COMPARING METHODS OF IDENTIFYING PLANTS VISITED BY POLLINATORS THROUGH MORPHOLOGICAL AND GENETIC TECHNIQUES

A thesis presented in partial fulfilment of the requirements

for the degree of

MASTER OF SCIENCE

IN

ZOOLOGY

School of Food Technology and Natural Sciences

Massey University

Palmerston North, New Zealand

Jasper Hobbs

2025

Abstract

Interest in bumblebee pollination is increasing as honeybee populations appear insufficient to meet agricultural pollination demand. Effective use of non-honeybee pollinators like bumblebees will also save farmers money, compared to managing commercial honeybees. Potentially, bumblebee populations can be increased by providing supplementary sources of nutrients (plants that blossom outside of crop flowering) on farms. To identify the most beneficial plant species, three methods were proposed that had been used to answer similar questions previously: (1) Using morphology of intentionally and unintentionally collected external pollen from the bumblebees to identify the plants they visited. (2) Using amplicon DNA sequencing (metabarcoding) on DNA extracted from pollen from bees to identify the plants they visited by comparing sequences against reference databases. (3) Using the same amplicon DNA sequencing method on the DNA extracted from bee stomach contents to identify the plants they visited.

These methods were compared using five bumblebees (*Bombus terrestris*), two native bees (*Leioproctus monticola*) and one honeybee (*Apis mellifera*). For method-1. 3627 individual pollen granules were counted and identified to 10 plant genera. For method-2 & 3 the target amplification region used for was the choloroplast gene trnL. 2,631,048 reads were obtained from seven pollen samples and for method-3. 916,868 reads were obtained from eight stomach DNA samples. 732 unique cpDNA sequences were recorded, however many of these referenced to the same genera, or to non-viridiplantae genera or did not reliably match at all (low match and coverage). After the bioinformatics, 35 New Zealand plant genera were identified although 13 represented <1% of the total reads, so were deemed irrelevant.

I found that the intentionally collected pollen in the bee pollen-baskets (corbiculae) did match the unintentionally collected pollen on the rest of the exterior body of bumblebees (in diversity and proportion, not number of grains). It was predicted that the unintentionally collected pollen (body pollen rather than corbicula pollen) would give more accurate representation of all the plant genera visited. The second main finding was that the visual morphology method was less effective than DNA. When comparing the two lists of genera gained from the pollen methods, the pollen DNA method had greater diversity, and is expected to be more accurate due to many pollen morphotypes sharing high similarity. Also, if repeated, now that the processes have been researched and developed, the DNA method would be far less time consuming, than the morphology method.

The third main finding was that the overlap in the list of plant genera visited for pollen and the list of plant genera visited for nectar was not significant enough to distinguish. Although only 14 out of 35 genera were found in both pollen and stomach contents, these 14 genera made up 96.96% of the reads compared to the

pollen and stomach exclusive genera, which contained only one genus with >1% which was 1.7%. These findings resulted in the claim that the plant genera visited for pollen does match the plant genera visited for nectar.

The main outcome of this study is that pollen DNA amplicon method was the most effective in terms of time and quality, and can be trusted to be representative of the species bumblebees feed on, as the third main finding showed using pollen did not exclude any significant genera only found in the stomach, or include significant genera that bumblebees do not consume. However, the failure of the metabarcoding approach to detect wild radish pollen suggests this approach is not perfect.

Key Words: Pollen, palynology, DNA amplicon sequencing, meta-barcoding, bumblebees, pollination

Acknowledgements

Completing my thesis has been an enormous and difficult task and it would not have been possible without the extensive support from my family, friends, cohort and tutors.

It would not have been possible without the tremendous support from my supervisors, Professor Mary Morgan-Richards and Professor Steve Trewick. You have provided constant support across all aspects of my degree such as the more traditional academic work like reviewing and writing literature, to applying for and providing scholarships and funding, and especially the practical skills and providing me with labs and equipment. It is thanks to all your help and support that I have enjoyed my master's degree experience.

I would like to thank Dr. Katherine Holt. I am super grateful for you letting me use your palynology lab and consumables, teaching me acetolysis and then confirming or correcting my pollen identifications.

I would also like to thank Nyasha Chikwature. Your knowledge and experience from your own similar project assisted me multiple steps throughout my thesis, and I am particularly grateful for the help with the bipartite graphs and presenting my results.

Thank you, Paul Barrett, for teaching me how to use the Retsch MM 400 miller.

Thank you, Lab techs Tracy Harris and Cleland Wallace, who gave me access to the labs and solved logistic and practical problems I encountered in them.

I would also like to thank the rest of the molecular ecology department who attended the weekly meetings and gave me advice and suggestions on explaining and presenting my findings.

Table of contents

List of Tables

List of Figures

1. Introduction

1.1 Pollination

Bee populations are on the decline; we're going to run out of food! Is a statement that has been reported by the media for years, but the issues are more complex than this. The global honeybee (*Apis mellifera*) population is increasing, but it is growing slower than agricultural demand for pollinators (Aizen & Harder, 2009; Mashilingi et al., 2022) and although honeybees contribute a large amount of the pollination, there are many crops dependent on other types of bees too (Leonhardt and Blüthgen, 2012).

About a third (35%) of total global crop production is dependent on pollinators (Aizen et al., 2019) as most of the base food such as wheat, rice, maize and oats are self-pollinating or wind pollinated. But 85% of the crop species humans enjoy and depend on are impacted or dependent on pollinators (Klein et al., 2007). This is because we produce far more wheat, rice and corn than we do fruit or vegetables (in terms of weight). Because carbohydrate crops typically do not require pollination, starvation is unlikely to be at risk if bee populations continue to decline, but the food we rely on for vitamins and minerals such as fruits and vegetables will be significantly affected which could then increase the cost of these products, impacting the health of those who cannot afford them.

Even crops that self-pollinate can have increased yields when pollinated, such as cotton in sub-Saharan Africa that saw a 62% increase of *Gossypium* cotton yield compared with an estimated 37% without bee pollination. (Stein et al., 2017).

This natural resource has been taken advantage of for thousands of years, but despite the wild bee population decline being identified 60+ years ago, this decline has still not stopped (Free and Butler, 1959). Due to a lack of data, the cause for this decline has not been identified, but is likely due to a combination of the mechanisation and monocultures of agriculture, extensive use of pesticides, and climate change (Cameron & Sadd, 2020).

The list of pollinator species is estimated to be over 200,000 (Mujeeb et al., 2023), the majority of these being insects such as bees (Hymenoptera), flies (Diptera) and butterflies (Lepidoptera) and many others. In modern agriculture honeybees are confirmed pollinators for 72.7% and suspected pollinators for an additional 10.2% of 1,330 crop species (Nabhan and buchman, 1997). The next biggest contributor are bumblebees which have been largely ignored historically as they appear to be inferior to honeybees in many ways, but when Europeans colonised New Zealand, they found that recreating European pastures for livestock was harder than they expected.

1.2 History of *Bombus* in New Zealand

None of the > 250 species of *Bombus* are native to New Zealand, which have their natural distribution in the Americas, Europe and Asia. A sample of bumblebees were introduced from England to New Zealand in 1885 (lye, 2020) and were released in Lyttelton (Graves, 2015). It is likely that this involved hibernating queens rather than whole colonies, but it is not known how they did it. It is also unknown exactly where in England the original sample of *Bombus* were collected or how many species or individuals were brought over, but it is reported that they thought there were at least six species. Four remain species remain, and recent genetic studies show all four species have all gone through population bottlenecks as well (Schmid-Hempel et al., 2007; Lye et al., 2011).

They were introduced for the agricultural industry, to pollinate red clover for cattle pastures. It is unknown when early settlers learnt that red clover requires bumblebee pollinators, but it is known that red clover was introduced long before the bumblebees. It is unknown exactly when clover was introduced to New Zealand but in 1835 Charles Darwin visited the mission station at Waimate (30 minutes south of Timaru) and wrote in his book 'Naturalist's Voyage Round the World', "*Fine crops of barley and wheat were standing in full ear; and in another part, fields of potatoes and clover*" (Darwin, 1860).

There was at least 50 years between the introduction of red clover and introduction of bumblebees (Parker, 2022). Although clover is capable of self-seeding, they produce fewer seeds and have low germination (Vleugels et al., 2019). So, until bumblebees were established in New Zealand, red clover seed had to be imported from overseas to maintain the population (Parker, 2022).

A study comparing genetic diversity of Tasmanian and New Zealand bumblebees found that bumbles bees taken from a single county in England had greater genetic diversity than the equivalent collected across New Zealand. This was calculated using allelic richness data from seven microsatellite markers. They found that in the New Zealand population the allelic richness was 4.24 and the expected heterozygosity was 0.729 compared to the Lincolnshire population which was 5.079 and the expected heterozygosity was 0.826. (Schmid-Hempel et al., 2007).

Using a using a Factorial Correspondence Analysis (FCA) method to calculate the overall genetic differentiation, the researchers made a model showing although there is overlap in the allelic richness, the New Zealand population extends beyond this suggesting it has evolved into a new population, and not just a subset.

This low diversity puts the New Zealand populations at risk of negative effects of low genetic diversity and inbreeding, but interestingly, studies such as Gerloff and Schmid-Hempel, 2005, have observed high

tolerance to inbreeding in laboratory tests with *B. terrestris*, which is likely due to their haplo-diploid sex determination. This is because haploid individuals only carry one set of alleles so cannot carry deleterious recessive alleles masked by dominant alleles. Because deleterious alleles are expressed fully in haploid males, natural selection is more effective at removing these alleles from a population. Inbreeding depression causes a reduction in survival because inbreeding in diploid organisms can lead to a build-up of deleterious recessive alleles. But bumblebees can avoid this thanks to haplo-diploid sex determination which is likely why they have high tolerance against negative effects of low genetic diversity (Haplodiploid Sex-determination, 2024).

1.3 Importance of *Bombus* to agriculture

Currently and historically, honeybees are the most important pollinators of most commercial crops (Newstrom-Lloyd, 2013). The main reason for this is volume as during pollination season, a well-managed hive can have up to 80,000 individuals (Baer, 2018). The number of hives placed on a farm for pollination varies, but a general rule of thumb is 2-3 hives per hectare, although for some crops such as avocado, farmers my place up to 8 (Syms & Denton,1996). This can result in over 1 million bees on a farm. The volume, efficiency and control of honeybees has made them the preference for commercial crop pollination, but recently, for the purpose of pollination, the attraction of bumblebees has been growing (Lihoreau et al. 2025). In nature bumblebees are generalist pollinators using many different cues to direct their foraging effort (Dahl 2025; Sommer et al., 2022) and so their role in agriculture is likely to be complex.

1.4 Greenhouse pollination

One area of agriculture where bumblebees are particularly important is in greenhouse agriculture. The use of greenhouses allows greater control over the climate allowing farmers to grow plants in locations and seasons where it was previously not possible to. Removing the crops from the environment allows farmers to farm more efficiently by reducing the crops lost to natural impacts like weather, birds and other pest. But one drawback is the loss of natural pollinators. Plants in greenhouses can be pollinated by humans through methods that range from manually using a brush to using automated robots. But all of these can have large time and monetary costs compared to allowing bees to do it naturally (Broussard & Martinsen, 2023). Because of this in most countries bees are used, but although honeybees maybe thought of as a first choice, they have been found to be ineffective and often struggle to survive. This is because they rely on landmarks and the direction of sunlight to orientate themselves, so the monotonous interior and diffused light often

causes bees to get disoriented and lost. Bumblebees however do not have this issue and are comfortable living in greenhouses (Bumblebee Biology, 2024). They can reportedly visit up to 450 flowers per hour, which is more effective than most artificial methods and a lot cheaper too (Kaftanoglu et al., 2000; Morandin, 2001). A recent study also found that in a greenhouse environment, bumblebee pollinated tomatoes had lower total soluble salts (TSS), pH and postharvest weight loss than that of self-pollinated and hand vibrated treatments, resulting in greater income for farmers (Zameer et al., 2022).

1.5 Solanaceae and buzz pollination

Another area of agriculture that is dependent on bumblebees is the pollination of Solanaceae plants. The Solanaceae or nightshade family contain Tomatoes, Capsicum, Chillies and other consumed fruits and they all rely on a method of pollination called buzz pollination (Nayak et al., 2020). Honeybees, due to their small and delicate nature, collect relatively small amounts of pollen when they visit certain clades of flower. In contrast, bumblebees, with their large size and more aggressive disturbance or constant 'buzzing', disrupt the stamen (the pollen releasing organ) more aggressively, causing a greater amount of pollen to be shaken off and collected (What's all the buzz about, 2024).

1.6 Orchard fruiting crops pollination

Bumblebees have been historically used in the previous two applications because honeybees were not suitable, however there is now growing use of bumblebees in fruiting crops pollination which is traditionally pollinated by honeybees. There are many benefits to using bumblebees in conjuncture or instead of honeybees, as bumblebees are far more efficient, and can handle greater environmental challenges. Honeybees need warm, dry days for effective operation while bumblebee's threshold for cold, wind and rain are much higher, allowing them to still pollinate effectively on bad days (Karbassioon et al., 2023). Their larger bodies and longer hair mean they can carry far more pollen than honeybees, and their buzzing and more aggressive nature disturbs more pollen too, even on species that do not require buzz pollination for successful breeding. Biobees Limited, the largest commercial bumblebee producer in New Zealand, claim "*Bumblebees are ten times more efficient in performing close pollination (within cultivar) and at least 20 times more efficient at cross-pollination than honeybees in avocados*." (What's all the buzz about, 2024).

Which is backed up by results found in studies such as (Nayak et al., 2020) and (Stern et al., 2021). However, what is likely the largest factor pushing these farmers and pollination companies towards bumblebees is the

impact of varroa on honeybees. There is increasing demand for pollinations services as wild populations continue to decrease, and at the same time, beekeepers are struggling to prevent varroa induced hive collapse across their colonies as treatment resistant varroa mites are increasing (Stahlmann-Brown et al., 2022). In New Zealand, if a pollination company intends to purchase colonies before the pollination season (usually because they cannot keep their own hives alive), this must be organised months in advance as the demand is so high (plus decent hives take 2–3 months to develop in spring). Varroa is evolving into a devastating problem in New Zealand (as well as globally) and so bumblebees' resistance is highly attractive to pollination companies.

1.7 Pasture and red clover pollination

In New Zealand we highly value fruiting crops, but what may be bumblebees' greatest impact on our current farming industries in New Zealand is the impact of pollinating red clover and lucerne, which are vital crops for grazing cattle. They are both perennial tap-rooted legumes which are high in protein and other nutrients required for cattle (Black et al., 2009). Red clover has its nectaries (nectar producing organs) at the base of the long corolla tube, which is too difficult to reach for honeybees, so they avoid them, but bumblebees have much longer tongues than honeybees which allows them to reach the nectar and pick up pollen in the process. The nectaries of other important feed crops like Lucerne (Alfalfa) are visited by honeybees, but the larger size and competitive behaviour bumblebees' makes them farm more efficient (Zhang et al., 2022).

1.8 Bumblebee domestication: why are they not used like honeybees?

Populating farms with honeybees is relatively easy as their hives are moveable, and their colonies can be to be split and sustained. But bumblebee colonies last less than 6 months making them incredibly hard to domesticate. Greenhouse farmer's that use bumblebees buy queens yearly instead of trying to sustain colonies beyond one season. They wait till their crop is in flower, then release the bees where they pollinate the plants for a few weeks and then die.

In New Zealand, once flowers run out at the end of summer, the colonies collapse as the queens, workers and males die, leaving only the new queens to hibernate over the winter and emerge sometime in spring. Once they emerge, the new queens immediately find nectar for energy and pollen that helps their ovaries develop, then they search for a suitable nest site and start developing new colonies.

This process is extremely difficult to recreate artificially. Although humans have been collecting and moving colonies for a couple hundred years, commercial rearing was not achieved until 1987 (Velthuis & Doorn, 2006). Even then, it is currently unfeasible to do within a farming setting and only two companies in New Zealand (Zonda and Biobees) have managed to replicate the environmental requirements, allowing them to breed queens and colonies at will.

1.9 Cost of pollination to farmers

Even with honeybees, hours of care are spent supporting the colonies, which a crop farmer may not have time for. Many hire pollination services to do this work for them which is an added cost. Fruiting crop farmers require pollination for short timeframes, so some just release new bumblebees each week for month or so, compared to cattle farmers which require clover pollination for months.

The cheapest and easiest option is to use wild bumblebees. They can already be found on farms and do lots of pollination for free. So instead of buying or trying to sustain domestic bees, why not just encourage and increase the number of wild ones to forage on the farm. When choosing a nesting site, *Bombus* queens prioritize foraging availability above all else, preferring locations where they can forage efficiently with minimal travel to conserve energy. Many crops flower later in spring, so in early spring queens may select sites outside these farms. Bumblebees forage up to a few hundred meters of the nest (Wolf & Moritz, 2008), so if a farm is more than a couple hectares wide (which many are), they may have low bumblebee density when the crops flower. Although studies have shown *B. terrestris* can travel up to around 800m while foraging (other species are typically much less) (Wolf & Moritz, 2008; Knight et al., 2005), so they likely will be found on all farms but at lower density if the nests are not on the property.

Many crops bloom later in spring, by which time *Bombus* queens have already chosen their nesting sites. To attract more queens to a farm, increasing the availability of flowers in early spring would have the greatest impact, so farmers could encourage *Bombus* queens to establish nests on their property by planting early blooming supplementary plants that provide nectar and pollen until the main crops flower. Identifying plants local bumblebees visit during this period would be key to maximizing their presence.

1.10 Methods of identifying the plants being visited

Bumblebees intentionally collect pollen in corbiculae and unintentionally collect it on the body hairs when feeding from nectaries, so a recently published study took advantage of this and used pollen collected from museum bumblebee specimens to identify plants being visited. "*This study reaffirms the potential of natural history collections to understand historic foraging patterns*." Knowles et al. (2022).

This study provides a list of plants that bumblebees visit but I predict this method provides skewed partial results of plants that Bumblebees forage on. They do include a lot of disclaimers such as pollen found on bees is not indicative of foraging as they found *Pinus* pollen, which being wind pollinated plants, means the pollen was most likely passively collected. But even excluding the unfeasible pollen from the foraging list, it still does not address the issue that adult bumble bees primarily feed on nectar. Although bumblebees do consume some pollen while foraging, it has been proposed that this is to quickly test pollen quality before collecting, rather than for their own nutrition (Mayberry et al., 2024). Identifying plants from pollen in corbiculae might miss the plants they are visiting for their own nutrients. I predict that the intentionally collected pollen would form the majority of pollen found on the samples, potentially skewing or excluding the species they visit for their own nutrients.

It is currently assumed that the plants bees visit for their own nectar consumption are the same plants they collect pollen from. If this is the case, then using pollen is a fair estimator of the species of plants being foraged on, but if this is not the case, then plants visited for their own nutrients are being excluded. A way to test this is to compare the stomach contents against the external pollen, using plant DNA sequences (markers). This type of study is called amplicon DNA sequencing (or metabarcoding) and has been used on pollen and stomach contents in similar studies (Mallott et al., 2018; Bontšutšnaja et al., 2021). There are a range of regions/markers that get used depending on the target organisms. One commonly used gene fragment is CO1 from the mitochondrial genome (mtDNA) and there is a comprehensive database of reference sequences for this marker (Porter & Hajibabaei, 2018). However, because we are interested in plants, targeting a chloroplast (cpDNA) region is appropriate, as these primers will not anneal to non-viridiplantae DNA (in theory). Five regions are commonly used in pollen studies: four are chloroplast genes (trnL, rbcL matK, and trnH-psbA) and one in the nuclear DNA - ITS (Lowe et al., 2022). For this study trnL primers that have been shown to outperform other primers in previous animal diet studies were available (Mallott et al., 2018).

Because bees consume mostly nectar, their stomach contents are expected to have low plant matter which will affect the DNA quantity, so the same amount of success may not be seen compared to other insect diet studies. To account for this, studies extracting DNA from honey rather than more traditional stomach

contents will likely give a better prediction and instruction of methods. In previous studies using DNA to investigate honey, trnL primers have typically been chosen further suggesting this was a good choice of marker in our study (Bayram, 2021; Milla et al., 2021).

Once the DNA from pollen and stomach samples have been extracted, amplified and sequenced, the results will be used to create lists of plant species (or plant genera) visited and these can be compared to determine whether there is significant overlap between the two approaches. The plant list from metabarcoding will be compared to the list gained from using pollen morphology to access accuracy.

1.11 Questions being asked

The original general question of this study was ‘How can farmers increase bumblebee populations on their land?’, which led to the question ‘What plants are bumblebees in New Zealand visiting?’. This then led to a study which addressed this question but raised questions about their methodology which led to the questions investigated in this study.

Question 1. Does the intentionally collected pollen in corbiculae (pollen baskets) match the unintentionally collected pollen on the rest of the exterior body?

Question 2. How effective is the method of using visual morphology when compared to DNA?

Question 3. Does the list of plants visited for pollen match the list for nectar?

1.12 Methodology

Bumblebees will be collected, and pollen baskets (corbiculae) will be taken. One will be used for pollen morphology and the other for DNA amplicon sequencing. The bees will then be washed, and the washed and corbicula pollen will undergo acetolysis for more accurate identification. The morphological types will be categorized, identified and confirmed by a palynologist. Then individual granules of each pollen type will be counted and recorded, resulting in a list of species visited by the bumblebees. The remaining pollen samples will be broken and have the DNA extracted and amplified. The washed bumble bees will then have their stomachs removed and the DNA will also be extracted and amplified. A plant specific marker will be used to exclude unwanted DNA. The pollen and stomach samples will be sequenced then the raw data will undergo read quality checks, cleaning and referencing, resulting in lists of species visited by the bumble bees.

The comparison of the results from the morphologically identified pollen from corbiculae and the washed exterior will answer question 1. The comparison of the results from the morphologically identified pollen, the DNA identified pollen will answer question 2. and the comparison of the results from the DNA identified pollen and stomach contents will answer question 3.

1.13 Expected outcomes

If the lists of species precured from the different methods have significant overlap, then it can be inferred that reasonably accurate representation of the plant species visited by bumblebee can be made from external pollen alone (Knowles et al. 2022). However, this is not predicted to be the case.

For question 1. It is predicted that the number of pollen granules in the intentionally collected pollen samples will be far greater than the number of pollen granules in unintentionally collected exterior body pollen samples. It is also predicted that the lists will have different plant diversity.

For question 2. It is predicted that the list of visited plant species from the DNA method will be greater in diversity and accuracy than the list of visited plant species from visual morphology. This is because pollen can be hard to distinguish within plant clades, while DNA can identify down to species (provided a there is a good reference database).

For Question 3. It is predicted that the list of plants visited for pollen will not match the list for nectar. If it is found that the list of plants visited for pollen does not match the list for nectar, then it can be inferred that the pollen collected from bumblebees is predominantly intentionally collected pollen, and not representative of species visited for nectar. If it is found that the list of plants visited for pollen does match the list for nectar, then it can be inferred that the assumption, external pollen is representative of the plant's bee visit for their own personal nutrition.

The results from this study will show the variation, accuracy, benefits and limitations of each method. The lists of plant species identified will give information about the diets of local bumblebees, but the small sample size will likely be too small to give a complete picture.

2. Materials and Methods

2.1 Bumblebee sampling

During the first sampling session, twenty-eight bumblebees were collected from different locations in the Manawatu region in March 2024. Using tongue length, all specimens were identified as *Bombus terrestris* (Goulson et al., 2002). Only four had pollen baskets (corbiculae) so only these four were used for the full analysis. The remaining bees from this sampling session were used for testing methods. Two of the four *Bombus terrestris* came from a small cattle-rearing farm in Opiki (40°28'17.3"S 175°25'47.0"E), the other two were collected between the Manawatu River and Massey Farmland (40°23'05.4"S 175°35'49.6"E). Bees were collected by placing a 100ml sample tube over the insect when it was on a flower, which makes them fly up in panic, allowing the lid to be slipped underneath. During later sampling sessions other bees were collected, including bumblebees (*Bombus terrestris*), Honeybees (*Apis mellifera*) and native bees (*Leioproctus monticola*) were collected from flowering plants such as koromiko (*Veronica salicifolia*) in the Manawatū (-40.4129, 175.662). The *Leioproctus* was identified by experts at Plant and Food Research. The sample names of the bees used in this study are R1, R2, 11, E9, 285, 286, 290 & 099. The first four came from the first sampling session, and the remaining came from later sessions. The species of each specimen and how it was used can be seen in table 1.

Table 1. Details of all samples used in this study from three bee species collected near Palmerston North, New Zealand.

Pollen morphology

Samples	Type of sample	Speciemen ID	Species
1	Leg pollen	R1	*Bombus terrestris*
2	Washed pollen	R1	*Bombus terrestris*
3	Leg pollen	R2	*Bombus terrestris*
4	Washed pollen	R2	*Bombus terrestris*
5	Leg pollen	11	*Bombus terrestris*
6	Washed pollen	11	*Bombus terrestris*
7	Leg pollen	E9	*Bombus terrestris*
8	Washed pollen	E9	*Bombus terrestris*
9	Washed pollen	099	*Bombus terrestris*
10	Washed pollen	096	*Bombus terrestris*
11	Washed pollen	277	*Bombus terrestris*
12	Leg pollen	293	*Leioproctus monitcola*
13	Washed pollen	293	*Leioproctus monitcola*
14	Leg pollen	285	*Leioproctus monitcola*
15	Washed pollen	285	*Leioproctus monitcola*
16	Leg pollen	286	*Apis mellifera*
17	Washed pollen	286	*Apis mellifera*

Pollen DNA

Samples	Type of sample	Speciemen ID	Species
1	Pollen DNA	R1	*Bombus terrestris*
2	Pollen DNA	R2	*Bombus terrestris*
3	Pollen DNA	11	*Bombus terrestris*
4	Pollen DNA	E9	*Bombus terrestris*
5	Pollen DNA	285	*Leioproctus monitcola*
6	Pollen DNA	293	*Leioproctus monitcola*
7	Pollen DNA	286	*Apis mellifera*

Stomach DNA

Samples	Type of sample	Speciemen ID	Species
1	Stomach DNA	R1	*Bombus terrestris*
2	Stomach DNA	R2	*Bombus terrestris*
3	Stomach DNA	11	*Bombus terrestris*
4	Stomach DNA	E9	*Bombus terrestris*
5	Stomach DNA	099	*Bombus terrestris*
6	Stomach DNA	285	*Leioproctus monitcola*
7	Stomach DNA	293	*Leioproctus monitcola*
8	Stomach DNA	286	*Apis mellifera*

2.2 Morphological identification of pollen

The bumblebees were stored at -18˚C degrees. Whole legs with pollen baskets (corbiculae), were taken from the bees, and these were stored in separate 1.5ml tubes. To collect external pollen trapped in the body hairs, the whole bees were submerged in water and detergent in 1.5ml tubes. They were shaken and then centrifuged at 10,000 rpm for 5 seconds. These tubes were also kept at -18 ºC. To allow comparison with

other species additional bees with corbiculae were collected: R1, R2, 11, E9 (*Bombus terrestris*), 286 (*Apis mellifera*) 285 and 293 (*Leioproctus monticola*) and were washed. Three *Bombus terrestris* (096, 099 and 277) had no corbiculae and so only provided washed pollen samples. This resulted in 17 samples total.

The pollen samples underwent acetolysis to chemically remove protoplasm making the pollen more visible for botanical identification. To separate the pollen from the legs, 1ml of Potassium hydroxide (KOH) was added to the seven samples, then these were kept at 80˚C for 10 minutes. The legs were then dipped up and down in the solution to shake off any remaining pollen. The samples were centrifuged at 3,000rpm for 3 minutes. The KOH was poured off, leaving the pellet behind, which was washed by vortexing with 1ml of distilled water. Samples were centrifuged at 3,000rpm for 3 minutes and the water discarded before addition of 1ml 10% Acetic acid. These were then vortexed and centrifuged again at 3,000rpm for 3 minutes and this step was repeated using Glacial Acetic acid instead of 10% Acetic acid, and the liquid was discarded. The acetolysis solution was made with 18ml of Acetic anhydride and 4ml of Concentrated sulfuric acid. 0.5ml of this was added to each sample, and these were incubated to 80 ºC for 4 minutes with the lids off. 1ml of glacial Acetic acid was added to each sample, and they were then vortexed and centrifuged at 3,000rpm for 3 minutes. The liquid was discarded and the step repeated with glacial Acetic acid, then with 10% Acetic acid, and finally with distilled water.

Samples were mounted on glass microscope slides with warm glycerine and red biological stain on a heated slide bed. The glycerine was warmed periodically to reduce viscosity.

Figure 2.1. Preparation of bumblebee pollen for microscopic examination. Left: Intact bee with pollen on hind legs. Right: Detail of pollen basket (corbicula) on hind leg. Right: Samples being heated to 80ºC.

Slide mounted pollen samples were examined on an Olympus BX51 microscope at 10x magnification. The slides were positioned with the edge of the cover slip aligned to the edge of the view of the camera, and then scrolled providing 20 consecutive photos in a transect to get a representation of all the pollen types and the proportion. If the sample had low pollen granule density, additionally transects were made until a pollen granules were located. Some slides had thousands of granules while others had <100. One sample was from a honeybee, which are extremely selective, so although this sample had high density, only a few photos were taken, as only one kind of pollen was present.

Temporary morphotypes were established, and these were used for taxonomic identification by comparison with available images of pollen types using online image databases and searches. Once confident identification had been made and verified by an experienced palynologist, the slides were re-examined and the numbers of each pollen taxon recorded for analysis.

2.3 DNA extraction, amplification and sequencing of pollen

DNA was extracted from pollen using the GeneJET Plant Genomic DNA Purification Kit following mechanical exine rupture. Pollen was scraped from each corbicula with a sterile scalpel into a separate 2ml prefilled 1mm zirconium ball high impact tube. These were agitated at 25 shakes per second dry for 15 minutes in a Retsch MM 400 miller. 350uL of Lysis Buffer A (GeneJET, Thermo scientific) was added, and these were then run for an additional 15 minutes.

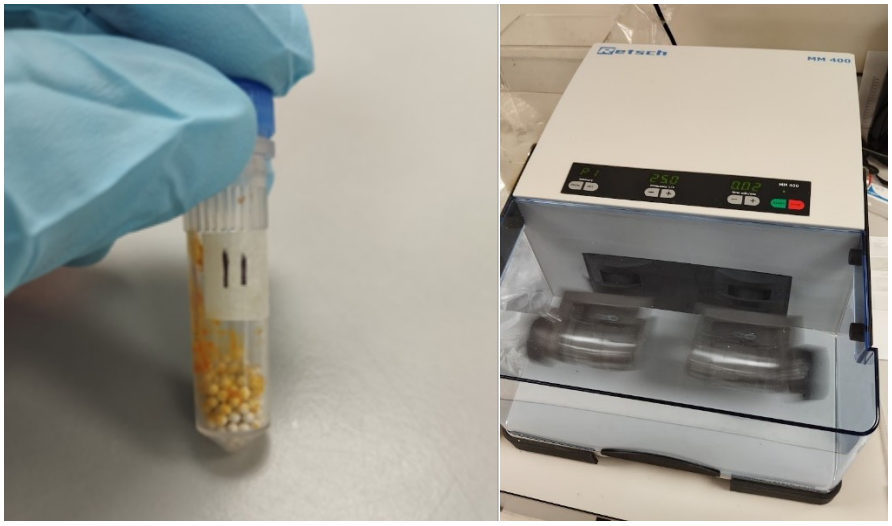

Figure 2.2 Preparation of bumblebee pollen for DNA extraction. Left: pollen and beads in 2ml tube (Sample 11p in tube before processing). Right: Retsch MM 400 miller required for bead-bashing to rupture pollen grains.

Lysed pollen in Lysis buffer A was pipetted to a clean 1.5ml tube with 50uL of Lysis buffer B and 20uL of RNase A (concentration). This was incubated for 10 minutes at 65˚C in a Allsheng MSC-100 Thermo Shaker Incubator. 130uL of Precipitation Solution was added, and mixed by inversion, and the samples were incubated on ice for 5 minutes. Samples were centrifuged at max RPM (14,000rpm), and the supernatant transferred to a clean tube containing 400uL of plant gDNA binding solution and 400uL of 96% ethanol. After thorough mixing half the solution was transferred to a 2ml spin column and centrifuged at 8,000 rpm for 1 minute. The flow-through was discarded and the remaining solution was run though the same column. 500uL of wash Buffer 1 was added to the spin column and this was centrifuged for 3 minutes at 10,000rpm, the flow-through discarded and 500ul of Wash Buffer 2 added. After being centrifuged for 3 minutes at 14,000rpm, the flow-though was discarded and the tube re-spun for another minute. The spin-column was then transferred to a clean microcentrifuge tube and the DNA was eluted from the membrane using 50uL of Elution buffer, which was left to incubate at room temperature for 5 minutes before being centrifuged at 10,000rpm for 1 minute.

To check the DNA extractions were suitable for amplicon sequencing trial amplifications were performed of the chloroplast gene trnL. For amplification polymerase chain reaction (PCR) 10uL reactions contained, 6.1uL H2O, 1uL Buffer, 1uL dNTP's, 0.4uL of each primer (trnLc- CGAAATCGGTAGACGCTACG;, trnLd GGGGATAGAGGGACTTGAAC; Taberlet et al. 2007), and 0.08uL of Taq DNA. Polymerase. Samples were then treated to the following thermal cycling conditions: 95˚C for 1 minute, 94˚C for 30 seconds, 50˚C for 15 seconds, 72˚C for 1 minute, then step 2 to 4 was repeated 34 times, then 72˚C for 10 minutes. DNA extractions were quantified using a Qubit fluorometer (table 2.1), before sending for amplicon sequencing.

Table 2. Details of samples used for DNA extracted from three bee species collected near Palmerston North, New Zealand.

Samples	Tube code	Name on report	Genus	Tissue	Total Amount	Conc. (ng/µL)	Volume (µL)
1	R1	R1P1	*Bombus*	Pollen	576	14.4	40
2	R1s	R1S2	*Bombus*	Stomach	2048	51.2	40
3	R2	R2P3	*Bombus*	Pollen	182.4	4.56	40
4	R2s	R2S4	*Bombus*	Stomach	1304	32.6	40
5	11	11P5	*Bombus*	Pollen	240	6	40
6	11s	11S6	*Bombus*	Stomach	2128	53.2	40
7	E9	E9P7	*Bombus*	Pollen	696	17.4	40
8	E9s	E9S8	*Bombus*	Stomach	692	17.3	40
9	285p	285P9	*Leioproctus*	Pollen	42.4	1.06	40
10	285s	285S10	*Leioproctus*	Stomach	848	21.2	40
11	293p	293P11	*Leioproctus*	Pollen	79.6	1.99	40
12	293s	293S12	*Leioproctus*	Stomach	78	1.95	40
13	286p	286P13	*Apis*	Pollen	640	16	40
14	286s	286S14	*Apis*	Stomach	1016	25.4	40
15	099s	099S15	*Bombus*	Stomach	135.2	3.38	40

Fifteen samples were amplicon sequenced by Custom Science Ltd. The primers used for the amplification were for the chloroplast gene trnL-c A49325 (forward) CGAAATCGGTAGACGCTACG; and trnL-d B49863 (reverse) GGGGATAGAGGGACTTGAAC; Taberlet et al. 2007) which are expected to yield a fragment of approximately 450 bp, although INDELS within the trnL intron result in length variation among species (Taberlet et al. 2007). cpDNA fragments were sequenced on a NovaSeq using 250 base pair (bp) Paired-Ends (PE).

2.4 DNA extraction, amplification and sequencing of stomach contents

The bee specimens were stored at -18˚C in separate plastic bags or vials. To soften for dissection the whole bag was submerged in warm water for 5 minutes, then dried. The bee was placed dorsally on a clean square of EVA Foam and two pins were pierced through the top of the thorax to hold it in place. Each abdominal sternite was removed exposing the organs, then the honey and main stomachs (fig. 2.3) were transferred to a 1.5ml tube. For DNA extraction, 350uL of Lysis buffer A was added to each tube, a clean pestle was used to crush the stomachs/contents to break up the tissue. The rest of the process from extraction to Qubit testing was the same as the pollen process above.

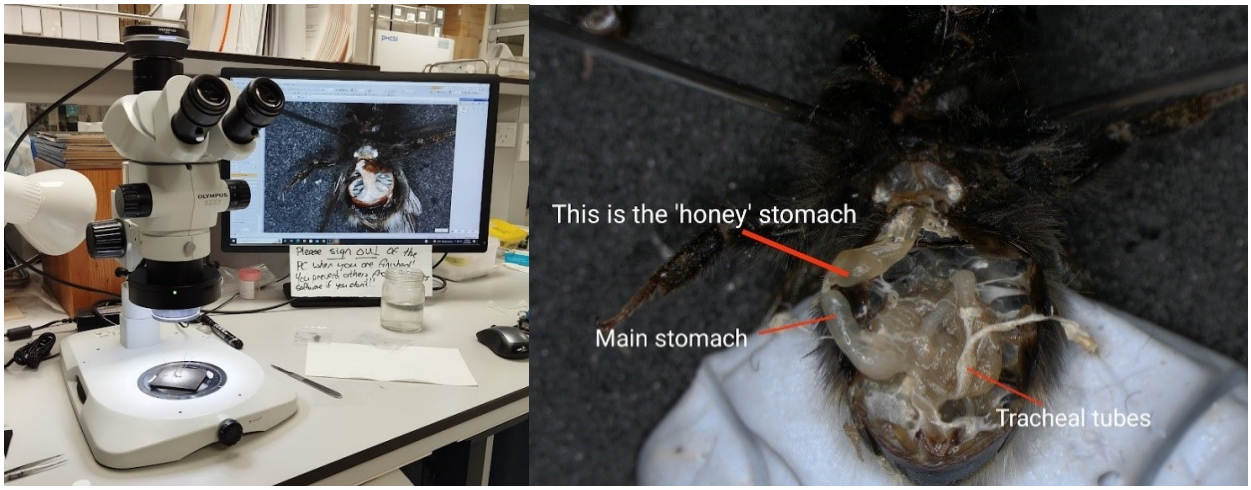

Figure 2.3. Bumblebee dissection to study plant visits using stomach contents. Left: Stomach extraction set up. Right: The bee digestive tract.

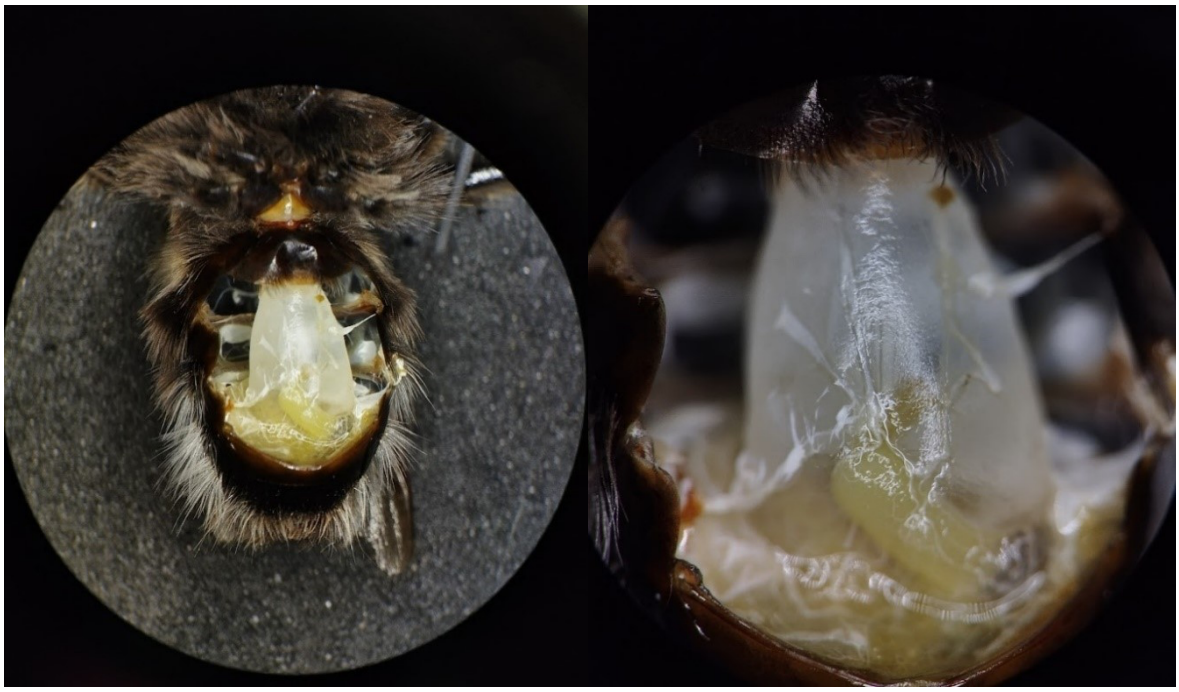

Figure 2.4. Bumblebee dissection to study plant visits using stomach contents. Left: Honey stomach in situ. Right: full honey stomach.

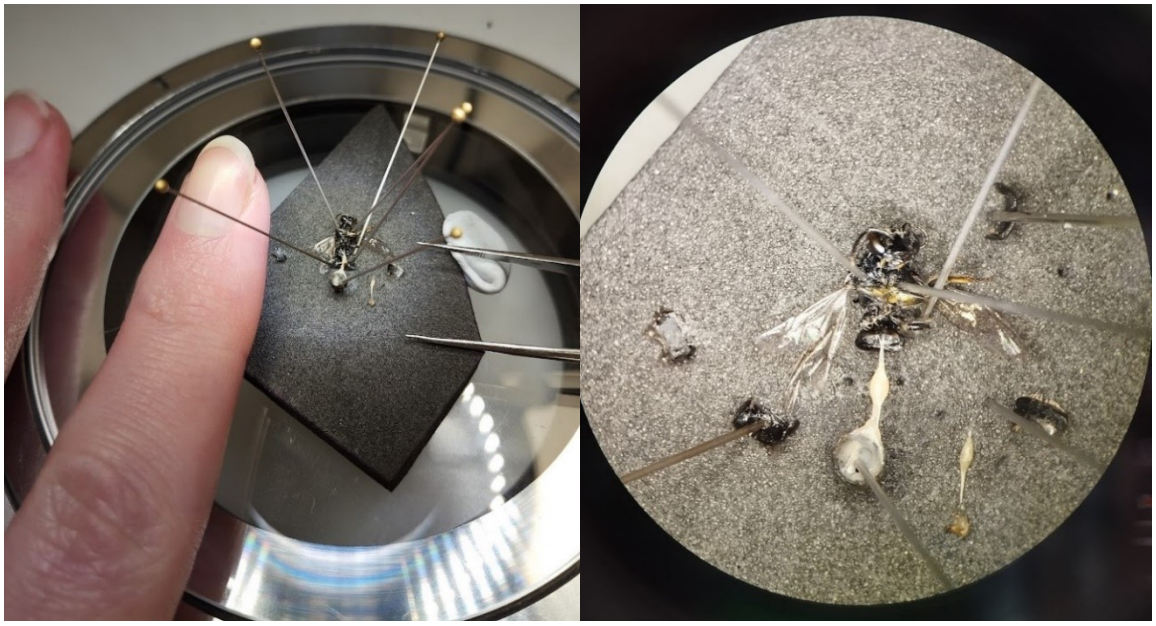

Figure 2.5. *Leioproctus monticola* dissection to study plant visits using stomach contents. Left: *L. monticola* with finger for reference. Right: showing removed sternites and exposed digestive tract.

2.5. Bioinformatics

2.5.1 Rationale

Amplicon sequencing generates millions of DNA sequence reads from the various taxa within the sample. Identification of these taxa is done by comparison with reference DNA sequences to infer their origin on the basis of similarity. This uses the Basic Local Alignment Search Tool (BLASTn) algorithm (Camacho et al., 2009) to comparing the new data with available published data. This is computationally intensive, so to reduce processing time the raw sequence reads are processed to make a list of sequence variants present in each sample. First, primers are removed from the reads and the poor-quality reads remove from the samples. The raw reads are dereplicated to leave only unique sequences. The number of reads for each unique sequence is recorded as this will later be used to assess relative abundance of the identified sequences in each sample. The remaining reads are made into Amplicon Sequence Variants (ASVs) that represent unique, error-corrected DNA sequences that merge very similar sequences and are equivalent to Operational Taxonomic Units (OTUs). Chimeric ASVs are removed leaving a set of ASVs that represent the error-free diversity in each sample. The number of reads associated with each ASV provides a read-depth score and infrequently occurring ASV's can be excluded to focus on the most abundant ones. The remaining were referenced against the NCBI database (Benson et al., 2012), and were identified using the closest matching species. 'Match' refers to the number of identical nucleotides between the databases' reference and the query ASV sequence, however, the degree of overlap between the ASV and the reference sequence (coverage) is also taken into account. If, for example, only 50bp of a 225bp ASV is aligned to the top-ranked reference species and is identical at those positions, the match will be recorded as 100%, even though a large proportion of the ASV was not compared. Therefore, both match and coverage percentages must be considered when assessing how accurate an identification is. After the BLAST searches, any dubious putative identifications are investigated further. Where these had low match or coverage they were excluded.

2.5.2 The DADA2 workflow.

The short DNA sequences (reads) from Custom Science Ltd. consisted of a zipped file (~5Gb unzipped) containing an Analysis report, the raw sequences and the cleaned sequences. However, many of the cleaned sequences still had primers attached so the raw sequences were used. This allowed greater control and ability to experiment to reduce the number of sequences lost in processing steps.

For example, for the sample R1P1, The Custom Science Ltd. cleaning pipeline reduced the reads from 523,042 to 104,360, while the bespoke pipeline used in this experiment kept 379,108 (72.5% of the raw data compared to 19.9%). During the identification stage this was reduced to 377,610, of which 99.6 percent of the identified sequences were useable, showing that the 'less conservative' pipeline was effective.

The first step in the sequence cleaning process was to remove the primers. Although Dada2 has this function, preliminary trials showed that it was not accurate enough. Cutadapt is a python program that accounts for partial primers remaining on sequences so is better at recognising and removing them. This is typically given instructions and run thought the terminal, but I had 15 samples so a python script called search_and_cutadapt.py was made with help from ChatGPT to run though the 15 sample files and output the primer-removed sequences into separate files. Once this was done, these files were imported into R studio and the sequence quality profiles were inspected. The FilterAndTrim function was used to clean them, by removing low quality reads and trimming the reads to 225 base pairs. The reads were reduced to 225 base pairs as quality profiles showed the bases dropped below a quality score of 30 beyond 225, meaning the reads up to 225 base pairs long were predicted to have a 99.9% call accuracy, which is why Illumina recommends excluding bases below Q30 (Illumina, 2025). This limitation likely had an impact on mapping accuracy, which is discussed further in the discussion.

The DNA reads were then dereplicated which collapses identical reads into unique sequences. This improves computational efficiency by reducing the number of reads and reduces noise by removing low-quality or spurious reads improving sequence inference accuracy.

This dataset was used to estimate error rates. The standard estimating function gave poor results as the observed error rates were showing a terrible fit to the Expected (Fitted) Error Model. This is usually due to insufficient data or poor sequencing quality, but due to the read quality profiles, this was unlikely the cause. This same issue has been found by others before and they found it was due to Illumina sequencers binning quality scores, which does improve efficiency but can impact dada2's error rate estimator function (Callahan, 2021). To solve this, they created their own functions, trialling multiple adjustments and they found that a function that altered loess function arguments (weights, span, and degree) and also enforced monotonicity (called Option 4 in my script) had the closest fit for the observed and predicted error rates. Four of these trial functions were published, so all were tested using my own data, and the same result was found so option 4 was used for the final pipeline.

Once a realistic error rate model was created, this was used to identify true biological sequences and create amplicon sequence variants (ASVs). Normally at this stage in the dada2 pipeline, the forward and reverse reads would be merged, which require a minimum overlap of 10bp (Qaigen, 2025). As the reads were

trimmed to 225 the maximum length and amplified region could be to successfully merge would be 42240bp, but the trnL region varies in length range of 332 to 654 bp; (Do & Zaveska, 2018). Because of this, early trials found merging lost a large proportion of data, so this step was skipped in the final script, and only the forward reads were used from this step onwards. The clean reads from the Custom Science Ltd. output data was comprised of only 20% of the original data, which may be because of the use of an overly conservative default pipeline that including the merging of forward and reverse reads.

DNA sequence cleaning process was complete when chimeric sequences were removed. To track the data lost at each step, a table was created to make sure a reasonable amount of data was kept. Table 2).

Table 3. Bioinformatic read exclusion. The number and proportion DNA sequence reads at each processing step of the pipeline.

Samples	Input	Filtered	Denoised Fwd	Denoised Rvs	Non-chimera	Final percent	Identified (id)	Percent kept	Usable id's	Percent kept
R1P1	523042	393738	393346	393366	379108	72.5	377610	99.6	377610	100.0
R1S1	104360	67129	66218	65956	65746	63	57591	87.6	4403	7.6
R2P3	1156804	886508	884337	885592	870119	75.2	868762	99.8	864661	99.5
R2S4	104081	66595	65650	65502	64599	62.1	55990	86.7	17642	31.5
S11P5	694265	506133	505701	505849	490880	70.7	489973	99.8	473059	96.5
S11S6	102367	68271	67435	67295	64403	62.9	56747	88.1	26848	47.3
E9P7	569798	399640	399490	399112	398572	69.9	397230	99.7	397230	100.0
E9S8	104200	66890	66058	66049	66049	62.9	56221	85.1	12907	23.0
S285P9	316310	229669	229557	229617	227778	72	226650	99.5	226650	100.0
S285S10	679181	426498	424962	424615	381813	56.2	373436	97.8	305043	81.7
S293P11	177616	141774	141604	141617	139406	78.5	138457	99.3	137859	99.6
S293S12	220646	175173	175006	175098	172019	78	170742	99.3	170742	100.0
S286P13	212997	163579	163433	163464	161649	75.9	160023	99.0	160023	100.0
S286S14	680133	464405	462705	462864	405138	59.6	398209	98.3	376135	94.5
S099S15	103144	75368	74980	74979	66482	64.5	64135	96.5	3572	5.6

2.6 Taxonomic identification process

The unique DNA sequences for each sample were extracted and added to a dataframe with their abundance (reducing ~4300 to ~3000 sequences). This was done to each of the 15 samples individually, as samples were expected to share similar plants so would contain the same sequences. The DNA sequence with the highest abundance had 34,570 copies. Most DNA sequences had fewer than 100 copies (reads), for example, 92% of R1S1 sequences had fewer than 100 copies and 41% had fewer than 10. All sequences with fewer than 100 reads (< 0.05%) were removed from the dataframe in R, reducing the list of unique sequences to 732. This list was then exported in FASTA format, with each sequence named with a placeholder, e.g. >Sequence_1 to >Sequence_732.

Using Blastn with the –remote function sequences can be uploaded to the NCBI servers directly from the terminal. So, a python script was created that uploaded each fasta sequence, then downloaded the xml file, then took the top identified species an added the name to the name in the fasta file. E.g. >Sequence_1 became >Sequence_1_Trifolium_radicosum in the fasta file. Each sequence was uploaded individually with a 30 second delay to avoid overloading the servers and getting blocked. This process recorded the time that passed between each sequence upload, download, and renaming, and most averaged 90 seconds. However, some sequences took far longer (the longest I saw when making and trialling the script was 40 minutes), so if the script went more than 10 minutes without a reply from NCBI, it skipped the sequence, then reported the skipped sequences once it was finished (Fig. 2.7).

This was done in two batches one day apart on 2/01/2025 and 3/01/2025 both took around 8 hours. It is likely the NCBI severs were under less use during this time.

Once this was done, the two lists were recombined into a spreadsheet in Excel (MicroSoft Excel 16.0.18623.20266) and separated into the 15 samples. All 732 unique sequences were matched to a reference, which was then manually checked and poor and non-viridiplantae identifications were excluded. Multiple species or genera identified had more than one unique DNA sequences within the same sample so these were combined into one for the curated genera lists. For each bee specimen a list of plant taxa was created from pollen morphology, external pollen DNA and DNA from stomach contents and graphed using Bipartite package in R (Figures 3.3 to 3.6).

Table 4. Custom Science Ltd.'s table showing the number of reads lost during their cleaning pipeline.

Sample	RawPE	Combined	Qualified	Nochime	Base(nt)	Avglen(nt)	GC	Q20	Q30
R1S1	104360	94943	93661	93654	28363242	302.85	40.39%	98.82%	95.48%
R2S4	104081	66354	65770	65769	16874737	256.58	46.18%	99.04%	96.35%
S11S6	102367	57888	57391	57386	15539911	270.80	43.96%	98.99%	96.19%
E9S8	104200	84200	83380	83365	23000526	275.90	45.24%	98.91%	95.92%
S285S10	679181	84529	82613	82588	22374659	270.92	43.40%	98.36%	94.78%
S286S14	680133	45850	45342	45341	10927628	241.01	46.80%	98.63%	95.50%
S099S15	103144	66669	65523	65486	21314303	325.48	51.85%	98.70%	95.29%
R1P1	523042	4231	4165	4165	1458018	350.06	37.41%	99.10%	96.74%
R2P3	1156804	6085	6036	6036	2106180	348.94	37.47%	99.15%	96.90%
S11P5	694265	23679	23454	23454	8150246	347.50	37.83%	99.16%	96.89%
E9P7	569798	531	476	476	198007	415.98	34.51%	96.69%	90.83%
S285P9	316310	1710	1457	1457	636675	436.98	34.67%	95.52%	88.22%
S293P11	177616	879	806	806	329736	409.10	34.70%	96.79%	91.13%
S293S12	220646	1572	1374	1374	583887	424.95	34.33%	96.20%	89.79%
S286P13	212997	439	413	413	166635	403.47	33.46%	97.49%	92.88%

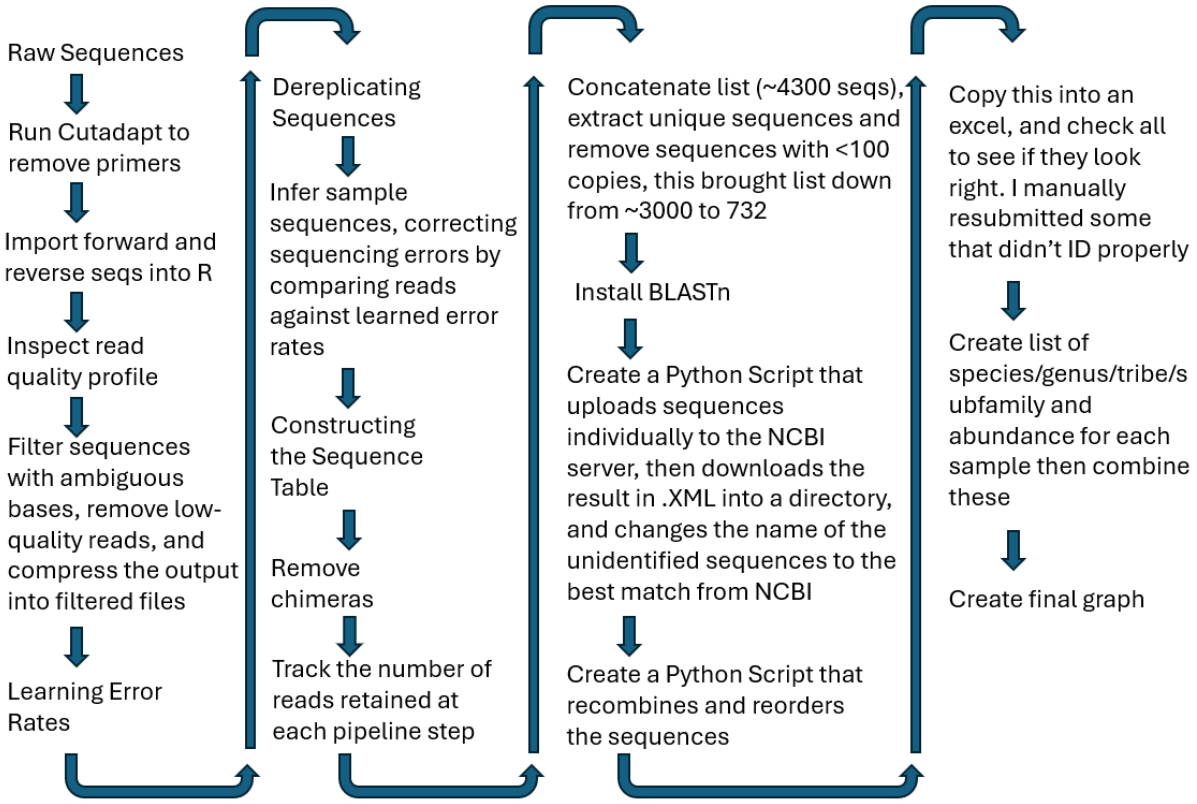

Figure 2.6. Bioinformatics pipeline for analysis of trnL amplicon DNA sequence data from bees.

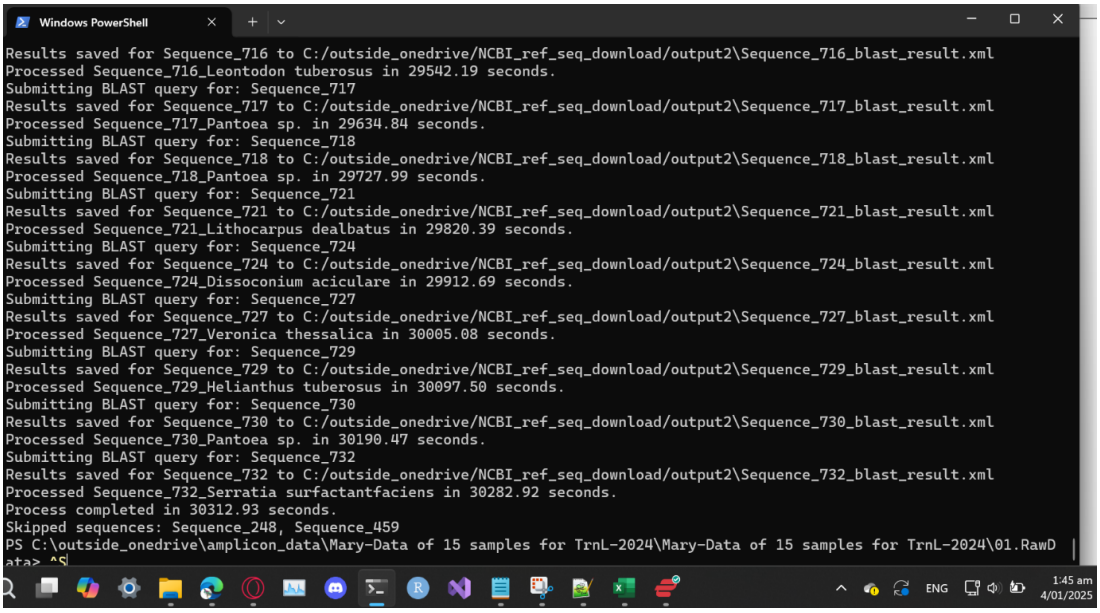

Figure 2.7 Screenshot of script uploading and downloading data to and from NCBI.

The python script takes the name of the species from the top match in the downloaded NCBI .xml file, no matter how close the match or coverage. As a result many of the putative reference matches were to species, genus and sometimes even higher taxonomic ranks that are not found in New Zealand. When the xml files were opened these sorts of identifications would typically have a match of around <50bp out of 225, so could match to almost anything. These were attributed to under representation of New Zealand species on NCBI and were excluded. Other excluded identifications were assumed to be correct (high match and coverage) but were not Viridiplantae, including bacteria and metazoa. The number of sequences kept during this step can be seen in the final column of table 2.

3. Results

3.1 Pollen morphology

3627 individual granules were identified and counted across the 17 slides. As predicted, most of the pollen granules came from the Leg pollen, which had 2846 (78.54%) from 7 slides (Table. 5.), compared to 776 from the total 10 washed samples 21.45% (Table 1; Fig. 3.1). The proportion and distribution of the plant genera identified was very similar using the two methods of pollen collection (Figure 3.1). Of the ten plant detected from washed pollen but not recorded from the corbiculae samples (Table 5).

Table 5. Exact counts of pollen granules from the 17 slides, comprising external pollen (washed) and pollen in corbiculae (leg).

	Slides	Bellis	Eucalyptus	Geranium	Hypochaeris	Lotus	Mentha	Passiflora	Raphanus	Trifolium	Veronica	Total
1	R1_Leg	0	0	11	12	1038	0	0	0	255	0	1316
2	R2_Leg	0	0	0	7	0	0	0	0	171	0	88
3	11_Leg	0	3	0	4	2	0	234	276	12	51	178
4	E9_Leg	0	0	0	1	651	0	0	0	14	0	22
5	285_Leg	0	0	0	0	21	0	0	0	0	0	582
6	286_leg	73	1	0	1	0	0	0	0	0	0	391
7	293_leg	0	0	0	0	8	0	0	0	0	0	666
	Total	73	4	11	25	1720	0	234	276	452	51	2846
8	R1_Washed	0	0	0	1	80	0	0	0	7	0	24
9	R2_Washed	0	0	0	0	3	0	0	0	19	0	21
10	11_Washed	0	0	4	2	0	0	24	357	3	1	21
11	E9_Washed	0	0	0	0	65	0	1	1	6	3	75
12	285_washed	0	0	0	0	21	0	0	0	0	0	
13	286_washed	100	0	0	1	0	0	0	0	0	0	101
14	293_washed	0	0	0	0	10	0	0	0	0	0	8
15	096_washed	0	3	0	7	8	8	0	0	0	0	10
16	099_washed	0	1	0	1	2	11	0	2	0	0	26
17	277_washed	4	0	0	0	20	0	0	0	0	0	17
	Total	104	4	4	12	209	19	25	360	35	4	776
	Total	177	8	15	37	1929	19	259	636	487	55	3622

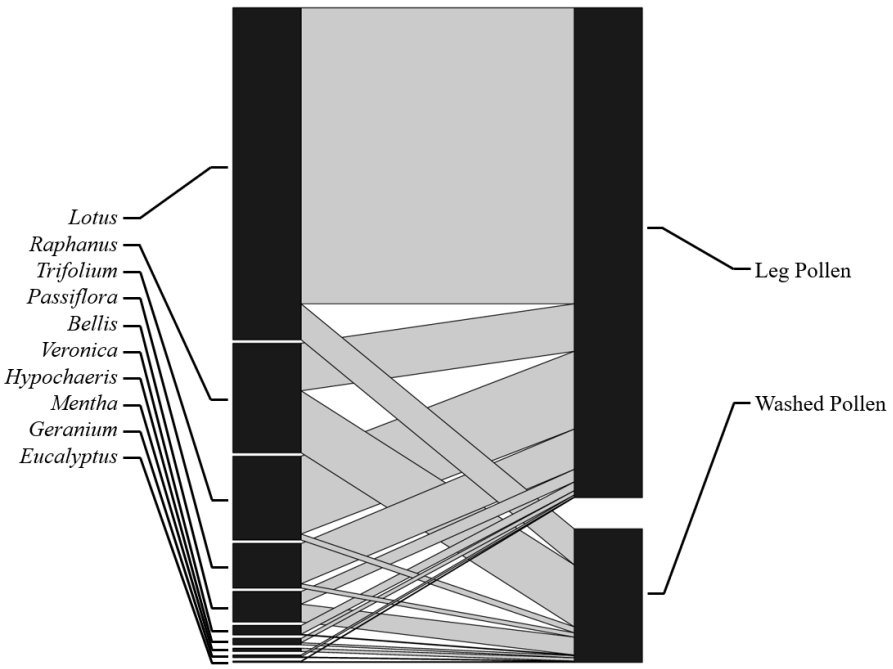

Figure 3.1. The proportion and distribution of the plant genera morphologically identified from pollen from bees collected in New Zealand, comparing two methods of pollen collection and illustrated with a bipartite graph. Left column shows the number pollen granules counted and identified to a genus and the right column shows the number pollen granules counted from each sample, then combined into the source of the samples.

Table. 5. Exact counts and percentages of pollen granules from *Bombus terrestris* legs and body identified for morphologically to plant genus.

Sample Type	Genus	Granules	Percent %	Sample Type	Genus	Granules	Percent %
Leg Pollen	*Lotus*	1720	60.436	Washed Pollen	*Lotus*	209	26.933
Leg Pollen	*Raphanus*	276	9.698	Washed Pollen	*Raphanus*	360	46.392
Leg Pollen	*Trifolium*	452	15.882	Washed Pollen	*Trifolium*	35	4.510
Leg Pollen	*Passiflora*	234	8.222	Washed Pollen	*Passiflora*	25	3.222
Leg Pollen	*Bellis*	73	2.565	Washed Pollen	*Bellis*	104	13.402
Leg Pollen	*Veronica*	51	1.792	Washed Pollen	*Veronica*	4	0.515
Leg Pollen	*Hypochaeris*	25	0.878	Washed Pollen	*Hypochaeris*	12	1.546
Leg Pollen	*Mentha*	0	0.000	Washed Pollen	*Mentha*	19	2.448
Leg Pollen	*Geranium*	11	0.387	Washed Pollen	*Geranium*	4	0.515
Leg Pollen	*Eucalyptus*	4	0.141	Washed Pollen	*Eucalyptus*	4	0.515

To compare the leg pollen and washed pollen, I separate the data into the individual samples (the four *Bombus terrestris* samples), revealing the variation among the bees using leg pollen or washed pollen (Fig 3.2). For example, pollen washed from the bee bodies has a lower proportion of radish (*Raphanus*) pollen than the pollen in the corbiculae, and only bee-11 had collected a high proportion of *Raphanus* pollen

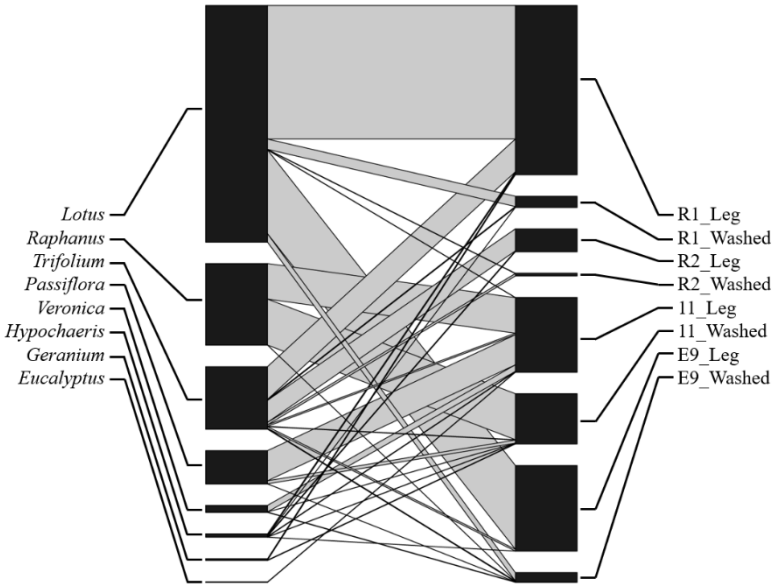

Figure 3.2. The proportion and distribution of the plant genera morphologically identified from pollen from four bumblebees (*Bombus terrestris*) collected in New Zealand illustrated with a bipartite graph. Left column shows the number pollen granules counted and identified to a genus and the right column shows the number pollen granules counted from the leg (corbicula) and washed pollen samples. The data from the additional bee samples can be found in the appendix (Fig A.1).

3.2 Amplicon DNA sequencing

A total of 2,631,048 reads were obtained from the seven pollen samples and 916,868 reads from the eight stomach DNA samples. From the eight bees (15 DNA samples) 732 unique cpDNA sequences were recorded, however many of these were later referenced to the same genera. Identification of these sequences required matching to reference databases, during which it was determined that many sequences were not from plant chloroplasts. For example, the stomach sample from *Leioproctus monticola* (285), contained sequences identified to flies, moths, trout, carp, shark, rays, and even a slow loris. Unique DNA sequences were removed from further analysis if the genus was not found in New Zealand, the overlap with reference was less than 50bp, or the sequence did not match a plant (Viridiplantae). The combined pollen samples had 1.04% of identified reads excluded while the combined stomach samples had 25.64% of identified reads excluded.

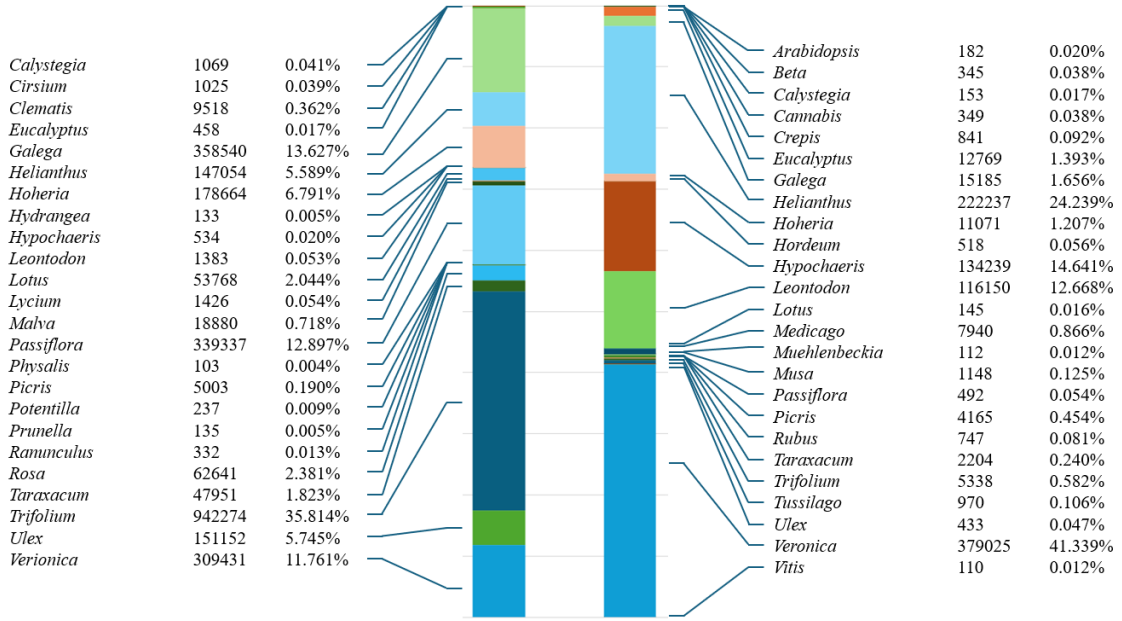

Figure 3.3. Combined trnL results from the pollen (left) and stomach DNA (right) of three species of bee (n =15) using relative read depth of amplicon data (stacked column chart).

The diversity of plant genera and proportion of reads obtained differed between the external pollen DNA and bee stomach DNA (Fig. 3.2). Ten plant genera were identified only from external pollen (e.g. *Clematis*) and eleven from only stomach DNA (e.g. *Medicago*) but in all cases these were relatively uncommon sequences (Fig. 3.3, 3.4). Fourteen plant genera were found in both pollen and stomach samples (Fig. 3.3).

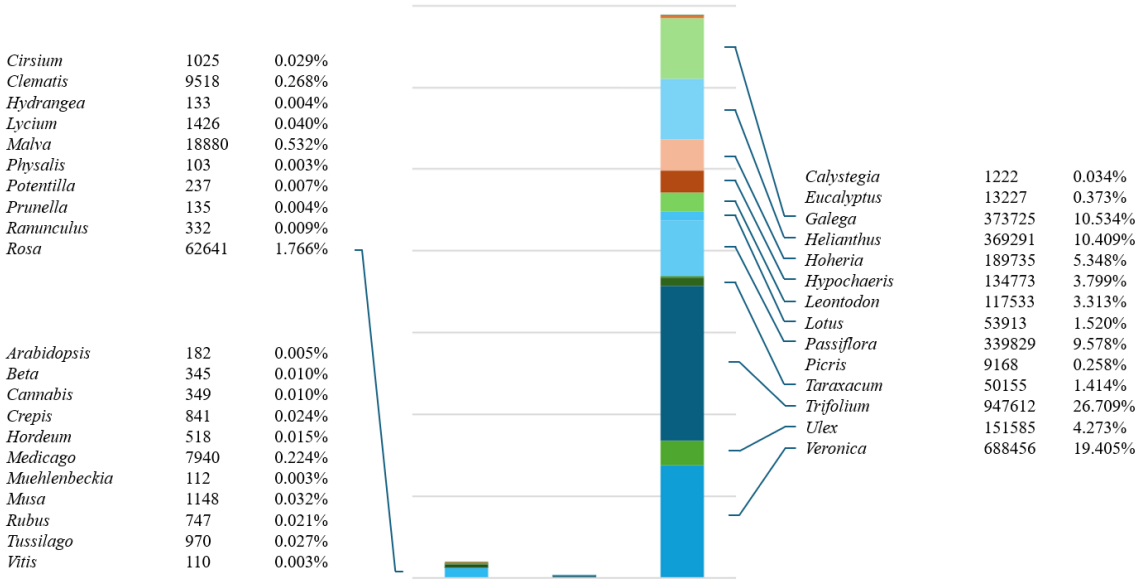

Figure 3.4. Plant genera identified using trnL amplicon DNA sequencing only in *Bombus terrestris* bee pollen samples (left), bee stomach samples (middle) or identified in both (right).

Pollen DNA is more convenient and reliable to use that stomach DNA, but using it relies on the assumption that exterior pollen is representative of the consumed plant genera. This figure shows that although using only pollen DNA will exclude some genera of plants, these stomach-exclusive genera are so low in the number of reads that they are insignificant, so it is safe to assume the exterior pollen is representative of the significant genera of plants bees visit for nutrients. Alternatively, using only pollen DNA will likely include genera not consumed, but again the number of reads for these pollen-exclusive genera are insignificant.

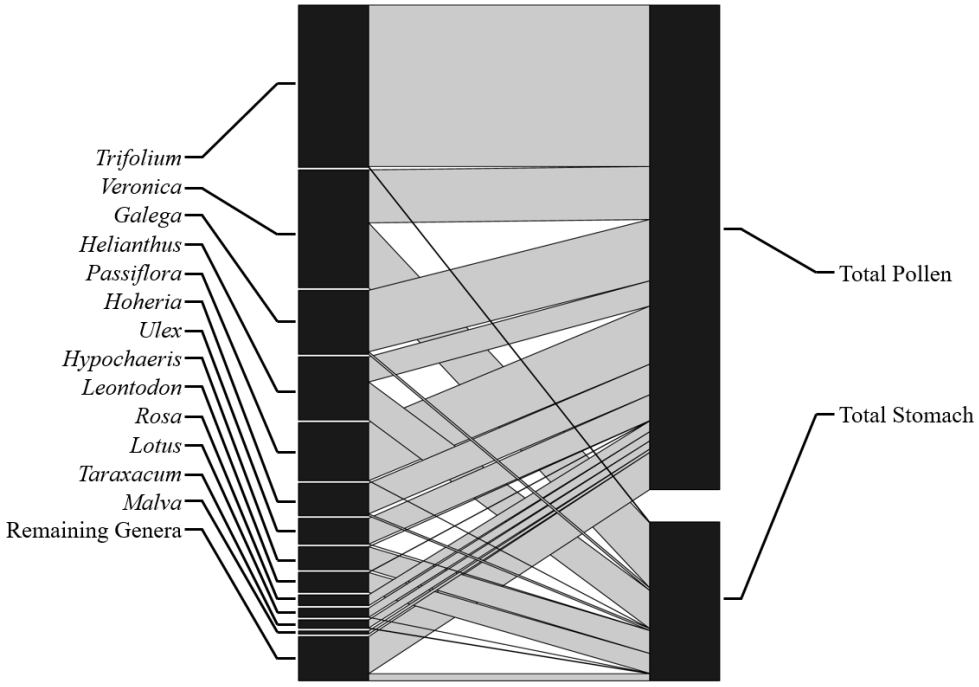

Figure 3.5. Bipartite graph of the genetically identified plant genera identified pollen and stomach contents from *Bombus terrestris*. Left column shows the number of reads referenced to a genus and the right column shows the number of useable reads sequenced from the two methods.

Nearly all the *Trifolium* (clover) reads came from the bee pollen samples, which also contained a high proportion of *Galega* (Goats rue; Fig 3.5). In contrast, bee stomach samples contained most of the dandelion-like catsear (*Hypochaeris*) and *Leontodon* (hawkbit). Most of the *Veronica* (koromiko) reads came from stomach samples, despite stomach samples having far fewer cpDNA reads than the pollen samples (Fig 3.4).

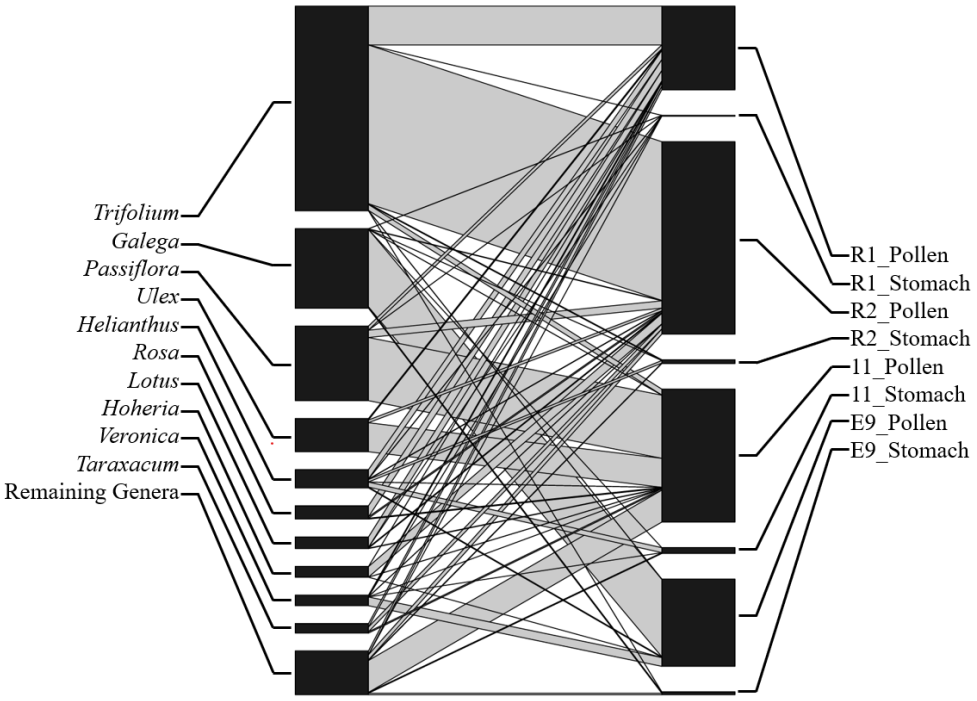

Figure 3.6. The proportion and distribution of the plant genera identified from DNA sequences from four bumblebees (*Bombus terrestris*) collected in New Zealand illustrated with a bipartite graph. Left column shows the number of reads referenced to a plant genus and the right column shows the relative number of reads sequenced from each sample (external pollen or stomach contents). This figure shows only the original four *Bombus terrestris* samples; the additional bee samples can be found in the appendix (Fig. A.2).

Comparing the pollen DNA and the stomach DNA for individual bee samples reveals variation among specimens (Fig. 3.6). The *Bombus terrestris* sample R2 was dominated by *Trifolium* (clover) pollen in the morphological and DNA datasets but the other three bumblebee samples showed less consistency. I recorded a high proportion of *Lotus* (Birdsfoot trefoil) pollen using morphology from the *Bombus terrestris* sample R1 and E9 but DNA sequences suggested a much higher relative *Trifolium* (clover) pollen load for R1 and *Galega* (goat's rue) for E9 (Fig. 3.6). *Bombus terrestris* sample R11 had pollen from *Passiflora* (passionfruit) in both DNA and morphological identifications but the high radish (*Raphanus*) pollen load observed from morphology was not detected in the DNA data (Fig 3.6).

Because the amount of DNA reads and number of pollen granules is not directly comparable, a Bray-Curtis dissimilarity metric allows comparison in addition to the visual comparison figures, as this compares the relative abundances of the genera found in the samples rather than direct counts. The mean Bray–Curtis dissimilarity (BCD) between methods (for samples from bumblebees R1, R2, 11 & E9) was 0.68 ± 0.10, indicating moderate overlap in community composition. Exact Bray–Curtis dissimilarity figures can be seen in the table below.

Table 6. Bray–Curtis dissimilarity figures for samples from bumblebees R1, R2, 11 & E9 showing the relationship between methods. (Microscopy/Metabarcoding).

Sample	Comparison	Bray–Curtis dissimilarity	Sample	Comparison	Bray–Curtis dissimilarity
R1_Leg/R1_Washed	Micro/Micro	0.123	11_Leg/11_Washed	Micro/Micro	0.449
R1_Leg/R1P1	Micro/Meta	0.678	11_Leg/11P5	Micro/Meta	0.572
R1_Leg/R1S1	Micro/Meta	0.890	11_Leg/11S6	Micro/Meta	0.996
R1_Washed/R1P1	Micro/Meta	0.793	11_Washed/11P5	Micro/Meta	0.930
R1_Washed/R1S1	Micro/Meta	0.890	11_Washed/11S6	Micro/Meta	0.997
R1P1/R1S1	Meta/Meta	0.708	11P5/11S6	Meta/Meta	0.988
R2_Leg/R2_Washed	Micro/Micro	0.136	E9_Leg/E9_Washed	Micro/Micro	0.124
R2_Leg/R2P3	Micro/Meta	0.163	E9_Leg/E9P7	Micro/Meta	0.997
R2_Leg/R2S4	Micro/Meta	0.767	E9_Leg/E9S8	Micro/Meta	0.979
R2_Washed/R2P3	Micro/Meta	0.159	E9_Washed/E9P7	Micro/Meta	0.958
R2_Washed/R2S4	Micro/Meta	0.807	E9_Leg/E9_Washed	Micro/Meta	0.938
R2P3/R2S4	Meta/Meta	0.802	E9P7/E9S8	Meta/Meta	0.565

Sample	Comparison	Bray–Curtis dissimilarity	Sample	Comparison	Bray–Curtis dissimilarity
Mean Leg/Washed	Micro/Micro	0.208	Mean Washed/Pollen	Micro/Meta	0.710
Mean Leg/Pollen	Micro/Meta	0.603	Mean Washed/Stomach	Micro/Meta	0.908
Mean Leg/Stomach	Micro/Meta	0.908	Mean Pollen/Somach	Meta/Meta	0.766

A BCD score of zero means there is complete overlap in composition, while a 1 is complete dissimilarity. The comparison of the Leg and washed pollen microscopy samples have the lowest dissimilarity (closest to 0) as they are comparing the same method and near same data (both pollen granule count), while the comparison of the pollen microscopy and stomach metabarcoding have the highest dissimilarity, which is expected as this is comparing different methods and different data (pollen granule count and stomach DNA reads).

4. Discussion

The focus of my study is to compare methods and related assumptions of bee diet studies, using bumblebees as a target organism, as honeybees have historically been prioritised over bumblebees despite the important contribution of the latter to pollination (Morandin, 2001; Zhang et al., 2022; Lihoreau et al. 2025). Because their importance for pasture and solanaceae pollination has been known for years, it is likely the decline in honeybees due to the Varroa mite is pushing this interest in alternative bees (Stahlmann-Brown et al., 2022). When this is combined with the difficulty of domesticating and maintain bumblebees, research into local populations has never been of more value.

Historically the cost of DNA sequencing has been preventative and visual studies like pollen morphology have been used, but with the cost or DNA sequencing and research decreasing, these decisions should now be reassessed as to whether cost/accuracy ratio of using DNA is worth it (Rohland & Reich, 2012; Li et al., 2019). To test this, three main questions were proposed (as seen in section 1.10) to test the methods and related assumptions of the following three approaches: visual pollen morphology, pollen DNA, and stomach DNA amplicon sequencing.

4.1 Key Findings

4.1.1 Question 1. Does the intentionally collected pollen in corbiculae match the unintentionally collected pollen on the rest of the exterior body of bumblebees?

I predicted that the number of pollen granules in the intentionally collected pollen samples would be far greater than the number of pollen granules in unintentionally collected exterior body pollen samples. This was found to be the case as 78.5% of the counted pollen granules came from the corbiculae (figure 3.1).

At the beginning of the project, I predicted that bumblebees would target different plants for pollen or nectar, so the unintentionally collected pollen (collected by washing it off the body, thus ‘washed pollen’) would be different to the intentionally collected leg pollen, which would prove it to be unrepresentative of species visited for nectar. This would disprove the claim that ‘leg pollen can be used to accurately identify the plants bumblebees are visiting for food’. However, only one of the ten plant genera was found in a single sample type. This genus was *Mentha,* and it was found only in washed pollen samples, however, the total 19 *Mentha* pollen grains came from two bees which did not have corbiculae, so I cannot conclude a difference in pollen between corbiculae, and other parts of the body. Thus, my prediction that the lists would have different plant diversity was not supported.

I have found that the intentionally collected (corbiculae) pollen diversity matches the incidental (body) pollen, and implies both sources represent the same pattern of flower visitation. Thus, the prediction that leg pollen can be used to accurately identify the plants bumblebees are visiting for food may potentially still be correct. There still remains the possibility that plants targeted for pollen-collection have stickier, larger, or more numerous pollen grains compared to plants targeted for nectar. However, plants produce nectar to attract pollinators, so a plant evolving attractive, high-quality nectar, with poor quality pollen is unlikely. On the other hand, many bees have nectar-robbing behaviours that gain access to nectar while bypassing the reproductive structures and avoid collecting pollen (Irwin et al., 2010). For example, this is known to happen when short-tongued bumblebees forage on clover, but clover is still successfully pollinated because robbing is a learned behaviour not exhibited by all short-tongued bumblebees and its more efficient for long-tongued bumblebees to feed the mutualistic way (Newman & Thomson, 2005). Thus, even if nectar robbing was occurring, some of the pollinators would be expected to collect in a mutualistic way, thus a suitable sample size would ensure pollen from nectar targeted plants would still be found on some of the bees.

For morphological identification of pollen, it appears that intentionally collected (corbiculae) pollen is representative of plant species visited for nectar.

4.1.2 Question 2. How effective is the method of using visual morphology when compared to DNA?

I predicted that the list of visited plant species from the DNA method would have higher diversity and accuracy than the list of visited plant species from visual morphology. The prediction regarding the diversity of identified species/genera was found to be the case as I only identified 10 plant genera in the pollen samples, compared to the 24 identified plant genera using pollen DNA. The prediction regarding the accuracy of correctly identifying species/genera was also found to be the case, as I had difficulty identifying pollen, even with assistance from a palynologist. In particular, the Asteraceae family has similar pollen across the clade and eight genera were found in the pollen samples using DNA (see Table 5), leading me to doubt the accuracy of identifying and distinguishing this pollen visually.

4.1.3 Question 3. Does the list of plants visited for pollen match the list for nectar from DNA sequences?

I predicted that the list of plants visited for pollen will not match the list for nectar. In hindsight this was a poor prediction, because why would a plant evolve a strategy of producing nectar, without the benefit of pollen distribution, even if nectar robbing was occurring, it would be highly unlikely to be significant

enough that limited pollen was being spread (thus the bees in my study would lack the pollen), as this would be unsustainable. Although this strategy has occurred before, it's generally in 'domesticated plants' like Calendula (Pot Marigold), Clarkia (Winecup) and Cosmos.

Although cpDNA sequences from 11 genera were identified from stomach exclusive DNA total reads from these 11 plant genera made up 0.37% of the total reads collected in this study, compared to 96.96% of reads which came from the 14 species present in the stomach and leg pollen. Some of the 11 plant genera detected in the stomach DNA are usually wind pollinated (e.g. *Vitis* sp.) so the presence of their pollen probably does not indicate nectar feeding on these plants by the bee. Although the list of plants visited for pollen does not match the list from the stomach, the abundant plant genera were a close match.

The lists of plant species identified will give information about the diets of local bumblebees, but the small sample size will likely be too small to give a complete picture.

Table 7. Families identified by DNA (The combined reads from both pollen and stomach samples).

Family	Genera detected	Common name	Reads	%
Amaranthaceae	Beta	Beets (Beetroot)	345	0.010
Asteraceae	Cirsium	Thistles	1025	0.029
Asteraceae	Crepis	hawksbeard	841	0.024
Asteraceae	Helianthus	Sunflower	369291	10.409
Asteraceae	Hypochaeris	Catsear	134773	3.799
Asteraceae	Leontodon	Hawkbit	117533	3.313
Asteraceae	Picris	Oxtongue	9168	0.258
Asteraceae	Taraxacum	Dandilion	50155	1.414
Asteraceae	Tussilago	Colts foot	970	0.027
Brassicaceae	Arabidopsis	Thale cress	182	0.005
Cannabaceae	Cannabis	Marijuana	349	0.010
Convolvulaceae	Calystegia	False bindweeds	1222	0.034
Fabaceae	Galega	Goat's rue	373725	10.534
Fabaceae	Medicago	Alfalfa	7940	0.224
Fabaceae	Ulex	Gorse	151585	4.273
Fabaceae	Trifolium	Clover	947612	26.709
Hydrangeaceae	Hydrangea	Hortensia	133	0.004
Lamiaceae	Prunella	Heal-all	135	0.004
Malvaceae	Hoheria	LaceBark	189735	5.348
Malvaceae	Malva	Mallow	18880	0.532
Musaceae	Musa	Plantain (Bananas)	1148	0.032
Myrtaceae	Eucalyptus	Gum Tree	13227	0.373
Nelumbonaceae	Lotus	Birdfoot Trefoil	53913	1.520
Passifloraceae	Passiflora	Passion flowers	339829	9.578
Plantaginaceae	Veronica	Speedwells(Hebe(Kc	688456	19.405
Poaceae	Hordeum	Barleys	518	0.015
Polygonaceae	Muehlenbeckia	Maidenhair	112	0.003
Ranunculaceae	Clematis	Leather flower	9518	0.268
Ranunculaceae	Ranunculus	Buttercup	332	0.009
Rosaceae	Potentilla	Cinquefoil	237	0.007
Rosaceae	Rosa	Rose	62641	1.766
Rosaceae	Rubus	Brambles	747	0.021
Solanaceae	Lycium	Boxthorn	1426	0.040
Solanaceae	Physalis	Groundcherry (Goos	103	0.003
Vitaceae	Vitis	Grapevines	110	0.003

4.2 Unexpected or contradictory findings

4.2.1 The Non-plant identifications

1.04% of identified pollen DNA reads were excluded while 25.64% of identified stomach contents DNA reads were excluded. Some of these identifications ranged from impossible (e.g. *Nycticebus* Slow Loris) to plausible, such as *Dyella* which is a genus of bacteria identified in 2005 in New Zealand. Most of these can be explained with poor coverage and match percentage, for example, the *Nycticebus* identified sequence (159 reads) had a query coverage of 14% (and was still the closest match). Therefore, in the bee stomach, the primers amplified DNA that was not from plant chloroplasts. Failure to find a close match reveals the limitations of the reference database, and most of the 'suspicious' identifications had less than 50% coverage with putative reference matches. Another possible limitation was the short length of reads used (225bp), which was touched on in the methods. However, the high mapping success of the pollen DNA leads me to believe that the impact of this was low and other limitations were the cause of the poor mapping in the stomach DNA samples.

Some of the sequences that were not attributed to Viridiplantae nevertheless had relatively high coverage and match such as 10 unique sequences that all were identified to *Providencia*, so to estimate the plausibility, I checked where in the reference genomes they were mapping.

My script that referenced the unique sequences was extremely basic, and did not record the region the primers mapped to, so to further understand, I manually referenced some of the plant and bacteria sequences to check where they were mapping. I used trnL-c and trnL-d primers, which are designed to amplify a broader range (~500bp) much wider than the target tRNA-Leucine region so I expected some of the sequences may have mapped to other regions (Probojati, et al., 2023).

The top NCBI match for most of the Trifolium sequences (that were manually checked) mapped to the intron within the trnL gene, trnL-UAA which was expected but some also had exact matches to regions that mapped to the trnH-GUG gene. These are both closely related groups so not that unsurprising. But some of the other plant genera sequences mapped to a broader range of regions, such as Helianthus which had some sequences map to many ribosomal protein genes such as 16s which is the gene target most commonly used for bacterial identification (Manaka et al., 2017). The sequences that were identified as bacteria (that were manually checked) also mapped to 16s and other ribosomal protein genes. Because of this I predict these bacterial identifications to be correct.

This is not surprising as chloroplast evolved from cyanobacteria through endosymbiosis (Stadnichuk & Kusnetsov, 2021), and many regions are so highly conserved (such as trnL) that they correspond with

bacterial plastomes (similar enough for some primers to anneal) (Micheli et al., 2014). Previous bacterial diversity studies have used trnL primers and found it less effective than 16s rRNA, but it was successfully used for Cyanobacteria (Micheli et al., 2014). Although *Providencia* is in Pseudomonadota rather than Cyanobacteria, I still believe the identification is correct as the unique sequences have a high coverage, a match of 100%, and they map to a plausible location.

Although the bacterial identifications are easy to explain, the stomach samples contained 90,784 reads identified to *Bombus*. As the stomach samples are from *Bombus* this seems plausible, but trnL primers should not amplify Metazoan DNA (in theory). Like the *Providencia*, these have high coverage and match (~100%), but the references show they are mapping to general assembled chromosomes rather than identified genes. One potential reason for this could be that these *Bombus* genomes the sequences are mapping to have bioinformatics errors that included bacterial DNA from the bee's microbiome into the genomes when they were assembled (Chrisman et al., 2022). Alternatively, fragments of mtDNA can be mistakenly incorporated into the nuclear DNA sequence during assembly too, or a phenomenon known as "NUMT" (Nuclear Sequence of Mitochondrial Origin), where mitochondrial DNA is actually in the nuclear genome (Hazkani-Covo et al., 2010). All of which may introduce ribosomal protein genes like 16s into the nuclear genomes.

All of these non-plant identifications are examples of errors and biases of sequencing and bioinformatics that need to be considered. These sorts of errors only occurred in the stomach samples so future studies would likely avoid them by using pollen samples.

4.2.2 Discrepancies between quantitative outcomes of metabarcoding and microscopy

If all the methods used in this study were perfect at identifying the plant species the bees visit, they would have identical diversity and proportions. Obviously, the lists of genera identified through pollen morphology, pollen DNA and stomach DNA vary, highlighting faults with at least two if not all three methods.

As seen in table 6. Mean Leg/Pollen (Microscopy/Metabarcoding) had a Bray–Curtis dissimilarity of 0.603 while Mean Leg/Stomach (Micro/Meta) had a score of 0.908. The washed pollen had similar results with 0.710 for Mean Washed/Pollen and 0.908 Mean Washed/Stomach. These results show moderate to almost no overlap in lists of genera identified from the different methods.

Some of these discrepancies can probably be explained though measuring biases. For example, the morphology method measures number of granules while DNA amplification measures the amount of

DNA. *Lotus* pollen is far smaller than many other pollen granules, such as *Trifolium*, so although it was the most numerous, the *Trifolium* may have more cpDNA overall. This would explain why the pollen morphology figures show *Lotus* was the most common genus, while only being 10th most common in the DNA figures.

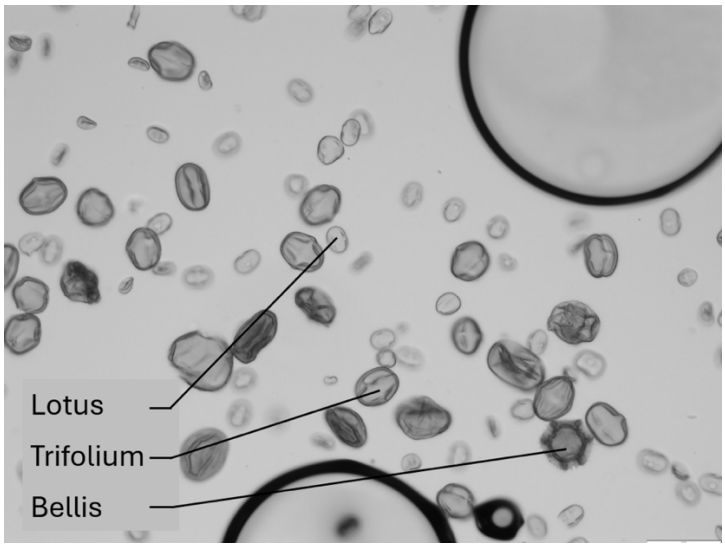

Figure 4.1. Three plant species represented by leg pollen from bee R1.

However, I predict identification inaccuracy is likely to have a bigger impact on plant list discrepancies. Only ten genera were identified in the morphology study compared to the 35 using DNA sequences. This was expected as pollen is difficult to identify and distinguish while DNA identification is generally more reliable. Even excluding the insignificant reads (sequences making up <1% of reads) the DNA list has 12 genera; thus, it can be inferred that the morphological identified plant list is missing genera, and the DNA list is the more accurate one.

4.2.3 The apparent lack of *Raphanus* in the DNA

However, in the pollen morphology counts, 17.5% of the pollen granules were identified to genus *Raphanus*. Nearly all the counted pollen granules came from *Bombus terrestris* bee-11, which was caught on a wild radish plant (*Raphanus raphanistrum*) and the pollen granules visually identified to *Raphanus* were confirmed by a trained palynologist, and yet despite this evidence none of the DNA reads matched to this genus using metabarcoding. Given my direct observations, I suggest that there was *Raphanus* in the pollen samples, and the lack of reads is due to a flaw in the DNA process. The first explanation for poor identification is a lack of references in the database, but *Raphanus* is found globally and grown for crops,

so there are plenty of references on NCBI. However, a previous study has found *Raphanus* to display high genetic variation and introgression (Zhang et al., 2021). This study sequenced 11 *Raphanus* plants from across the northern hemisphere and they found extensive gene flow, genome reshuffling, and large-fragment translocation between domesticated and wild *Raphanus* plants. Although most of this would affect the nuclear rather than the chloroplast genomes, studies have shown hybridization can lead to the creation of cpDNA polymorphism in the hybrid offspring (Hollingsworth et al., 1999; Bhandari & Park, 2022). This combined with most references coming from northern hemisphere species means even if *Raphanus* has many references on NCBI, local wild *Raphanus* may have become too distant and could be being mis-identified by NCBI and the python script.

Another possibility is that the problem occurred before the sequencing. The GeneJET Plant Genomic DNA Purification Kit used in this study warned that phenolics-rich plants like oilseed rape may have reduced effectiveness. *Raphanus* is also high polyphenols and polysaccharides like oilseed rape and most other plants within the Brassicaceae family, and so, may explain the lack of identified reads (Gamba et al., 2021). This is further supported by the lack of other brassicas, as only 182 reads (0.005% of total reads) were identified to genera within Brassicaceae, and they were all from one unique sequence identified to *Arabidopsis* with poor coverage so are probably something else. Failure to detect Brassicaceae pollen might be a real disadvantage with the protocol used here and should be further explored.

4.3 Limitations

The goal of this project was to compare three methods of identifying bumblebee diet and determine the best. I have found that all three have large drawbacks affecting things like time, cost and accuracy, but despite all this, it appears that using pollen DNA is the best method.

4.3.1 Method 1. Pollen Morphology Identification

I found that if starting with the fresh bees, and ending with the list plants/genera, this process took a week of full days. The collection of pollen, acetolysis, and creation of slides, took a full day and the rest of the time was spent identifying the types of pollen and then counting them. Hours of this week were spent learning so a trained palynologist would take a much shorter time. The cost of this method is relatively low as the consumables (slides and acids) used a far cheaper than those used in genetic work. As well as this access to high quality microscopes and a palynology lab is also required (although the acetolysis process could be done in most labs).

There are many steps along this process where accuracy can be lost. Pollen can be distributed through the air so to prevent/reduce contamination, the bees were placed in plastic bags when caught and then the pollen was collected in a microbiology lab and immediately transferred to 1.5mL micro tubes, to reduce air exposure. The acetolysis and slide mounting was done in a dedicated palynology lab with positive air pressure to prevent pollen and spores from entering the room. All this work was done in autumn which has lower air pollen counts than the other seasons too. Because of these measures I have high confidence in the samples having little to no contamination. However, I found accurate and consistent pollen identification to be difficult. The GNS New Zealand Bee Pollen Catalogue (Raine & Newstrom-Lloyd, 2022) was used as well as online photos to identify the types and then refence photos were made using online and my own photos to compare against when counting the pollen in the 258 photos.

Most of the pollen granules (91%) were identified to be *Lotus, Raphanus, Trifolium* and *Passiflora* pollen which was confirmed by an experienced palynologist. DNA reads from *Lotus, Trifolium* and *Passiflora* were also matched in the pollen DNA samples further supporting these identifications to be correct. However, 17.5% of the pollen granules were identified to genus *Raphanus*.

4.3.2 Method 2. Pollen DNA identification

This method took months from start to finish, but most of that time was spent experimenting and developing a workflow, plus waiting for the reads to come back from the sequencing company (Custom Science Ltd.). Now that the workflow for extraction and bioinformatics is established, if I was to do this experiment with more pollen samples, it would take a couple hours, before and after the sequencing. That length of time is up to the method/company you choose. The cost of this method, in comparison to the previous, is relatively high as this method involves far more consumables, and DNA extraction and amplification consumables are often highly specialised and expensive. On top of this mass amplicon sequencing is also often expensive. Like the previous method, there are many steps where accuracy can be lost, but this is reduced with good practices like using clean equipment and labs. During PCR a control was used each time to monitor contamination. The other main limitation is the quality of the referencing databases. Over 99% of the reads in the pollen samples were successfully identified which resulted is a combined 2,658,705 reads. As discussed in the results, some of these had to be excluded due to poor match/coverage on the identification (like 40bp out of 225 or a match below 90%, although some were far lower). But 98.95% were kept and used in the final figures. The only concern I have with this (other than cost) is DNA extraction and amplification bias, as seen with *Raphanus*, if using this method in future, testing should be done to see if it works on *Raphanus* or other Brassicas.

4.3.3 Method 3. Stomach contents DNA identification

This method had the same time and cost limitations as method 2. but it had even further accuracy limitations. The stomach removal requires extremely fine motor skills (which may be an issue for some), and if the bee has not feed in a while when it is caught, there is almost no stomach contents, as was found in some of the practice bees. The pollen method also potentially has this issue, but at least you can see how much pollen it has when caught. Even if the stomach is full, its mostly nectar, the trnL primers amplify the chloroplast which comes from accidently ingested plant material (like pollen) which is why the stomach samples amplified less DNA than the pollen. Custom science reported difficulty amplify the stomach DNA samples and had to do multiple rounds of amplification and increase from 35 to 40 cycles. Once they were sequenced, they had lower referencing success than the pollen, the lowest 85.1% (see Table 2. Bioinformatic read exclusion). Again, like the pollen samples, reads had to excluded due to poor match/coverage, then exclude even more reads despite their high match percentage and coverage, as they were not Viridiplantae, like Bacteria and Metazoa, further reducing the usable DNA. The worst example was for bee 099 (S099S15) which only 5.6% of the originally identified DNA made it into the final results (see Table 2.). This could potentially be reduced using different primers, as trnL should (in theory) only amplify chloroplast DNA.

Like all things, the best method will depend on a project's time and funding, but with all these limitations considered, Method 2. Pollen DNA identification is deemed the optimal approach. The results show that the original concerns about excluding species being visited for nectar by using pollen are not supported as 96.96% of the reads came from species found using both methods (pollen and stomach), and the stomach exclusive species had so few reads (<1% of total reads) that they are not considered to be an important food source. Once the pollen exine is broken, the DNA extraction, amplification, sequencing and bioinformatics is simple, compared to the stomach contents, which has issues at every step.

5. Conclusion

The original general question of this study was ‘How can farmers increase bumblebee populations on their land?’, which led to the question ‘What plants are bumblebees in the in New Zealand visiting?’. Now that the sub questions of this project have been answered, there is evidence to show:

The intentionally collected pollen in corbiculae does match (in both diversity and proportion) the unintentionally collected pollen on the rest of the exterior body.

Identifying pollen using visual morphology is cheap but time consuming, requires palynology education and even then, is not particularly reliable, compared to DNA identification, which is expensive but consistent, reliable and despite the limitations, still reasonably accurate.

When excluding insignificant proportions of the reads (genera making up <1%) the list of plants visited for pollen has significant overlap with the list of plants visited for nectar (in both diversity and proportion).

Now that the pollen DNA has shown to be the optimal method, and the processing and bioinformatic tools have been developed, a larger scale project with a large sample size could be run efficiently. Although this project has provided a list of the most significant plant species for the sampled bees, five bumblebees, two native *Leioproctus* bees and one honeybee is too small a sample size to consider representative. A future project could also investigate patterns of plant selection, such as protein content. Bumblebees are known to have less complex foraging behaviour than honeybees, but they still show evidence of communication and non-random selection (Herascu, 2017), and so farmers would still likely benefit from the knowledge of what plants bumblebees in New Zealand visit.

References

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

Aizen, M. A., Aguiar, S., Biesmeijer, J. C., Garibaldi, L. A., Inouye, D. W., Jung, C., ... & Seymour, C. L. (2019). Global agricultural productivity is threatened by increasing pollinator dependence without a parallel increase in crop diversification. Global change biology, 25(10), 3516-3527.

Baer, B. (2018) The honeybee hive. Australian Academy of Science. The honeybee hive - Curious

Bayram, N. E. (2021). Molecular approach to detecting pollen types in honey: DNA barcoding. Hacettepe Journal of Biology and Chemistry, 49(1), 37-41.

Callahan, B. (2021). Binned quality scores and their effect on (non-decreasing) trans rates. Github. https://github.com/benjjneb/dada2/issues/1307

Bhandari, G. S., & Park, C. W. (2022). Molecular evidence for natural hybridization between Rumex crispus and R. obtusifolius (Polygonaceae) in Korea. Scientific Reports, 12(1), 5423.

Black, A. D., Laidlaw, A. S., Moot, D. J., & O'Kiely, P. (2009). Comparative growth and management of white and red clovers. Irish Journal of Agricultural and Food Research, 149-166.

Bontšutšnaja A, Karise R, Mänd M, Smagghe G. Bumble bee foraged pollen analyses in spring-time in Southern Estonia shows abundant food sources. Insects. 2021 Oct 9;12(10):922. doi: 10.3390/insects12100922

Broussard, M. A., Coates, M., & Martinsen, P. (2023). Artificial pollination technologies: A review. Agronomy, 13(5), 1351.

Cameron, S. A., & Sadd, B. M. (2020). Global trends in bumble bee health. Annual review of entomology, 65, 209-232.

Dahl, O., A. (2025). Floral Cues and Foraging Preferences in Various Species of Wild Bumblebees. The International Journal of Ecopsychology (IJE): Vol. 10: Iss. 1, Article 5

Darwin, C. (1860). A naturalist's voyage round the world. Journal of researches.

Do, D., & Zaveska, D, L. (2018). Herbarium tale: the utility of dry specimens for DNA barcoding Juncaceae. Plant Systematics and Evolution, 304(2), 281-294.

Free, J. B., & Butler, C. G. (1959). Bumblebees. (No Title).

Gamba, M., Asllanaj, E., Raguindin, P. F., Glisic, M., Franco, O. H., Minder, B., ... & Muka, T. (2021). Nutritional and phytochemical characterization of radish (Raphanus sativus): A systematic review. Trends in Food Science & Technology, 113, 205-218.

Goulson D, Hanley ME (2004) Distribution and forage use of exotic bumblebees in South Island, New Zealand. NZ J Ecol 28:225–232. https://www.jstor.org/stable/24056431.

Hazkani-Covo, E., Zeller, R. M., & Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS genetics, 6(2), e1000834.

Herascu, R. A. L. U. C. A. (2017). Bumblebee navigation and foraging behaviour: a short review. Field Studies, 491.

Hollingsworth, M. L., BAILEY, J. P., Hollingsworth, P. M., & FERRIS, C. (1999). Chloroplast DNA variation and hybridization between invasive populations of Japanese knotweed and giant knotweed (Fallopia, Polygonaceae). Botanical Journal of the Linnean Society, 129(2), 139-154.

Illumina. (2025). Measuring sequencing accuracy. Illumina.

Kaftanoglu, O., Sommeijer, M., & de Ruitjer, A. (2000). The diversity and faunistics of bumblebees for pollination in greenhouses. Insect pollination in greenhouse. Netherlands: Department of Social Insects, 73-82.

Karbassioon, A., Yearlsey, J., Dirilgen, T., Hodge, S., Stout, J. C., & Stanley, D. A. (2023). Responses in honeybee and bumblebee activity to changes in weather conditions. Oecologia, 201(3), 689-701.

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.

Leonhardt, S.D., Blüthgen, N. The same, but different: pollen foraging in honeybee and bumblebee colonies. Apidologie 43, 449–464 (2012).

Li, H., Wu, K., Ruan, C., Pan, J., Wang, Y., & Long, H. (2019). Cost-reduction strategies in massive genomics experiments. Marine Life Science & Technology, 1, 15-21.

Lihoreau, M., Monchanin, C., Lacombrade, M. et al. (2025). Why bumblebees have become model species in apidology: A brief history and perspectives. Apidologie 56, 19

Lowe, A., Jones, L., Witter, L., Creer, S., & de Vere, N. (2022). Using DNA metabarcoding to identify floral visitation by pollinators. Diversity, 14(4), 236.

Lye, G. C., Lepais, O., & Goulson, D. (2011). Reconstructing demographic events from population genetic data: the introduction of bumblebees to New Zealand. Molecular Ecology, 20(14), 2888-2900.

Mallott, E. K., Garber, P. A., & Malhi, R. S. (2018). trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (Cebus capucinus, Primates). PloS one, 13(6), e0199556.

Manaka, A., Tokue, Y., & Murakami, M. (2017). Comparison of 16S ribosomal RNA gene sequence analysis and conventional culture in the environmental survey of a hospital. Journal of Pharmaceutical Health Care and Sciences, 3, 1-6.

Mashilingi, S. K., Zhang, H., Garibaldi, L. A., & An, J. (2022). Honeybees are far too insufficient to supply optimum pollination services in agricultural systems worldwide. Agriculture, Ecosystems & Environment, 335, 108003.

Micheli, C., Cianchi, R., Paperi, R., Belmonte, A., & Pushparaj, B. (2014). Antarctic cyanobacteria biodiversity based on ITS and trnL sequencing and its ecological implication. Open Journal of Ecology, 4(8), 456-467.

Milla, L., Sniderman, K., Lines, R., Mousavi-Derazmahalleh, M., & Encinas-Viso, F. (2021). Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. Ecology and Evolution, 11(13), 8683-8698.

Morandin, L. A., Laverty, T. M., & Kevan, P. G. (2001). Bumble bee (Hymenoptera: Apidae) activity and pollination levels in commercial tomato greenhouses. Journal of economic entomology, 94(2), 462-467.

Mujeeb, S. A., Divya, P., Chaitanya, G., & Manisha, B. L. (2023). The Vital Role of Pollinators in Agriculture and Ecosystems. Integrated Publications TM New Delhi, 25.

Nabhan, G. P., & Buchmann, S. L. (1997). Services provided by pollinators. Nature's Services: societal dependence on natural ecosystems, 133-150.

Nayak, R. K., Rana, K., Bairwa, V. K., Singh, P., & Bharthi, V. D. (2020). A review on role of bumblebee pollination in fruits and vegetables. Journal of Pharmacognosy and Phytochemistry, 9(3), 1328-1334.

Parker, H. E. (2022). Honey and humble: Bee introductions, environment, and ideology in Aotearoa New Zealand, 1839-1900 (Doctoral dissertation, The University of Waikato).

Porter, T. M., & Hajibabaei, M. (2018). Automated high throughput animal CO1 metabarcode classification. Scientific Reports, 8(1), 4226.

Probojati, R. T., Hadiyanti, N., & Hapsari, L. (2023). Assessment for identification of Stelechocarpus burahol and sister species complex of Annonaceae family using trnL intron sequences. Jurnal Biologi Tropis, 23(4), 643-649.

Qaigen. (2025). CLC Genomics Workbench 25.0.2. 37.4 Merge Overlapping Pairs. https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/User_Manual.pdf#page=1188.31

Rahmati-Joneidabad, M., Behbahani, B. A., Taki, M., Hesarinejad, M. A., & Toker, O. S. (2024). Evaluation of the probiotic, anti-microbial, anti-biofilm, and safety properties of Levilactobacillus brevis Lb13H. LWT, 207, 116636.

Raine, J.I., Li, X., & Newstrom-Lloyd, L. (2022). New Zealand bee pollen catalogue.

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome research, 22(5), 939-946.

Schmid-Hempel, P., Schmid-Hempel, R., Brunner, P. C., Seeman, O. D., & Allen, G. R. (2007). Invasion success of the bumblebee, Bombus terrestris, despite a drastic genetic bottleneck. Heredity, 99(4), 414-422

Sommer, J., Rao, V. & Sprayberry, J. (2022) Deconstructing and contextualizing foraging behavior in bumble bees and other central place foragers. Apidologie 53, 32.

Stahlmann-Brown, P., Hall, R. J., Pragert, H., & Robertson, T. (2022). Varroa appears to drive persistent increases in New Zealand colony losses. Insects, 13(7), 589.

Stein, K., Coulibaly, D., Stenchly, K., Goetze, D., Porembski, S., Lindner, A., Konaté, S., & Linsenmair, E. K. (2017). Bee pollination increases yield quantity and quality of cash crops in Burkina Faso, West Africa. *Scientific reports*, *7*(1), 17691. https://doi.org/10.1038/s41598-017-17970-2

Stern, R. A., Rozen, A., Eshed, R., Zviran, T., Sisai, I., Sherman, A., ... & Sapir, G. (2021). Bumblebees (Bombus terrestris) improve 'Hass' avocado (Persea americana) pollination. Plants, 10(7), 1372.

Syms, P. & Denton, M. (1996). Bees & Trees. New Zealand Tree Crops Association. Bees & Trees - New Zealand Tree Crops Association (NZTCA)

Velthuis, H. & Doorn, Adriaan. (2006). A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. http://dx.doi.org/10.1051/apido:2006019. 37. 10.1051/apido:2006019.

Vleugels, T., Amdahl, H., Roldán-Ruiz, I., & Cnops, G. (2019). Factors Underlying Seed Yield in Red Clover: Review of Current Knowledge and Perspectives. Agronomy, 9(12), 829. https://doi.org/10.3390/agronomy9120829

What's all the buzz about. (2024). BioBees. What's all the buzz about? – Biobees Ltd

Zameer, S., Ali, M., Sajjad, A., Saeed, S., Matloob, A., Bashir, M. A., ... & Atta, S. (2022). Foraging behavior and visit optimization of bumblebees for the pollination of greenhouse tomatoes. Journal of King Saud University-Science, 34(1), 101744.

Zhang, C. J., Gao, Y., Kim, D. S., Yu, J., Jiang, C., Wang, Y., ... & Yan, X. (2022). Bumblebees are more efficient than honeybees to facilitate wind-blown pollen dispersal of alfalfa (Medicago sativa L.). Agriculture, Ecosystems & Environment, 340, 108161.

Zhang, X., Liu, T., Wang, J., Wang, P., Qiu, Y., Zhao, W., ... & Li, X. (2021). Pan-genome of Raphanus highlights genetic variation and introgression among domesticated, wild, and weedy radishes. Molecular Plant, 14(12), 2032-2055.

Appendix

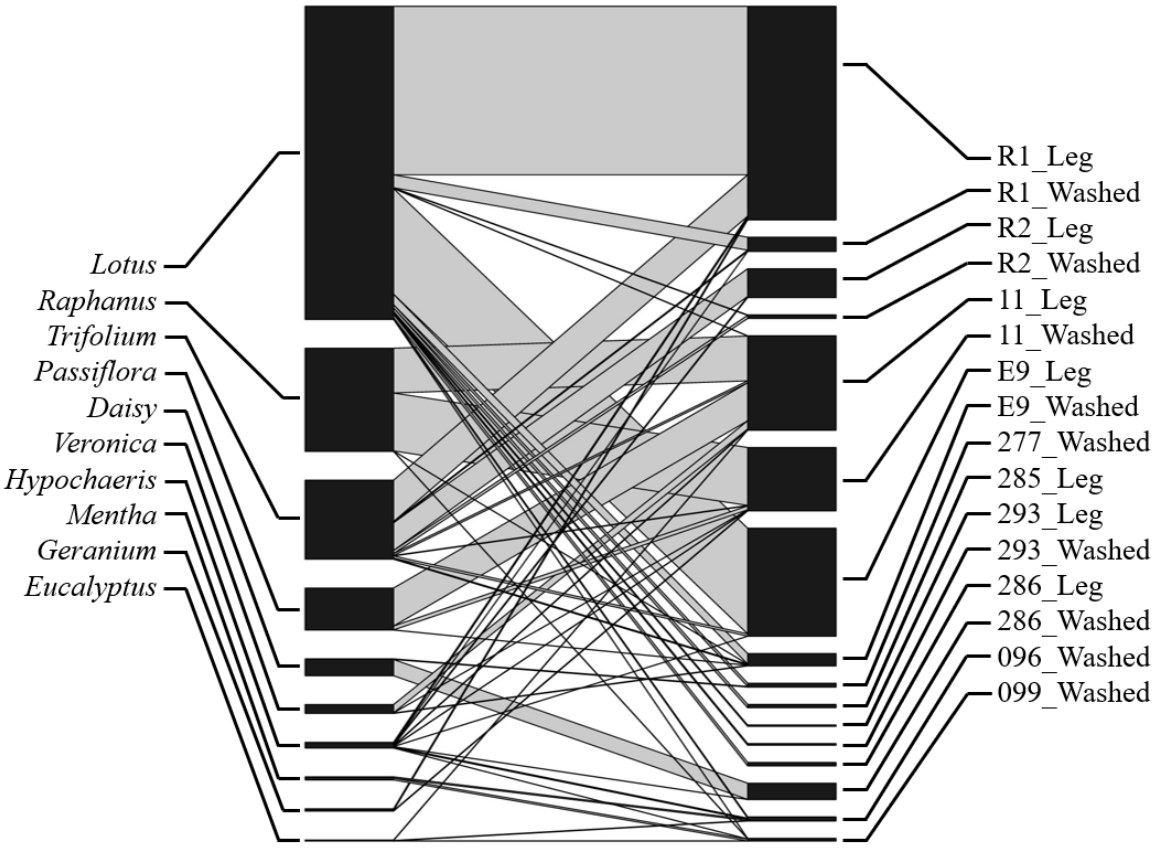

Figure A.1 The proportion and distribution of the plant genera morphologically identified from pollen from all samples from *Bombus terrestris, Apis mellifera* and *Leioproctus monticola* collected in New Zealand illustrated with a bipartite graph. Left column shows the number pollen granules counted and identified to a genus and the right column shows the number pollen granules counted from the leg and washed pollen samples.

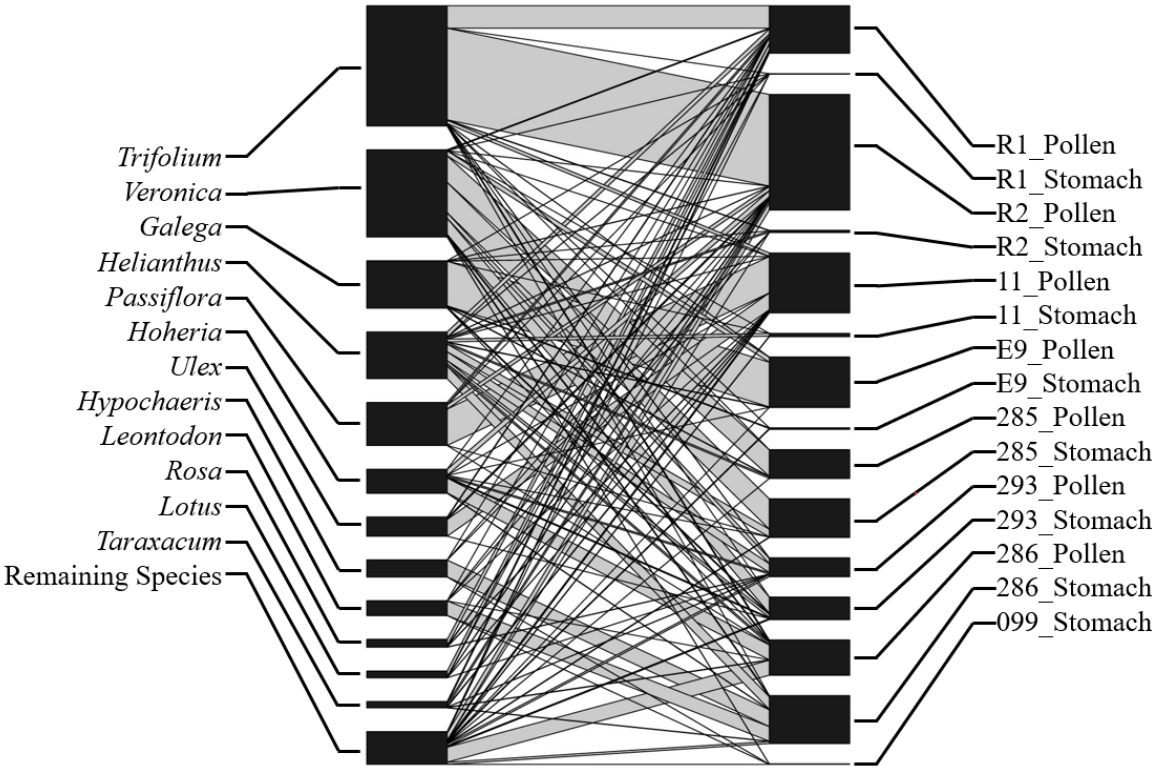

Figure A.2 The proportion and distribution of the plant genera identified from DNA sequences from all samples from *Bombus terrestris, Apis mellifera* and *Leioproctus monticola* collected in New Zealand illustrated with a bipartite graph. Left column shows the number of reads referenced to a plant genus and the right column shows the relative number of reads sequenced from each sample (external pollen or stomach contents).