1557–1565. https://doi.org/10.1890/14-1863.1

Moerman, R., Roger, N., De Jonghe, R., Michez, D., & Vanderplanck, M. (2016). Interspecific Variation in Bumblebee Performance on Pollen Diet: New Insights for Mitigation Strategies. *PLOS ONE*, *11*(12), e0168462. https://doi.org/10.1371/JOURNAL.PONE.0168462

Moerman, R., Vanderplanck, M., Fournier, D., Jacquemart, A. L., & Michez, D. (2017). Pollen nutrients better explain bumblebee colony development than pollen diversity. *Insect Conservation and Diversity*, *10*(2), 171–179. https://doi.org/10.1111/icad.12213

Moerman, R., Vanderplanck, M., Roger, N., Declèves, S., Wathelet, B., Rasmont, P., Fournier, D., & Michez, D. (2016). Growth Rate of Bumblebee Larvae is Related to Pollen Amino Acids. *Journal of Economic Entomology*, *109*(1), 25–30. https://doi.org/10.1093/JEE/TOV279

Moore, P. D., Webb, J. A., & Collinson, M. E. (1991). *Pollen Analysis*. Blackwell Scientific Publications, Oxford, UK.

Morandin, L. A., & Kremen, C. (2013). Hedgerow restoration promotes pollinator populations and exports native bees to adjacent fields. *Ecological Applications*, *23*(4), 829–839. https://doi.org/10.1890/12-1051.1

Murray, T. E., Coffey, M. F., Kehoe, E., & Horgan, F. G. (2013). Pathogen prevalence in commercially reared bumble bees and evidence of spillover in conspecific populations. *Biological Conservation*, *159*, 269–276. https://doi.org/10.1016/J.BIOCON.2012.10.021

Nichols, R. N., Goulson, D., & Holland, J. M. (2019). The best wildflowers for wild bees. *Journal of Insect Conservation*, *23*(5–6), 819–830. https://doi.org/10.1007/s10841-019-00180-8

Nicholson, C. C., & Ricketts, T. H. (2018). *Wild pollinators improve production, uniformity, and timing of blueberry crops*. https://doi.org/10.1016/j.agee.2018.10.018

Nicholson, C. C., Ricketts, T. H., Koh, I., Smith, H. G., Lonsdorf, E. v., & Olsson, O. (2019). Flowering resources distract pollinators from crops: Model predictions from landscape simulations. *Journal of Applied Ecology*, *56*(3), 618–628. https://doi.org/10.1111/1365-2664.13333

Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C., & Jordano, P. (2011). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1706), 725–732. https://doi.org/10.1098/RSPB.2010.1371

Ollerton, J., Erenler, H., Edwards, M., & Crockett, R. (2014). Pollinator declines. Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. *Science (New York, N.Y.)*, *346*(6215), 1360–1362. https://doi.org/10.1126/science.1257259

Osborne, J. L., Clark, S. J., Morris, R. J., Williams, I. H., Riley, J. R., Smith, A. D., Reynolds, D. R., & Edwards, A. S. (1999). A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar. *Journal of Applied Ecology*, *36*(4), 519–533. https://doi.org/10.1046/J.1365-2664.1999.00428.X

Osborne, J. L., Martin, A. P., Carreck, N. L., Swain, J. L., Knight, M. E., Goulson, D., Hale, R. J., & Sanderson, R. A. (2008). Bumblebee flight distances in relation to the forage landscape. *Journal of Animal Ecology*, *77*(2), 406–415. https://doi.org/10.1111/J.1365-2656.2007.01333.X

Osborne, J. L., Smith, A., Clark, S. J., Reynolds, D. R., Barron, M. C., Lim, K. S., & Reynolds, A. M. (2013). The Ontogeny of Bumblebee Flight Trajectories: From Naïve Explorers to Experienced Foragers. *PLOS ONE*, *8*(11), e78681. https://doi.org/10.1371/JOURNAL.PONE.0078681

Patrizia, W. K., Andrea, K. W., Catrin, H., Nico, W., Thomas, B. T., Sara, E., Leonhardt, D., Kämper, W., Werner, P. K., Eltz, Á. T., Hilpert, A., Blüthgen, Á. N., & Leonhardt, S. D. (2016). How landscape, pollen intake and pollen quality affect colony growth in Bombus terrestris. *Landscape Ecology*, *31*, 2245–2258. https://doi.org/10.1007/s10980-016-0395-5

Paula, D. P., Linard, B., Andow, D. A., Sujii, E. R., Pires, C. S. S., & Vogler, A. P. (2015). Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics. *Molecular Ecology Resources*, *15*(4), 880–892. https://doi.org/10.1111/1755-0998.12364

Pearman, W. S., Freed, N. E., & Silander, O. K. (2020). Testing the advantages and disadvantages of short- And long- read eukaryotic metagenomics using simulated reads. *BMC Bioinformatics*, *21*(1). https://doi.org/10.1186/s12859-020-3528-4

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*, *10*(10), 1690–1701. https://doi.org/10.1111/2041-210X.13265

Pełka, K, Worobo, R.W., Walkusz, J., & Szweda, P. (2021). Bee Pollen and Bee Bread as a Source of Bacteria Producing Antimicrobials. *Antibiotics*, *10*(6):713. https://doi.org/10.3390/antibiotics10060713

Pellicer, J., Fls, M. F. F., & Fls, I. J. L. (2010). *The largest eukaryotic genome of them all?b oj_1072 10..15*. https://academic.oup.com/botlinnean/article/164/1/10/2418538

Pellicer, J., & Leitch, I. J. (2020). The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist*, *226*(2), 301–305. https://doi.org/10.1111/NPH.16261

Pendegraft, A. H., Guo, B., & Yi, N. (2019). Bayesian hierarchical negative binomial models for multivariable analyses with applications to human microbiome count data. *PLOS ONE*, *14*(8), e0220961. https://doi.org/10.1371/JOURNAL.PONE.0220961

Persson, A. S., Mazier, F., & Smith, H. G. (2018). When beggars are choosers—How nesting of a solitary bee is affected by temporal dynamics of pollen plants in the landscape. *Ecology and Evolution*, *8*(11), 5777–5791. https://doi.org/10.1002/ece3.4116

Piko, J., Keller, A., Geppert, C., Batáry, P., Tscharntke, T., Westphal, C., & Hass, A. L. (2021). Effects of three flower field types on bumblebees and their pollen diets. *Basic and Applied Ecology*, *52*. https://doi.org/10.1016/j.baae.2021.02.005

Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, *15*(4), 819–830. https://doi.org/10.1111/1755-0998.12355

Polling, M., Sin, M., de Weger, L. A., Speksnijder, A. G. C. L., Koenders, M. J. F., de Boer, H., & Gravendeel, B. (2022). DNA metabarcoding using nrITS2 provides highly qualitative and quantitative results for airborne pollen monitoring. *Science of The Total Environment*, *806*, 150468. https://doi.org/10.1016/J.SCITOTENV.2021.150468

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., Barrio-Amorós, C. L., Salazar-Valenzuela, D., & Prost, S. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*, *7*(4). https://doi.org/10.1093/gigascience/giy033

Pornon, A., Andalo, C., Burrus, M., & Escaravage, N. (2017). DNA metabarcoding data unveils invisible pollination networks. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-16785-5

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*(February), 1–12. https://doi.org/10.1038/srep27282

Potter, C., de Vere, N., Jones, L. E., 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*, *2019*(2). https://doi.org/10.7717/PEERJ.5999

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 and Evolution*, *25*(6), 345–353. https://doi.org/10.1016/J.TREE.2010.01.007

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*(7632), 220–229. https://doi.org/10.1038/nature20588

Powney, G. D., Carvell, C., Edwards, M., Morris, R. K. A., Roy, H. E., Woodcock, B. A., & Isaac, N. J. B. (2019). Widespread losses of pollinating insects in Britain. *Nature Communications*, *10*(1038). https://doi.org/10.1038/s41467-019-08974-9

Prior, R. L., Cao, G., Martin, A., Sofic, E., Mcewen, J., O'brien, C., Lischner, N., Ehlenfeldt, M., Kalt, W., Krewer, G., & Mike, C. (1998). Antioxidant Capacity As Influenced by Total

Phenolic and Anthocyanin Content, Maturity, and Variety of Vaccinium Species. *Journal of Agricultural and Food Chemistry*, *46*, 2686–2693.

Pywell, R. F., Meek, W. R., Hulmes, L., Hulmes, S., James, K. L., Nowakowski, M., & Carvell, C. (2011). Management to enhance pollen and nectar resources for bumblebees and butterflies within intensively farmed landscapes. *Journal of Insect Conservation*, *15*, 853–864. https://doi.org/10.1007/s10841-011-9383-x

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., … Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*(7589), 228–232. https://doi.org/10.1038/nature16996

Rahl, M. (2008). Microscopic Identification and Purity Determination of Pollen Grains. In *Methods in molecular medicine* (Vol. 138, pp. 263–269). https://doi.org/10.1007/978-1-59745-366-0_22

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. In *Genome Biology* (Vol. 19, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13059-018-1462-9

Rasmont, P., & Iserbyt, S. (2012). The bumblebees scarcity syndrome: Are heat waves leading to local extinctions of bumblebees (Hymenoptera: Apidae: Bombus)? *Annales de La Societe Entomologique de France*, *48*(3–4), 275–280. https://doi.org/10.1080/00379271.2012.10697776

Ravoet, J., de Smet, L., Meeus, I., Smagghe, G., Wenseleers, T., & de Graaf, D. C. (2014). Widespread occurrence of honey bee pathogens in solitary bees. *Journal of Invertebrate Pathology*, *122*, 55–58. https://doi.org/10.1016/J.JIP.2014.08.007

Reddington, K., Eccles, D., O'Grady, J., Drown, D. M., Hansen, L. H., Nielsen, T. K., Ducluzeau, A.-L., Leggett, R. M., Heavens, D., Peel, N., Snutch, T. P., Bayega, A., Oikonomopoulos, S., Ragoussis, J., Barry, T., Van Der Helm, E., Jolic, D., Richardson, H., Jansen, H., … Brown, B. L. (2020). Metagenomic analysis of planktonic riverine microbial consortia using nanopore sequencing reveals insight into river microbe taxonomy and function Metagenomic analysis of planktonic riverine microbial consortia using nanopore sequencing. *GigaScience*, *9*, 1–12. https://doi.org/10.1093/gigascience/giaa053

Redhead, J. W., Dreier, S., Bourke, A. F. G., Heard, M. S., Jordan, W. C., Sumner, S., Wang, J., & Carvell, C. (2016). Effects of habitat composition and landscape structure on worker foraging distances of five bumble bee species. *Ecological Applications*, *26*(3), 726–739. https://doi.org/10.1890/15-0546

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*(4), 881–890. https://doi.org/10.1890/14-1011.1

Requier, F., & Leonhardt, S. D. (2020). Beyond flowers: including non-floral resources in bee conservation schemes. *Journal of Insect Conservation*, *24*, 5–16. https://doi.org/10.1007/s10841-019-00206-1

Richardson, R. T., Curtis, H. R., Matcham, E. G., Lin, C. H., Suresh, S., Sponsler, D. B., Hearon, L. E., & Johnson, R. M. (2019). Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. *Molecular Ecology*, *28*(3), 686–697. https://doi.org/10.1111/mec.14975

Richardson, R. T., Eaton, T. D., Lin, C. H., Cherry, G., Johnson, R. M., & Sponsler, D. B. (2021). Application of plant metabarcoding to identify diverse honeybee pollen forage along an urban–agricultural gradient. *Molecular Ecology*, *30*(1), 310–323. https://doi.org/10.1111/MEC.15704

Richardson, R. T., Lin, C., Quijia, J. O., Riusech, N. S., Goodell, K., & Johnson, R. M. (2015a). Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in Plant Sciences*, *3*(11), 1500043. https://doi.org/10.3732/APPS.1500043

Richardson, R. T., Lin, C.-H., Sponsler, D. B., Quijia, J. O., Goodell, K., & Johnson, R. M. (2015b). Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. *Applications in Plant Sciences*, *3*(1), 1400066. https://doi.org/10.3732/apps.1400066

Ricketts, T. H., Regetz, J., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., Bogdanski, A., Gemmill-Herren, B., Greenleaf, S. S., Klein, A. M., Mayfield, M. M., Morandin, L. A., Ochieng', A., & Viana, B. F. (2008). Landscape effects on crop pollination services: are

there general patterns? *Ecology Letters*, *11*(5), 499–515. https://doi.org/10.1111/j.1461-0248.2008.01157.x

Riedinger, V., Renner, M., Rundlöf, M., Steffan-Dewenter, I., & Holzschuh, A. (2014). Early mass-flowering crops mitigate pollinator dilution in late-flowering crops. *Landscape Ecology*, *29*(3), 425–435. https://doi.org/10.1007/s10980-013-9973-y

Riley, J. R., Smith, A. D., Reynolds, D. R., Edwards, A. S., Osborne, J. L., Williams, I. H., Carreck, N. L., & Poppy, G. M. (1993). Wireless World May 117-122. *J. Chem. Ecol*, *19*, 53–88.

Roger, N., Michez, D., Wattiez, R., Sheridan, C., & Vanderplanck, M. (2017). Diet effects on bumblebee health. *Journal of Insect Physiology*, *96*, 128–133. https://doi.org/10.1016/J.JINSPHYS.2016.11.002

Rogers, S. R., Tarpy, D. R., & Burrack, H. J. (2014). Bee Species Diversity Enhances Productivity and Stability in a Perennial Crop. *PLoS ONE*, *9*(5), e97307. https://doi.org/10.1371/journal.pone.0097307

Rollin, O., Bretagnolle, V., Decourtye, A., Aptel, J., Michel, N., Vaissière, B. E., & Henry, M. (2013). Differences of floral resource use between honey bees and wild bees in an intensive farming system. *Ecosystems and Environment*, *179*, 78–86. https://doi.org/10.1016/j.agee.2013.07.007

Rotheray, E. L., Osborne, J. L., & Goulson, D. (2017). Quantifying the food requirements and effects of food stress on bumble bee colony development. *Http://Dx.Doi.Org/10.1080/00218839.2017.1307712*, *56*(3), 288–299. https://doi.org/10.1080/00218839.2017.1307712

Roulston, T. H., & Cane, J. H. (2000). Plant Systematics and Evolution Pollen nutritional content and digestibility for animals. In *Plant Syst. Evol* (Vol. 222).

Roulston, T. H., Cane, J. H., & Buchmann, S. L. (2000). What governs protein content of pollen: Pollinator preferences, pollen-pistil interactions, or phylogeny? *Ecological Monographs*, *70*(4), 617–643. https://doi.org/10.1890/0012-9615(2000)070[0617:WGPCOP]2.0.CO;2

Ruedenauer, F. A., Spaethe, J., & Leonhardt, S. D. (2016). Hungry for quality—individual bumblebees forage flexibly to collect high-quality pollen. *Behavioral Ecology and Sociobiology*, *70*(8), 1209–1217. https://doi.org/10.1007/s00265-016-2129-8

Rundlöf, M., Andersson, G. K. S., Bommarco, R., Fries, I., Hederström, V., Herbertsson, L., Jonsson, O., Klatt, B. K., Pedersen, T. R., Yourstone, J., & Smith, H. G. (2015). Seed

coating with a neonicotinoid insecticide negatively affects wild bees. *Nature 2015 521:7550*, *521*(7550), 77–80. https://doi.org/10.1038/nature14420

Samnegård, U., Persson, A. S., & Smith, H. G. (2011). Gardens benefit bees and enhance pollination in intensively managed farmland. *Biological Conservation*, *144*(11), 2602–2606. https://doi.org/10.1016/J.BIOCON.2011.07.008

Samuelson, A. E., Gill, R. J., Brown, M. J. F., & Leadbeater, E. (2018). Lower bumblebee colony reproductive success in agricultural compared with urban environments. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1881). https://doi.org/10.1098/RSPB.2018.0807

Sawyer, R., Chandler, J., & Rennison, D. (2018). *Pollen Identification for Beekeepers & An Illustrated Key*. Northern Bee Books & Harrogate and Ripon BKA.

Sellappan, S., Akoh, C. C., & Krewer, G. (2002). *Phenolic Compounds and Antioxidant Capacity of Georgia-Grown Blueberries and Blackberries*. https://doi.org/10.1021/jf011097r

Senapathi, D., Fründ, J., Albrecht, M., Garratt, M. P. D., Kleijn, D., Pickles, B. J., Potts, S. G., An, J., Andersson, G. K. S., Bänsch, S., Basu, P., Benjamin, F., Bezerra, A. D. M., Bhattacharya, R., Biesmeijer, J. C., Blaauw, B., Blitzer, E. J., Brittain, C. A., Carvalheiro, L. G., … Klein, A. M. (2021). Wild insect diversity increases inter-annual stability in global crop pollinator communities. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1947). https://doi.org/10.1098/rspb.2021.0212

Senapathi, D., Goddard, M. A., Kunin, W. E., & Baldock, K. C. R. (2017). Landscape impacts on pollinator communities in temperate systems: evidence and knowledge gaps. *Functional Ecology*, *31*(1), 26–37. https://doi.org/10.1111/1365-2435.12809/SUPPINFO

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, *11*(10), e0163962. https://doi.org/10.1371/JOURNAL.PONE.0163962

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I., & Keller, A. (2015a). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, *15*, 20. https://doi.org/10.1186/s12898-015-0051-y

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I., & Keller, A. (2015b). Increased efficiency in identifying mixed pollen

samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, *15*(1). https://doi.org/10.1186/S12898-015-0051-Y

Smart, M., 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*(1), 38–49.

Somerville, D. C. (2001). *Nutritional Value of Bee Collected Pollens: A Report for the Rural Industries Research and Development Corporation*. RIRDC.

Somme, L., Moquet, L., Quinet, M., Vanderplanck, M., Michez, D., Lognay, G., & Jacquemart, A. L. (2016). Food in a row: urban trees offer valuable floral resources to pollinating insects. *Urban Ecosystems*, *19*(3), 1149–1161. https://doi.org/10.1007/S11252-016-0555-Z

Stadhouders, R., Pas, S. D., Anber, J., Voermans, J., Mes, T. H. M., & Schutten, M. (2010). The Effect of Primer-Template Mismatches on the Detection and Quantification of Nucleic Acids Using the 5' Nuclease Assay. *The Journal of Molecular Diagnostics*, *12*(1), 109–117. https://doi.org/10.2353/JMOLDX.2010.090035

Srivathsan, A., Sha, J. C. M., Vogler, A. P., & Meier, R. (2015). Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (Pygathrix nemaeus). *Molecular Ecology Resources*, *15*(2), 250–261. https://doi.org/10.1111/1755-0998.12302

Stout, J. C., & Dicks, L. V. (2022). From science to society: implementing effective strategies to improve wild pollinator health. *Philosophical Transactions of the Royal Society B*, *377*(1853). https://doi.org/10.1098/RSTB.2021.0165

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, *99*(2), 349–364. https://doi.org/10.3732/ajb.1100335

Stevens, B.M., Creed, T.B., Reardon, C.L. & Manter, D. (2023) Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. *Scientific Reports 13*(9323). https://doi.org/10.1038/s41598-023-36101-8

Stubbs, C. S., & Drummond, F. A. (2001). Bombus impatiens (Hymenoptera: Apidae): An Alternative to Apis mellifera (Hymenoptera: Apidae) for Lowbush Blueberry Pollination. *Apiculture and Social Insects J. Econ. Entomol*, *94*(3), 609–616. https://academic.oup.com/jee/article/94/3/609/2217314

Swenson, S. J., & Gemeinholzer, B. (2021). Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. *PLOS ONE*, *16*(2), e0245611. https://doi.org/10.1371/JOURNAL.PONE.0245611

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E. D., Wang, J., Yang, C., Bruce, C., Nevard, T., Potts, S. G., Zhou, X., & Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution*, *6*(9), 1034–1043. https://doi.org/10.1111/2041-210X.12416

Tew, N. E., Memmott, J., Vaughan, I. P., Bird, S., Stone, G. N., Potts, S. G., & Baldock, K. C. R. (2021). Quantifying nectar production by flowering plants in urban and rural landscapes. *Journal of Ecology*, *109*(4), 1747–1757. https://doi.org/10.1111/1365-2745.13598

Thomas B., Murphy D. J., & Murray B. (2003). *Encyclopedia of Applied Plant Sciences*. Elsevier Academic Press

Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., & Trites, A. W. (2016). Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, *16*(3), 714–726. https://doi.org/10.1111/1755-0998.12490

Timberlake, T. P., Vaughan, I. P., Baude, M., & Memmott, J. (2021). Bumblebee colony density on farmland is influenced by late-summer nectar supply and garden cover. *Journal of Applied Ecology*, *58*(5), 1006–1016. https://doi.org/10.1111/1365-2664.13826

Timberlake, T. P., Vaughan, I. P., & Memmott, J. (2019). Phenology of farmland floral resources reveals seasonal gaps in nectar availability for bumblebees. *Journal of Applied Ecology*, *March*, 1365-2664.13403. https://doi.org/10.1111/1365-2664.13403

Tommasi, N., Ferrari, A., Labra, M., Galimberti, A., & Biella, P. (2021). Harnessing the power of metabarcoding in the ecological interpretation of plant—pollinator DNA data:

Strategies and consequences of filtering approaches. *Diversity*, *13*(9), 437. https://doi.org/10.3390/D13090437/S1

Toshack, M., & Elle, E. (2019). Wild bumble bee foraging preferences and fat content in highbush blueberry agro-ecosystems. *Apidologie*, *50*(4), 425–435. https://doi.org/10.1007/s13592-019-00654-3

Tuell, J. K., Ascher, J. S., & Isaacs, R. (2009). Wild bees (Hymenoptera: Apoidea: Anthophila) of the Michigan Highbush blueberry agroecosystem. *Annals of the Entomological Society of America*, *102*(2), 275–287. https://doi.org/10.1603/008.102.0209

Ummenhofer, C. C., & Meehl, G. A. (2017). Extreme weather and climate events with ecological relevance: A review. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 372, Issue 1723). Royal Society Publishing. https://doi.org/10.1098/rstb.2016.0135

Valentini, A., Miquel, C., Taberlet, P., Valentini, A., Miquel, C., & Taberlet, P. (2010). DNA Barcoding for Honey Biodiversity. *Diversity*, *2*(4), 610–617. https://doi.org/10.3390/d2040610

van den Berge, S., Tessens, S., Baeten, L., Vanderschaeve, C., & Verheyen, K. (2019). Contrasting vegetation change (1974–2015) in hedgerows and forests in an intensively used agricultural landscape. *Applied Vegetation Science*, *22*(2), 269–281. https://doi.org/10.1111/AVSC.12424

van der Reis, A. L., Beckley, L. E., Olivar, M. P., & Jeffs, A. G. (2023). Nanopore short-read sequencing: A quick, cost-effective and accurate method for DNA metabarcoding. *Environmental DNA*, *5*(2), 282–296. https://doi.org/10.1002/EDN3.374

van Engelsdorp, D., & Meixner, M. D. (2010). A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology*, *103*, S80–S95. https://doi.org/10.1016/J.JIP.2009.06.011

Vanderplanck, M., Moerman, R., Rasmont, P., Lognay, G., Wathelet, B., Wattiez, R., Michez, D., & Hughes, W. (2014). How Does Pollen Chemistry Impact Development and Feeding Behaviour of Polylectic Bees? *PLOS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0086209

Vanderplanck, M., Zerck, P. L., Lognay, G., & Michez, D. (2020). Sterol addition during pollen collection by bees: another possible strategy to balance nutrient deficiencies? *Apidologie*, *51*(5), 826–843. https://doi.org/10.1007/S13592-020-00764-3/TABLES/4

Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., Tapolczai, K., & Domaizon, I. (2018). Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. *Methods in Ecology and Evolution*, *9*(4), 1060–1069. https://doi.org/10.1111/2041-210X.12960

Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M., & Bouchez, A. (2017). Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? *Freshw. Sci.*, *36*(1), 162–177. https://doi.org/10.1086/690649

Vaudo, A. D., Farrell, L. M., Patch, H. M., Grozinger, C. M., & Tooker, J. F. (2018). Consistent pollen nutritional intake drives bumble bee (Bombus impatiens) colony growth and reproduction across different habitats. *Ecology and Evolution*, *8*(11), 5765–5776. https://doi.org/10.1002/ece3.4115

Vaudo, A. D., Patch, H. M., Mortensen, D. A., Tooker, J. F., & Grozinger, C. M. (2016). Macronutrient ratios in pollen shape bumble bee (Bombus impatiens) foraging strategies and floral preferences. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), E4035–E4042. https://doi.org/10.1073/pnas.1606101113

Vaudo, A. D., Stabler, D., Patch, H. M., Tooker, J. F., Grozinger, C. M., & Wright, G. A. (2016). Bumble bees regulate their intake of essential protein and lipid pollen macronutrients. *Journal of Experimental Biology*, *219*(24), 3962–3970. https://doi.org/10.1242/JEB.140772/262426/AM/BUMBLE-BEES-REGULATE-THEIR-INTAKE-OF-THE-ESSENTIAL

Vaudo, A. D., Tooker, J. F., Patch, H. M., Biddinger, D. J., Coccia, M., Crone, M. K., Fiely, M., Francis, J. S., Hines, H. M., Hodges, M., Jackson, S. W., Michez, D., Mu, J., Russo, L., Safari, M., Treanore, E. D., Vanderplanck, M., Yip, E., Leonard, A. S., & Grozinger, C. M. (2020). Pollen Protein: Lipid Macronutrient Ratios May Guide Broad Patterns of Bee Species Floral Preferences. *Insects 2020, Vol. 11, Page 132*, *11*(2), 132. https://doi.org/10.3390/INSECTS11020132

Velthius, H. W., & Doorn, A. Van. (2006). A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie*, *37*, 421–451.

Vicens, N., & Bosch, J. (2000). Weather-Dependent Pollinator Activity in an Apple Orchard, with Special Reference to *Osmia cornuta* and *Apis mellifera* (Hymenoptera: Megachilidae and Apidae). In *Environmental Entomology* (Vol. 29, Issue 3). https://doi.org/10.1603/0046-225X-29.3.413

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482/METRICS

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*(3), 471–474. https://doi.org/10.1111/J.2041-210X.2012.00190.X

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. In *Nature Biotechnology* (Vol. 39, Issue 11, pp. 1348–1365). Nature Research. https://doi.org/10.1038/s41587-021-01108-x

Warton, D. I., Wright, S. T., & Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, *3*(1), 89–101. https://doi.org/10.1111/J.2041-210X.2011.00127.X

Westphal, C., Steffan-Dewenter, I., & Tscharntke, T. (2006). Foraging trip duration of bumblebees in relation to landscape-wide resource availability. *Ecological Entomology*, *31*(4), 389–394. https://doi.org/10.1111/j.1365-2311.2006.00801.x

Whitehorn, P. R., O'Connor, S., Wackers, F. L., & Goulson, D. (2012). Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science*, *336*(6079), 351–352. https://doi.org/10.1126/SCIENCE.1215025/SUPPL_FILE/WHITEHORN.SOM.PDF

Wietzke, A., Westphal, C., Gras, P., Kraft, M., Pfohl, K., Karlovsky, P., Pawelzik, E., Tscharntke, T., & Smit, I. (2018). Insect pollination as a key factor for strawberry physiology and marketable fruit quality. *Agriculture, Ecosystems and Environment*, *258*, 197–204. https://doi.org/10.1016/J.AGEE.2018.01.036

Williams, N. M., & Kremen, C. (2007). Resource distributions among habitats determine solitary bee offspring production in a mosaic landscape. *Ecological Applications*, *17*(3), 910–921. https://doi.org/10.1890/06-0269

Williams, N. M., Regetz, J., & Kremen, C. (2012). Landscape-scale resources promote colony growth but not reproductive performance of bumble bees. *Ecology*, *93*(5), 1049–1058. https://doi.org/10.1890/11-1006.1

Winfree, R., Aguilar, R., Vázquez, D. P., Lebuhn, G., & Aizen, M. A. (2009). A meta-analysis of bees' responses to anthropogenic disturbance. *Ecology*, *90*(8), 2068–2076. https://doi.org/10.1890/08-1245.1

Winfree, R., Reilly, J. R., Bartomeus, I., Cariveau, D. P., Williams, N. M., & Gibbs, J. (2018). Species turnover promotes the importance of bee diversity for crop pollination at regional scales. *Science*, *359*(6377), 791–793. https://doi.org/10.1126/SCIENCE.AAO2117

Wood, T. J., Holland, J. M., & Goulson, D. (2015). Pollinator-friendly management does not increase the diversity of farmland bees and wasps. *Biological Conservation*, *187*, 120–126. https://doi.org/10.1016/J.BIOCON.2015.04.022

Wood, T. J., Holland, J. M., & Goulson, D. (2017). Providing foraging resources for solitary bees on farmland: current schemes for pollinators benefit a limited suite of species. *Journal of Applied Ecology*, *54*(1), 323–333. https://doi.org/10.1111/1365-2664.12718

Wood, T. J., Holland, J. M., Hughes, W. O. H., & Goulson, D. (2015). Targeted agri-environment schemes significantly improve the population size of common farmland bumblebee species. *Molecular Ecology*, *24*(8), 1668–1680. https://doi.org/10.1111/mec.13144

Wood, T. J., Holland, J. M., & Goulson, D. (2016). Diet characterisation of solitary bees on farmland: dietary specialisation predicts rarity. *Biodiversity and Conservation*, *25*(13), 2655–2671. https://doi.org/10.1007/s10531-016-1191-x

Wood, T. J., Kaplan, I., & Szendrei, Z. (2018). Wild bee pollen diets reveal patterns of seasonal foraging resources for honey bees. *Frontiers in Ecology and Evolution*, *6*(DEC), 210. https://doi.org/10.3389/FEVO.2018.00210/BIBTEX

Woodcock, B. A., Edwards, M., Redhead, J., Meek, W. R., Nuttall, P., Falk, S., Nowakowski, M., & Pywell, R. F. (2013). Crop flower visitation by honeybees, bumblebees and solitary bees: Behavioural differences and diversity responses to landscape. *Agriculture, Ecosystems and Environment*, *171*, 1–8. https://doi.org/10.1016/j.agee.2013.03.005

UN DESA. (2022). *World Population Prospects 2022: Summary of Results*. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf

Yourstone, J., Varadarajan, V. & Olsson, O. (2023) Bumblebee flower constancy and pollen diversity over time. *Behavioral Ecology*, *34*(4), 602 – 612. https://doi.org/10.1093/beheco/arad028

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zeng, J., Cai, H., Peng, H., Wang, H., Zhang, Y., & Akutsu, T. (2020). Causalcall: Nanopore Basecalling Using a Temporal Convolutional Network. *Frontiers in Genetics*, *10*, 1332. https://doi.org/10.3389/FGENE.2019.01332/BIBTEX

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J., & Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, *2*(1), 1–12. https://doi.org/10.1186/2047-217X-2-4

Appendix: Presentations, media and publications

Conference presentations

September 2022. Oral presentation at the Association of Applied Biologists conference, Shaping the future for pollinators: innovations in farmed landscapes. Presentation title: "MinION sequencing of pollen loads using the RevMet approach allows semi-quantitative characterization of commercial bumblebee diets on UK blueberry farms."

October 2022. Oral presentation at the Scandinavian Association for Pollination Ecology (SCAPE). Presentation title: "Commercial bumblebees for blueberry pollination supplement their pollen diet from the wider landscape."

December 2022: Oral presentation at the British Ecological Society. Presentation title: "Commercial bumblebees for blueberry pollination supplement their pollen diet from the wider landscape."

Media

Through the Earlham Institute, some of the work included in this thesis was featured in the 2021 Royal Society Summer Exhibition, entitled "What is a bee's favourite flower?"
I was involved in several different aspects of the exhibition, where I was able to talk about this research. The materials I participated in include:

- YouTube documentary (https://www.youtube.com/watch?v=VyQT8T_YkE0&ab_channel=TheRoyalSociety)
- The Naked Scientists podcast (20/07/2021)
- COP26 Glasgow, 2021. Q&A with members of the public

Publications

The following are publications that I contributed to, but were not connected to the research carried out in this thesis:

Lewis, R., Bell, E., Kent, E.S. (2022). 'Why are pollinators declining?' in Sá-Pinto, X., Beniermann, A., Børsen, T., Georgiou, M., Jeffries, A., Pessoa, P., Sousa, B., & Zeidler, D.L. (Eds.). *Learning Evolution Through Socioscientific Issues*. UA Editora, pp.165-181.

Luke, S. H., Roy, H. E., Thomas, C. D., Tilley, L. A. N., Ward, S., Watt, A., Carnaghi, M., Jaworski, C. C., Tercel, M. P. T. G., Woodrow, C., Aown, S., Banfield-Zanin, J. A., Barnsley, S. L., Berger, I., Brown, M. J. F., Bull, J. C., Campbell, H., Carter, R. A. B., Charalambous, M., … Dicks, L. V. (2023). Grand challenges in entomology: Priorities for action in the coming decades. *Insect Conservation and Diversity*, *16*(2), 173–189. https://doi.org/10.1111/icad.12637