THE EVOLUTION OF DIET WIDTH IN *LASIOGLOSSUM* (*CHILALICTUS*) (HALICTIDAE, APOIDEA), IN ASSOCIATION WITH SPECIATION AND SENSORY MORPHOLOGY

Trace Akankunda

M.Bio (PB), The University of Adelaide

Thesis submitted to The University of Adelaide for the degree of Doctor of Philosophy

Department of Plant Protection

School of Agriculture, Food & Wine

Faculty of Sciences, The University of Adelaide

Waite Campus

April 2019

Table of contents

List of tables

List of figures

List of supplementary material

Table

Figures

List of electronic files

Figure S3.1. A maximum clade credibility tree of the Bayesian phylogenetic analysis of 107 *L. (Chilalictus)* species based on 80 morphological characters. Well supported species clades that are consistent with the molecular phylogeny are marked with an asterisk (*). Species that consistently group together in the molecular phylogeny are highlighted with the same colour. The arrow shows a well-supported clade of species comprising *L. supralucens, L. mirandum, L. calophyllae* and *L. tamburinei,* which is not fully recovered in our molecular phylogeny.

Figure S3.2. A maximum clade credibility tree of the Bayesian phylogenetic analysis of 51 *L. (Chilalictus)* species based on 2 nuclear genes (WNT and EF 1 alpha; 2nd copy) and one mitochondrial gene (CO1) and showing all species replicates (sequences from 91 specimens). Posterior probabilities < 99% are shown above the nodes.

Figure S4.1. The mean number of placodea on the antennae of polylectic and broadly polylectic species, summed over flagella segments 5 – 11.

Figure S4.2. The mean number of trachodea on the antennae of polylectic and broadly polylectic species, summed over flagella segments 5 – 11.

Data S3.1 List of morphological characters used to assemble the data matrix for phylogenetic analysis of Lasioglossum (*Chilalictus*).

Data S3.2 Likelihood estimates of the ancestral state reconstruction without model restrictions imposed using MrBayes.

Data S3.3 Likelihood estimates of the ancestral state reconstruction with model restrictions imposed using MrBayes.

Data S4.1 Morphological measurements of 51 species of *L.* (*Chilalictus*) used for comparative phylogenetic analysis.

Summary

Bees are highly dependent on flowers as a source of food, from which they collect oil, nectar or pollen. While most bees rely on multiple plants for nectar, some species specialize on specific pollen hosts while other are generalists. Robertson (1925) was the first to document this observation and he coined the terms monolecty, oligolecty and polylecty to categorise bees based on their pollen host breadth. Since then, several researchers of pollination systems have debated the definitions and biological relevance of these categories. The most significant advancement in this debate comes from the work of Cane and Sipes (2006) who subdivided the categories of diet width further and revised the lexicon used for classifying bees based on their diet width. While their revised classification system remains contentious, it provides a useful framework for representing the various levels of pollen host specialization and generalization observed in bees. Of particular interest in this study, is the category of broadly polylectic bees because despite their economic, ecological and evolutionary significance, as far as we know, no work has been done to understand the factors driving or maintaining the evolution of broad polylecty in bees.

Flower visitation records of bees suggest that broad polylecty is rare across the families of bees and is predominantly observed in two families: Apidae and Halictidae. In order to develop a hypothesis on the factors influencing the evolution of broadly polylectic behaviour in bees, I reviewed literature on the evolution of pollen host breadth in bees and phytophagy in herbivorous insects. As in herbivores, the evolution of diet width in bees is bidirectional towards oligolecty and polylecty depending on different ecological factors operating within a given microhabitat. However, the selection of pollen hosts appears to be constrained by neurological, physiological or morphological factors in

both oligolectic and polylectic species. Given that broadly polylectic behaviour appears to be taxonomically conserved, I hypothesize that genetic constraints limit the evolution of polylecty to broad polylecty in most bee taxa and that these constraints can be deduced by comparing morphological, neurological and digestive physiological differences between polylectic and broadly polylectic species. To test this hypothesis, I studied the evolution of broad polylecty in a native Australian subgenus of *Lasioglossum* (*Chilalictus*). In accordance with observations from the literature, broad polylecty is rare in this subgenus and I found that it has evolved independently at least four times. However, in examining the differences between broadly polylectic species and species in lower pollen host breadth categories, I found no evidence of morphological adaptations for broad polylecty in the peripheral olfactory sensory organs of *L.* (*Chilalictus*) species. Possible directions of further research on the factors influencing the evolution of broad polylecty in bees are explored.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Trace Akankunda Date: 4/ 12 /2019

Acknowledgements

First, I want to thank the University of Adelaide for sponsoring my PhD programme and giving me access to numerous resources that have enabled me to finish well. I then wish to express my sincere gratitude to my supervisors Katja Hogendoorn, Carlos Rodriguez Lopez and Remko Leijs for guiding me through the hills and valleys of my PhD programme. They have been tireless in demanding for the best quality of work from me, right from the inception of project ideas, through to the execution and presentation. They have made the PhD programme a rich learning experience for me and set me up well for whatever career path I pursue going forward.

I also want to thank specific people who have been involved in my PhD projects; Ken Walker of the Museums Victoria for giving me access to his treasured collection of museum specimens and data that have enabled me to study my research questions, Tessa Bradford of the South Australian Regional Facility for Molecular Ecology and Evolution for supporting me in conducting molecular laboratory experiments, Hien To and her colleagues from the Bioinformatics hub at the University of Adelaide for providing me with support in data analysis, Lisa Anne O'Donovan and Ken Neubauer of the Adelaide microscopy for supporting me in conducting scanning electron microscopy.

I also wish to thank people who have boosted my morale and picked me up from time to time in the course of the PhD programme; members of Katja's research group, Andrew Austin's research group, staff and students at the Waite campus and the Australian Centre for Evolutionary Biology & Biodiversity at the University of Adelaide. Finally I thank my immediate family for loving me from across the seas and my extended family at Trinity church Adelaide for keeping me grounded.

Statement of Authorship

Title of Paper	Why is broad polylecty rare in bees?
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	NA

Principal Author

Name of Principal Author (Candidate)	Trace Akankunda
Contribution to the Paper	Developed the concept, synthesised literature, wrote and revised the the manuscript
Overall percentage (%)	85%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 9/04/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Katja Hogendoorn
Contribution to the Paper	Contributed to concept development, edited and revised the manuscript
Signature	Date 1/4/2019

Name of Co-Author	Carlos Rodriguez Lopez
Contribution to the Paper	Contributed to concept development, edited and revised the manuscript
Signature	Date 3/4/2019

Name of Co-Author	Remko Leijs
Contribution to the Paper	Contributed to concept development, edited and revised the manuscript
Signature	Date 4/4/2019

Chapter 1: General Introduction - Why is broadly polylectic behaviour rare in bees?

Diet width is an important attribute of bees for several reasons. Diet width determines the economic importance of species as pollinators of crops (Klein et al. 2007), as well as the number and strength of interactions that species have with both plants and animals in their habitats and therefore their ecological significance (Martín González et al. 2010). Furthermore, diet width has been found to determine the geographical range size of species (Slatyer et al. 2013), their effective population size (Packer et al. 2005), and their capacity to cope with changing environments (Burkle et al. 2013), which in turn affect their risk of extinction.

The diet width of bees is a continuum, with specialists that forage on a limited number of related plant species and generalists that forage on multiple unrelated plant species (Cane & Sipes 2006). The continuum of diet width has been partitioned into classes that are defined based on the number and taxonomic diversity of the pollen hosts visited by a species. Two similar classification systems have been previously used to group bees into seven categories (Müller & Kuhlmann 2008; Cane & Sipes 2006). However the definitions, biological and ecological relevance of these categories remain contentious (Vossler 2018; Dötterl & Vereecken 2010). Here we use the classification method suggested by Cane & Sipes (2006), because it defines the category of broadly polylectic bees (Table 1.1), which is the focus of this review. This category was not used by Müller & Kuhlmann (2008) because of the difficulty of applying it to species without the exact knowledge of all the flowering plants in a study site.

Table 1.1. Lexicon and criteria for classifying species according to their diet width (Cane & Sipes 2006).

Class of specialisation	Number of plant taxa		
	Species	Genera	Tribes or families
Broad Polylecty	Many	Many	> 25% of available families
Polylecty	Many	Many	4 to < 25% of available families
Mesolecty	Many	>4	1 - 3 families or big tribes
Oligolecty	> 1	1 - 4	1
-Eclectic	> 1	2 - 4	2 - 3
-Narrow	> 1	1	1
Monolecty	1	1	1

Flower visitation records of bees suggest that monolectic behaviour is rare in bees (Minckley & Roulston 2006; Waser et al. 1996). While some species are strict monoleges (González-Varo et al. 2016; Strickler 1979), most oligolectic species will forage on alternative congeneric plant species in the absence of their primary hosts (Cane & Sipes 2006; Wcislo & Cane 1996), but it is not known whether such emergency resources are nutritionally sufficient (Wcislo & Cane 1996). Likewise, broad polylecty might be rare in species of bees because it has been predominantly observed in two of seven bee families, namely Apidae and Halictidae (Vossler 2018; Dalmazzo & Vossler 2015; Cane & Sipes 2006). However, some species from the family Megachilidae, e.g. *Osmia cornuta* and *O. bicornis* have also been identified as broadly polylectic (Haider et al. 2014; Sedivy et al. 2011).

Several hypotheses have been put forward that can explain the rarity of monolecty. These include: (a) the difficulty of synchronising the lifecycles of pollinators with the

flowering periods of individual pollen hosts; (b) the fluctuating abundance of flowers over time and space, which does not guarantee a steady food supply for monolectic species (Waser et al. 1996) and (c), the vulnerability of monolectic species to population bottlenecks and extinction due to narrow geographic distributions (Slatyer et al. 2013) and reduced effective population sizes (Packer et al. 2005), that often characterise species with a narrow diet width. Because of these difficulties in maintaining a monolectic diet, it is not surprising that most of the evolutionary changes in the pollen host breadth of bees occur from oligolecty to polylecty, although reversals also happen (reviewed in Dötterl & Vereecken 2010 and Danforth et al. 2013). Given the evolutionary significance of a broad diet, one would expect polylectic species to adapt to broadly polylectic behaviour. Broadly polylectic behaviour confers a number of evolutionary advantages including resilience to changing environments (Burkle et al. 2013) and the capacity to have a wide geographic distribution (Gupta 2014; Slatyer et al. 2013), which reduces susceptibility to population bottlenecks and extinction, but see Scheper et al. (2014), who found no correlation between diet width and species decline in Netherlands. The disparity between the perceived evolutionary advantage of broad polylecty and its rare occurrence in bee families demands an explanation.

The need for clarification of the evolution of broadly polylectic behaviour may come as a surprise, because studies have in the past defined oligolecty as a derived trait, and in need of an explanation (reviewed by Muller and by Waser et al. in 1996). However, this assumption was challenged based on field observation studies (Minckley & Roulston 2006; Waser et al. 1996) and phylogenetic analysis (Danforth et al. 2013; Danforth et al. 2006; Muller 1996), showing that oligolecty is the most likely ancestral character state in bees. The long-standing misconception about the evolutionary

trajectory of diet width in bees has several possible causes. First, the erroneous definition of diet width as a dichotomy instead of a continuum (Cane & Sipes 2006; Waser et al. 1996), may have contributed to the misunderstanding of the direction of the evolution. Second, the higher-level phylogeny of bees was, until recently poorly understood (Danforth et al. 2006), and hence the directionality of the diet width could not be interpreted correctly. Third, confusion about overall directionality is caused by the fact that all bee families contain oligolectic and polylectic species, but in some families, oligolecty is considered an ancestral trait while in others it is thought to be derived (Minckley & Roulston 2006; Wcislo & Cane 1996). Therefore it is necessary now to examine the factors influencing the evolution of polylectic behaviour and broadly polylectic behaviour in bees.

To try to understand the factors influencing the evolution of broadly polylectic behaviour in bees, we reviewed the literature on the evolution of diet width in bees. Most of the literature focuses on the evolutionary drivers of oligolectic behaviour (Minckley & Roulston 2006; Wcislo & Cane 1996). Because the hypotheses with regards to the evolution of broad polylecty in bees are scant, we have broadened our review to include the evolution of broad diets in herbivorous insects in general, which is justified by the similarities in patterns of plant host use between herbivorous insects and bees (Sedivy et al. 2011; Sedivy et al. 2008; Sipes & Tepedino 2005).

Drivers of polyphagy in herbivorous insects

Similar to the studies of pollen host breadth in bees, studies on the evolution of diet width in herbivorous insects mostly focus on the evolution of host specialization (Singer 2008). This is likely because most herbivorous insects are specialists and specialization is believed to be selectively advantageous in part because it allows species to master the use of specific plants and offers protection from predators (Bernays 1998). However, polyphagy is common in the order Orthoptera and occurs less frequently in other insect orders (Bernays 1998). Several reasons have been provided to explain the evolution of polyphagy in herbivorous insects. In grasshoppers (Orthoptera), polyphagy offers physiological benefits such as increased nymph survival and adult body mass (Miura & Ohsaki 2004), growth rate and fecundity (Bernays & Minkenberg 1997a). Additionally, in laboratory experiments using *Schistocerca americana* (Orthoptera), individuals which were fed on low protein diets, mixed their diets more frequently compared to those fed on high protein diets (Bernays & Raubenheimer 1991; Bernays & Bright 1991). This suggests that polyphagy in grasshoppers is driven by the need to achieve a nutritional balance (Miura & Ohsaki 2004; Bernays & Bright 1991), although it could also be an adaptive response to unpredictable food sources (Singer 2001).

In contrast to species in the order Orthoptera, there were no nutritional benefits observed in feeding experiments using polyphagous Lepidopterans and Hemipterans (Adler 2004; Singer 2001; Bernays & Minkenberg 1997). Instead, other factors such as predator and parasite avoidance, the availability of abundant food choices and the need to dilute toxins from different foods were suggested explanations for polyphagy in these orders (Adler 2004; Bernays & Minkenberg 1997). Based on experiments

investigating the drivers of a mixed diet in caterpillars of two arctiid moth species *Grammia geneura* and *Estigmene acrea* (Singer et al. 2004), Singer (2008) suggested that no single ecological factor might solely be responsible for the patterns of polyphagy observed in herbivores. Instead he hypothesized that a trade-off between several factors might drive herbivores to forage on multiple plant species where they maximize fitness benefits and reduce fitness costs within their microhabitats. In turn, specialization could result when a single host reliably offered the best chances of safety from parasitoids or predators (Bernays 1998) and nutrient quality in comparison to other potential hosts. For example, *Parrhasius polibetes* (Lycaenidae) has achieved both lower exposure to parasitism and a high nutrient diversity by consuming three host plant species (Rodrigues et al. 2010).

In addition, Bernays (1998) stressed that diet expansion in herbivorous insects generally happens under certain dietary constraints, normally associated with the presence or absence of specific secondary metabolites in the newly acquired hosts. For example, with the exception of *Papilio glaucus* (tiger swallowtail), most species in the genus *Papilio* shift between different hosts but appear on the whole to be restricted within five plant families that share similar chemical profiles, namely; Rutaceae, Apiaceae, Asteraceae, Annonaceae and Lauracea (Murphy & Feeny 2006; Feeny 1991).

Drivers of oligolectic behaviour in bees

Similar to the research on herbivorous insects, research into plant-bee interactions has focused on the evolution of oligolectic behaviour in bees (Minckley & Roulston 2006; Wcislo & Cane 1996). The factors that are thought to drive the evolution of oligolecty in bees include: resource partitioning to reduce competition between closely related

species, increase in foraging efficiency, adaptation to plant/flower chemistry (toxins and nutrient content), parasitism and predation, and synchronization with the most dominant and reliable pollen hosts especially in xeric habitats (Minckley & Roulston 2006; Wcislo & Cane 1996; Muller 1996). So far, few studies have provided exclusive support for a single hypothesis as being the driver for oligolecty in bees, while many studies present inconclusive or contradictory results (Minckley & Roulston 2006). It is likely that as suggested by (Singer 2008) for the evolution of polyphagy, no single ecological factor might solely drive the evolution or maintenance of oligolecty in bees. Instead a combination of factors acting together within the microhabitats of bee species might influence their foraging behaviour towards oligolecty or polylecty (Sedivy et al. 2008; Minckley & Roulston 2006; Sipes & Tepedino 2005). In addition, contradictory results may partly be due to the fact that often no distinction is made between primary oligolecty, as a maintained ancestral trait for a species with an oligolectic ancestor, and secondary oligolecty, as a reversal from a polylectic ancestor for an oligolectic species (Minckley & Roulston 2006), while these may not be the result of similar selective pressures.

Nonetheless, there is a general consensus among researchers that pollen host breadth is strongly influenced by the parental and ancestral foraging behaviour (Müller & Kuhlmann 2008; Minckley & Roulston 2006) and is highly conserved in bees (Litman et al. 2011; Dötterl & Vereecken 2010; Larkin et al. 2008; Sedivy et al. 2008; Michez et al. 2008). By mapping host preferences to phylogenies of five genera of Melittidae s.l., Michez et al. (2008) observed that host plant associations were retained in most of the clades. He also noted however, that switches to unrelated pollen hosts from different families were common. Similarly, Sedivy et al. (2008) demonstrated that

pollen host preferences were maintained in the species of the genus *Chelostoma*, a genus that mainly specializes on plants from the family Campanulaceae. He also noted that some of the host switches observed in the *Chelostoma* species were to pollen hosts already utilized by closely related species, a common phenomenon in bees (Sipes & Tepedino 2005) and other herbivores (Futuyma et al. 1995; Janz et al. 2001; Dötterl & Vereecken 2010). Furthermore, in several cases where pollen host switches to distantly related plants had occurred in the group of *Chelostoma* specialists, flowers of several species in the newly incorporated pollen host taxa had a striking resemblance to flowers of the primary host plants (Sedivy et al. 2008), with the exception of *C. florisomne*, a specialist on Ranunculaceae which has no resemblance to the flowers of Campanulaceae. Combining observation from his study with findings on the poor larval performance of *Chelostoma* species reared on non-host pollen (Praz et al. 2008b), Sedivy et al. (2008) proposed an alternative hypothesis for the evolution of foraging behaviour in bees, the 'constraint hypothesis'.

The constraint hypothesis for pollen host selection

According to the constraint hypothesis, the foraging behaviour of bee species can change towards either oligolecty or polylecty depending on local selection forces acting within the species' microhabitat. However, the choice of pollen host plants utilized in either phase of the species' foraging behaviour is constrained by its physiological and or neurological capabilities (Sedivy et al. 2008). A parallel hypothesis termed 'the oscillation hypothesis' was proposed for the evolution of foraging behaviour in phytophagous insects (Janz & Nylin 2008). These authors suggest that herbivores exhibit an evolutionary developmental plasticity towards polyphagy or oligophagy depending on local selection pressures but are constrained in the choice of host plant. This hypothesis was supported by studies on the evolution of

diet width within butterflies in the tribe Nymphalini, where several phases of expansion and contraction in host breadth were observed but also evidence for highly constrained host plant selection towards plants used by ancestral and closely related species (Weingartner et al. 2006; Janz et al. 2001).

Evidence using a quantitative genetics approach showed that constraints in host plant selection of phytophagous insects could have a strong genetic component (Futuyma et al. 1995; Futuyma et al. 1994; Futuyma et al. 1993). Larval performance in four species of the beetle genus *Ophraella* (Chrysomelidae) on their congener's host plants was correlated with their genetic distance to the congener as well as the taxonomic relatedness of their primary host and the congener's host plant. Similar genetic constraints on pollen host selection have been demonstrated in the oligolectic bee *Heriades truncorum* (Megachilidae), as mature bees, regardless of their larval diet, strictly foraged on the primary host family for pollen (Praz et al. 2008b). Likewise, phylogenetic constraints were observed in the pollen host selection of an oligolectic clade in the genus *Andrena*, as host shifts were mostly restricted to the same plant tribe (Larkin et al. 2008). Together, these results suggest that genetic constraints determine the fundamental host plant breadth of a species (in contrast to the realized/observed host breadth determined by locally operating ecological factors) and could manifest in the species' morphology, neurology or physiology (Dötterl & Vereecken 2010; Futuyma & Moreno 1988).

Neurological, digestive physiological and morphological constraints for pollen host selection

In support of the constraint hypothesis, neurological and digestive physiological constraints have been observed in the patterns of pollen host selection in bees (Dötterl

& Vereecken 2010; Sedivy et al. 2008; Praz et al. 2008a; Minckley & Roulston 2006). In addition, morphological adaptations for pollen host use, such as in the vestiture, mouthparts and scopa, that have been observed in several oligolectic bee species (Wcislo & Cane 1996; Walker 1995; Thorp 1979; Parker 1978), suggest that species without such adaptations might be constrained in their pollen host selection.

Neurological constraints for pollen host selection

Evidence for the effects of neurological constraints on pollen host selection was reviewed by Dötterl & Vereecken (2010). Here we have highlighted a few examples and included evidence not cited in their review. Experimental evidence of the effects of imprinting on pollen host selection in *Heriades truncorum* (Megachilidae), has shown that the species is neurologically constrained to its host family of Asteraceae (Praz et al. 2008a). This conclusion was based on the fact that although *H. truncorum* larvae successfully developed on pollen from *Campanula rapunculi* (Campanulaceae) and *Echium vulgare* (Boraginaceae), the emerging adults refused to collect pollen from flowers of the two plants provided and only collected pollen from *Buphthalmum salicifolium* and *Tanacetum vulgare*, both in the primary host family Asteraceae. The neurological constraints associated with pollen host restrictions in bees are visual and olfactory (Dötterl & Vereecken 2010; Sedivy et al. 2008; Praz et al. 2008a). Visual constraints were suggested to explain oligolecty in *Chelostoma* species as distantly related pollen host plants that became incorporated into the diet of some species had strikingly similar floral shapes, morphology or color to the ancestral pollen hosts of the respective species (Sedivy et al. 2008). Similar patterns for pollen host expansion to distantly related plants by oligolectic bees were reported for species in the tribe Anthidiini and genera *Macrotera* (Andrenidae) and *Diadasia* – Apidae (referenced in

Sedivy et al. 2008).

In addition, flowers are known to produce volatile chemicals in their pollen kits or corolla that act as olfactory cues which might constrain bees to particular pollen hosts (Milet-Pinheiro et al. 2013; Dötterl & Vereecken 2010). For example, following observations that *Hoplitis adunca* was attracted to the floral scent produced by *E. vulgare* and not to that produced by *Anchusa officinalis* (both in the family Boraginaceae), an experimental study using artificial flowers revealed that 1,4-benzoquinone was the floral scent compound used by naïve *H. adunca* for host recognition whereas the more experienced individuals recognised more complex mixtures of volatile compounds (Burger et al. 2012). In addition, *Protodiscelis palpalis* (Colletidae) a specialist on *Hydrocleys martii* (Alismataceae) was found to use p-methylanisole for pollen host recognition (Carvalho et al. 2014). These results suggest that constraints in visual and olfactory receptivity could influence pollen host selection in bees (Dötterl & Vereecken 2010).

Furthermore, research on food aversion in honeybees showed that there is a conceivable link between the olfactory stimulus and digestive physiology in bees (Wright et al. 2010). Using associative conditioning of the proboscis extension response (PER) Wright et al. (2010) showed that honey bees could use a post ingestive mechanism via serotonin receptors in the gut to commit odours from toxic plants to memory. A post ingestive mechanism works by associating the harmful effects of pollen or nectar with the host's floral odour (Wright et al. 2010). It follows that neurological signals for pollen host identification in bees (olfactory or visual) could be secondary stimuli following a digestive physiological constraint.

Digestive physiological constraints for pollen host selection

The digestive physiological constraints that have been proposed to explain pollen host specialization in bees have been reviewed by Praz et al. (2008b). Studies on the digestion of pollen in Megachilidae showed that species vary in their efficiency of digesting specific pollen hosts (Suárez-Cervera et al. 1994) and that oligoleges might be more effective at digesting pollen collected from their primary than secondary pollen hosts (Dobson & Peng 1997 and references therein). Additionally, in feeding experiments using larvae of oligolectic bees from the genus *Osmia* (Megachilidae), species failed to develop on non-host pollen (Praz et al. 2008b). Furthermore, bees of the species group *Colletes succinctus* are attracted to host-plants that are highly divergent in color and scent but produce pollen with similar chemical composition indicating that they might be physiologically rather than neurologically limited to exploit alternative flowers (Vanderplanck et al. 2017). However, some oligolectic species have been shown to survive on non-host pollen (reviewed by Wcislo & Cane 1996) and host switches to pollen with diverse chemical profiles have been reported in an oligolectic species group of *Melitta leporina* (Vanderplanck et al. 2017). This shows that digestive physiological constraints may not explain oligolectic behaviour in all cases. Nonetheless, it is evident that pollen from certain plants contains chemical properties that might constrain pollen host selection for bees (Müller & Kuhlmann 2008; Praz et al. 2008b).

Digestive physiology in bees is a collaborative function between the host enzymes and gut bacteria (Kwong & Moran 2015; Engel & Moran 2013). Moreover, research shows that communities of gut bacteria differ in their functional capabilities (Engel & Moran 2013). In addition, the composition of gut bacterial communities is taxonomically

conserved in the corbiculate bees i.e honey bees, bumblebees, orchid bees and stingless bees (Koch et al. 2013; Martinson et al. 2011) and it has been experimentally shown that the gut environment in bees might select for the composition of bacterial communities, possibly through adjusting physiological conditions of the gut to favour some bacterial species or strains over others (Kwong & Moran 2015; Kwong et al. 2014). This has given rise to the hypothesis that differences in composition of the gut bacteria might explain the observed variation in their metabolic capabilities (Engel et al. 2014; Kwong et al. 2014) and therefore pollen host selection might be, at least in part, a consequence of differences in the gut bacterial communities.

Morphological adaptations to pollen hosts

While limited research has been done to show morphological constraints to the collection of pollen from certain hosts, morphological adaptations for pollen collection and harvesting have been reported in several oligolectic species (Thorp 2000; Wcislo & Cane 1996; Walker 1995; Thorp 1979). In his study on pollen host specialization in the genus *Rophitini* (Halictidae), Patiny et al. (2008) observed that facial spines in the species *Rophites* were used to collect pollen from the nototribic flowers of Lamiaceae and that the elongated head and mouth parts in females of *Conanthalictus* (Halictidae) allowed them to access pollen in tube-shaped flowers of *Nama*. Similar adaptations were observed in a range of Australian native species that forage on *Eremophila* (Houston 1983, 1990; Exley 1998). In addition, Sipes & Tepedino (2005) reported that the clades of *Diadasia* foraging on different host plants differed with respect to degrees of branching of the fore basitarsal hairbrush and in the length and density of branching of their scopal hairs. These morphological adaptations were linked to the respective species' ability to forage on host plants with pollen of varying sizes and shape (Sipes & Tepedino 2005). In Australia, mesoventral hair shafts with one sided branched hairs

were observed in several oligolectic species of the native Australian subgenus *Lasioglossum (Chilalictus*). This characteristic has been suggested to aid in the collection of the relatively large echinate pollen grains occurring in the genus *Wahlenbergia* (Walker 1995). These studies show that certain flowers are difficult to handle and manipulate for pollen collection and bees that exploit them require morphological adaptations (Patiny et al. 2008; Muller 1996). Therefore species that lack such morphological adaptations for pollen collection might be constrained in their pollen host selection.

The evolution of broadly polylectic behaviour

Based on the constraint hypothesis (Sedivy et al. 2008), polylectic species should be capable of evolving towards either broad polylecty or oligolecty depending on prevailing selection pressures. While polylectic species in diverse bee families have been observed to specialize on pollen hosts (Haider et al. 2014; Danforth & Ji 2001; Muller 1996; Tasei & Picart 1973), only few species exhibit broadly polylectic behaviour. Broad polylecty is predominantly observed in two bee families of Apidae and Halictidae (Vossler 2018; Dalmazzo & Vossler 2015; Cane & Sipes 2006), although some species in the family Megachilidae are also broadly polylectic (Haider et al. 2014). This might suggest that for most polylectic species, there is a selective advantage for host specialization rather than host broadening. However, it is not immediately clear how the advantages of host specialization could outweigh those of host broadening (Minckley & Roulston 2006). Alternatively, given that broadly polylectic behaviour is beneficial but rare, it seems plausible that polylectic species in most bee taxa are genetically constrained from becoming broadly polylectic. Indeed genetic constraints for pollen host selection have been demonstrated in several polylectic and broadly polylectic species, which will be discussed below.

Genetic constraints for pollen host selection in polylectic and broadly polylectic species

Using pollen loads of museum specimens, Haider et al. (2014) reported conserved floral preferences in two pairs of geographically separated, polylectic sister species: the western Mediterranean *O. emaginata* and eastern Mediterranean *O. mustelina*, and the western Palearctic *O. bicornis* and Eastern Palearctic *O. pedicornis*. Therefore, he hypothesized that, similar to oligolectic species, polylectic and broadly polylectic species too might have diet restrictions caused by genetic constraints in the ability to recognize, collect or digest pollen from particular floral resources. This hypothesis is supported by feeding experiments using two polylectic species of *O. cornuta* and *O. bicornis* (Sedivy et al. 2011), which showed that *O. bicornis* developed well on *Ranunculus* pollen but failed to develop on *Echium* while *O. cornuta* developed on *Echium* pollen but failed on *Ranunculus*. Both species developed well on *Sinapis* but none on *Tanacetum*. Additionally, while studying the pollen hosts of the bee genus *Colletes*, Müller & Kuhlmann (2008) found that polylectic species avoided pollen from the family Asteraceae in their foraging trips. Furthermore, the only three polylectic species found to utilize pollen from this family shared ancestry with species that were oligolectic on the family Asteraceae. Thus they concluded that genetic constraints were limiting other polylectic species from exploiting pollen from this family (Müller & Kuhlmann 2008). More recently, a study to determine what pollen traits are unfavourable for broadly polylectic species showed that bumble bees might lack the ability to digest pollen from the family Asteraceae (Vanderplanck et al. 2016). Together, these studies support the hypothesis that genetic constraints could exclude certain plant families from the diet of polylectic species, while broadly polylectic species have overcome these constraints.

Conclusion

The evolutionary trajectory of diet width in bees has been from oligolecty to polylecty, with numerous reversals to oligolecty. Despite the surmised evolutionary advantages of a broad diet, broadly polylectic behaviour appears to be rare across bee families. The factors constraining the evolution of broad polylecty in bees have not yet been fully explored. Based on the research presented, pollen host selection in both oligolectic and polylectic species appears to be genetically constrained. The genetic factors controlling pollen host selection could be neurological, digestive physiological or morphological. Given that broadly polylectic behaviour is taxonomically conserved, we hypothesize that the evolution from polylectic to broadly polylectic behaviour might be genetically constrained in several bee species. These genetic constraints should be deducible by comparing morphological, digestive and neurological differences associated with pollen host breadth between closely related polylectic and broadly polylectic species.

Aims and overview

In order to test the hypothesis that the evolution of broadly polylectic behaviour in bees is influenced by genetic constraints, it was necessary to select a group of species that is closely related but varies in pollen host breadth. This would reduce the number of confounding factors that could explain the differences observed in pollen host breadth between the selected species. The native Australian subgenus of *Lasioglossum* (*Chilalictus*) was selected as the study object because it has species that span the pollen host breadth spectrum from oligolecty to broad polylecty (Walker 1995). It also has well documented flower visitation records (Atlas of Living Australia website at http://www.ala.org.au. Accessed June 2018), that are necessary to establish the pollen host breadth categories of individual species. In addition, species of *L.* (*Chilalictus*) are important pollinators of both crops and native vegetation in Australia (Walker 1995), and this study would contribute to the body of knowledge on the life history of species in this subgenus.

To select species of *L.* (*Chilalictus*) that are suitable for comparing genetic differences between broadly polylectic and polylectic species, I studied the evolutionary history of pollen host breadth in this subgenus using a molecular phylogenetic approach. This approach allows one to map pollen host breadth categories of individual species onto a phylogeny and reconstruct the ancestral states of the nodes in the phylogeny using Bayesian inference (Ronquist 2004). A molecular phylogeny of *L.* (*Chilalictus*) has been previously published with 24 species (Gibbs et al. 2012), which is not sufficient to represent the range of pollen host breadth represented within this subgenus.

Therefore the aim of this study was to expand the current molecular phylogeny of *L.* (*Chilalictus*) using well identified specimens. In order to include reliably identified specimens in the phylogenetic analysis, a method needed to be developed that allowed

the amplification of multiple genes using DNA from older pinned specimens. Therefore, an amplicon sequencing method was developed in chapter 2. The sequences generated were then used to construct a molecular phylogeny of *L.* (*Chilalictus*) and analyse the evolution of broad polylecty in this subgenus. The association with geographical range size and speciation is treated in chapter 3. Finally, closely related species that differed in their pollen host breadth were examined in chapter 4 to analyse the association between broad polylecty and olfactory sensory morphology in *L.* (*Chilalictus*).

References

Adler, L. S., 2004. Size-mediated performance of a generalist herbivore feeding on mixed diets. *The Southwestern Naturalist*, 49(2), pp. 189–196.

Bernays, E.A., 1998. Evolution of feeding behaviour in insect herbivores. *Bioscience*, 48(1), pp. 35–44.

Bernays, E.A. & Bright, K.L., 1991. Dietary mixing in grasshoppers: switching induced by nutritional imbalances in foods. *Entomologia experimentalis et applicata*, 61(3), pp. 247–253.

Bernays, E.A. & Minkenberg, O., 1997. Insect herbivores: different reasons for being a generalist. *Ecology*, pp. 1157–1169.

Bernays, E.A. & Raubenheimer, D., 1991. Dietary mixing in grasshoppers: Changes in acceptability of different plant secondary compounds associated with low levels of dietary protein (Orthoptera: Acrididae). *Journal of insect behaviour*, 4(5), pp. 545–556.

Burger, H. et al., 2012. An arthropod deterrent attracts specialised bees to their host plants. *Oecologia*, 168(3), pp. 727–736.

Burkle, L. A., Marlin, J.C. & Knight, T.M., 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science*, 339(6127), pp. 1611–1615.

Cane, J.H. & Sipes, S., 2006. Characterizing floral specialization by bees: analytical methods and a revised lexicon for oligolecty. In N. M. Waser, Ollerton Jeff, ed. *Plant-pollinator interactions: from specialization to generalization*. Chicago University of Chicago Press, pp. 99–122.

Carvalho, A.T., Dötterl, S. & Schlindwein, C., 2014. An aromatic volatile attracts oligolectic bee pollinators in an interdependent bee-plant relationship. *Journal of Chemical Ecology*, 40(10), pp. 1126–1134.

Dalmazzo, M. & Vossler, F.G., 2015. Pollen host selection by a broadly polylectic halictid bee in relation to resource availability. *Arthropod-plant interactions*, 9(3), pp. 253–262.

Danforth, B.N. et al., 2006. The history of early bee diversification based on five genes plus morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 103(41), pp. 15118–15123.

Danforth, B.N. et al., 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual Review of Entomology*, 58(1), pp. 57–78.

Danforth, B.N. & Ji, S., 2001. Australian *Lasioglossum* + *Homalictus* form a monophyletic group: resolving the "Australian enigma." *Systematic Biology*, 50(2), pp. 268–283.

Dötterl, S. & Vereecken, N.J., 2010. The chemical ecology and evolution of bee–flower interactions: a review and perspectives. *Canadian journal of zoology*, 88(7), pp. 668–697.

Engel, P. & Moran, N.A., 2013. Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. *Gut microbes*, 4(1), pp. 60–65.

Engel, P., Stepanauskas, R. & Moran, N.A., 2014. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. *PLoS Genetics*, 10(9), p.e1004596.

Exley, E.M., 1998. New *Euryglossa* (*Euhesma*) bees (Hymenoptera: Colletidae: Euryglossinae) associated with the Australian plant genus *Eremophila* (Myoporaceae). *Records of the Western Australian Museum*, 18, pp. 419–437.

Feeny, P., 1991. Chemical constraints on the evolution of swallowtail butterflies. In P. W. Price et al., eds. *Plant-animal interaction; evolutionary ecology in tropical and temperate regions* . New York: John Wiley and Sons, pp. 315–340.

Futuyma, D.J. et al., 1994. Genetic variation in a phylogenetic context: Responses of two specialized leaf beetles (Coleoptera: Chrysomelidae) to host plants of their congeners. *Journal of Evolutionary Biology*, 7(2), pp. 127–146.

Futuyma, D.J., Keese, M.C. & Funk, D.J., 1995. Genetic constraints on macroevolution: the evolution of host affiliation in the leaf beetle genus *Ophraella*. *Evolution*, 49(5),p. 797.

Futuyma, D.J., Keese, M.C. & Scheffer, S.J., 1993. Genetic constraints and the phylogeny of insect-plant associations: responses of *Ophraella communa* (Coleoptera: Chrysomelidae) to host plants of its congeners. *Evolution*, 47(3),p. 888.

Futuyma, D.J. & Moreno, G., 1988. The evolution of ecological specialisation. *Annual Review of Ecology, Evolution, and Systematics*.

Gibbs, J. et al., 2012. Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Molecular Phylogenetics and Evolution*, 65(3), pp. 926–939.

González-Varo, J.P., Ortiz-Sánchez, F.J. & Vilà, M., 2016. Total bee dependence on one flower species despite available congeners of similar floral shape. *Plos*

One, 11(9),p. e0163122.

Gupta, R.K., 2014. Taxonomy and distribution of different honeybee species. In R. K. Gupta et al., eds. *Beekeeping for poverty alleviation and livelihood security*. Dordrecht: Springer Netherlands, pp. 63–103.

Haider, M. et al., 2014. Phylogeny and floral hosts of a predominantly pollen generalist group of mason bees (Megachilidae: Osmiini). *Biological journal of the Linnean Society. Linnean Society of London*, 111(1), pp. 78–91.

Houston, T., 1990. Descriptions of new paracolletine bees associated with flowers of *Eremophila* (Hymenoptera: Colletidae). *Records of the Western Australian Museum* 14(4), pp. 583–621.

Houston, T.F., 1983. An extraordinary new bee and adaptation of palpi for nectar-feeding in some Australian Colletidae and Pergidae (Hymenoptera). *Australian Journal of Entomology*, 22(3), pp. 263–270.

Janz, N., Nyblom, K. & Nylin, S., 2001. Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. *Evolution*, 55(4), pp. 783–796.

Janz, N. & Nylin, S., 2008. The oscillation hypothesis of host-plant range and speciation. In K. Tilmon, ed. *Specialization, speciation, and radiation. The evolutionary biology of herbivorous insects*. University California Press, pp. 203–215.

Klein, A.-M. et al., 2007. Importance of pollinators in changing landscapes for world crops. *Proceedings. Biological Sciences / the Royal Society*, 274(1608), pp. 303–313.

Koch, H. et al., 2013. Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Molecular Ecology*, 22(7), pp. 2028–2044.

Kwong, W.K. et al., 2014. Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, 111(31), pp. 11509–11514.

Kwong, W.K. & Moran, N.A., 2015. Evolution of host specialization in gut microbes: the bee gut as a model. *Gut microbes*, 6(3), pp. 214–220.

Larkin, L. L., Neff, J.L. & Simpson, B.B., 2008. The evolution of a pollen diet: Host choice and diet breadth of *Andrena* bees (Hymenoptera: Andrenidae). *Apidologie*, 39(1), pp. 133–145.

Litman, J.R. et al., 2011. Why do leafcutter bees cut leaves? New insights into the early evolution of bees. *Proceedings. Biological Sciences / the Royal Society*, 278(1724), pp. 3593–3600.

Martín González, A.M., Dalsgaard, B. & Olesen, J.M., 2010. Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity*, 7(1), pp. 36–43.

Martinson, V.G. et al., 2011. A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, 20(3), pp. 619–628.

Michez, D. et al., 2008. Phylogeny and host-plant evolution in Melittidae *s.l.* (Hymenoptera: Apoidea). *Apidologie*, 39(1), pp. 146–162.

Milet-Pinheiro, P. et al., 2013. The chemical basis of host-plant recognition in a specialized bee pollinator. *Journal of Chemical Ecology*, 39(11–12), pp. 1347–1360.

Minckley, R.L. & Roulston, T.H., 2006. Incidental mutualism and pollen specialization among bees. Plant-pollinator interactions. In N. M. Waser, Ollerton Jeff, ed. *From specialization to generalization*. Chicago: The University of Chicago Press, pp. 69–98.

Miura, K. & Ohsaki, N., 2004. Diet mixing and its effect on polyphagous grasshopper nymphs. *Ecological research*, 19(3), pp. 269–274.

Muller, A., 1996. Host-plant specialization in western palearctic anthidine bees (Hymenoptera: Apoidea: Megachilidae). *Ecological monographs*, 66(2), pp. 235–257.

Müller, A. & Kuhlmann, M., 2008. Pollen hosts of western palaearctic bees of the genus *Colletes* (Hymenoptera: Colletidae): the Asteraceae paradox. *Biological Journal of the Linnean Society*, 95(4), pp. 719–733.

Murphy, S.M. & Feeny, P., 2006. Chemical facilitation of a naturally occurring host shift by *Papilo machaon* butterflies (Papilionidae). *Ecological Monographs*.

Packer, L. et al., 2005. Conservation genetics of potentially endangered mutualisms: reduced levels of genetic variation in specialist versus generalist bees. *Conservation Biology*, 19(1), pp. 195–202.

Parker, F.D., 1978. Biology of the bee genus *Proteriades titus* (Hymenoptera: Megachilidae). *Journal of the Kansas Entomological Society*, 51(2), pp. 145–173

Patiny, S., Michez, D. & Danforth, B.N., 2008. Phylogenetic relationships and host-plant evolution within the basal clade of Halictidae (Hymenoptera, Apoidea). *Cladistics : the international journal of the Willi Hennig Society*, 24(3), pp.

255–269.

Praz, C.J., Müller, A. & Dorn, S., 2008a. Host recognition in a pollen-specialist bee: evidence for a genetic basis. *Apidologie*, 39(5), pp. 547–557.

Praz, C.J., Müller, A. & Dorn, S., 2008b. Specialized bees fail to develop on non-host pollen: do plants chemically protect their pollen? *Ecology*, 89(3), pp. 795–804.

Robertson, C., 1925. Heterotropic bees. *Ecology*, 6(4), pp. 412–436.

Rodrigues, D. et al., 2010. Trade-offs underlying polyphagy in a facultative ant-tended florivorous butterfly: the role of host plant quality and enemy-free space. *Oecologia*, 163(3), pp. 719–728.

Ronquist, F., 2004. Bayesian inference of character evolution. *Trends in Ecology & Evolution*, 19(9), pp. 475–481.

Scheper, J. et al., 2014. Museum specimens reveal loss of pollen host plants as key factor driving wild bee decline in The Netherlands. *Proceedings of the National Academy of Sciences of the United States of America*, 111(49), pp. 17552–17557.

Sedivy, C. et al., 2008. Patterns of host-plant choice in bees of the genus *Chelostoma*: the constraint hypothesis of host-range evolution in bees. *Evolution*, 62(10), pp. 2487–2507.

Sedivy, C., Müller, A. & Dorn, S., 2011. Closely related pollen generalist bees differ in their ability to develop on the same pollen diet: evidence for physiological adaptations to digest pollen. *Functional ecology*, 25(3), pp. 718–725.

Singer, M.S., 2001. Determinants of polyphagy by a woolly bear caterpillar: a test of the physiological efficiency hypothesis. *Oikos*, 93(2), pp. 194–204.

Singer, M.S. et al., 2004. Disentangling food quality from resistance against parasitoids: diet choice by a generalist caterpillar. *The American Naturalist*, 164(3), pp. 423–429.

Singer, M.S., 2008. Evolutionary ecology of polyphagy. In KJ Tilmon, ed. *Specialization, speciation, and radiation: The evolutionary biology of herbivorous insects*. University of California Press, pp. 29–42.

Sipes, S.D. & Tepedino, V.J., 2005. Pollen-host specificity and evolutionary patterns of host switching in a clade of specialist bees (Apoidea: *Diadasia*). *Biological Journal of the Linnean Society*, 86(4), pp. 487–505.

Slatyer, R.A., Hirst, M. & Sexton, J.P., 2013. Niche breadth predicts geographical range size: a general ecological pattern. *Ecology Letters*, 16(8), pp. 1104–1114.

Strickler, K., 1979. Specialization and foraging efficiency of solitary bees. *Ecology*, 60(5), pp. 998–1009.

Suárez-Cervera, M. et al., 1994. An ultrastructural study of pollen grains consumed by larvaeof *Osmia* bees (Hymenoptera, Megachilidae). *Grana*, 33(4–5), pp. 191–204.

Tasei, J.N. & Picart, M., 1973. Le comportement de nidification chez *Osmia* (*Osmia*) *cornuta* Latr. et *Osmia* (*Osmia*) *rufa* L. (Hymenoptera Megachilidae). *Apidologie*, 4, p.195 –225 .

Thorp, R.W., 1979. Structural, behavioural, and physiological adaptations of bees (Apoidea) for collecting pollen. *Annals of the Missouri Botanical Garden*, 66(4),p. 788.

Thorp, R.W., 2000. The collection of pollen by bees. *Plant Systematics and Evolution*,

222(1–4), pp. 211–223.

Vanderplanck, M. et al., 2016. Is non-host pollen suitable for generalist bumblebees? *Insect science*, 25(2), pp. 259–272.

Vanderplanck, M. et al., 2017. The importance of pollen chemistry in evolutionary host shifts of bees. *Scientific reports*, 7, p.43058.

Vossler, F.G., 2018. Are stingless bees a broadly polylectic group? an empirical study of the adjustments required for an improved assessment of pollen diet in bees. In P. Vit, S. R. M. Pedro, & D. W. Roubik, eds. *Pot-Pollen in Stingless Bee Melittology*. Cham: Springer International Publishing, pp. 17–28.

Walker, K.L., 1995. Revision of the Australian native bee subgenus *Lasioglossum* (*Chilalictus*). *Memoirs of Museum Victoria*, 55(1 & 2), pp. 1–423.

Waser, N.M. et al., 1996. Generalization in pollination systems and why it matters. *Ecology*, 77(4),p. 1043–1060.

Wcislo, W.T. & Cane, J.H., 1996. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. *Annual Review of Entomology*, 41, pp. 257–286.

Wright, G.A. et al., 2010. Parallel reinforcement pathways for conditioned food aversions in the honeybee. *Current Biology*, 20(24), pp. 2234–2240.

Statement of Authorship

Title of Paper	Multilocus barcoding of pinned insect specimens using MiSeq
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	NA

Principal Author

Name of Principal Author (Candidate)	Trace Akankunda
Contribution to the Paper	Developed the concept, conducted sequencing of specimens, conducted 50% of data analysis, wrote and revised the manuscript
Overall percentage (%)	80%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 9/04/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Katja Hogendoorn
Contribution to the Paper	Contributed to concept development, identified specimens, provided guidance on laboratory experiments, revised the manuscript
Signature	Date 1/4/2019

Name of Co-Author	Carlos Rodriguez Lopez
Contribution to the Paper	Contributed to concept development, provided guidance on laboratory experiments, revised the manuscript
Signature	Date 3/4/2019

Name of Co-Author	Remko Leijs		
Contribution to the Paper	Contributed to concept development, provided guidance on laboratory experiments, revised the manuscript		
Signature		Date	4/4/2019

Name of Co-Author	Thu-Hien To		
Contribution to the Paper	Conducted 50% of the sequence data analysis, revised the manuscript		
Signature		Date	1 April 2019

Chapter 2: A method to generate multi-locus barcodes of pinned insect specimens using MiSeq

Trace Akankunda[1], Hien To[4], Carlos Rodriguez Lopez[1,2], Remko Leijs[3], Katja Hogendoorn[1,5]

katja.hogendoorn@adelaide.edu.au

[1]The University of Adelaide, School of Agriculture, Food and Wine Adelaide SA 5005

[2]Environmental Epigenetics and Genetics Group; Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington, KY 40546

[3]South Australian Museum, North Terrace Adelaide SA 5000

[4]The Bioinformatics Hub, The University of Adelaide, SA 5000, Australia

[5] Corresponding author

Abstract

For molecular insect identification, amplicon sequencing methods are recommended because they offer a cost effective approach for targeting small sets of informative genes from multiple samples. High-throughput multilocus amplicon sequencing has been achieved using the MiSeq sequencing platform. However, the MiSeq generates short gene fragments of less than 500 bp, which then have to be overlapped using bioinformatics to achieve longer sequence lengths. This increases the risk of generating chimeric sequences or leads to the formation of incomplete loci. Here, we propose a modified nested amplicon sequencing method for targeting multiple loci from pinned insect specimens using the MiSeq Illumina platform. The modification exists in using a three-step nested PCR approach targeting near full length loci in the initial PCR and subsequently amplifying short fragments between 300 and 350 bp for next generation sequencing using Illumina chemistry. Using this method, we generated 407 high quality sequences of three loci from 86% of all the specimens sequenced. Out of 103 pinned bee specimens of replicated species, 71% passed the 95% sequence similarity threshold between species replicates. This method worked best for pinned specimens aged between 0 - 5 years, with a limit of 10 years for pinned and 14 years for ethanol preserved specimens. Hence, our method overcomes some of the challenges of amplicon sequencing using short read Next Generation Sequencing and improves possibilities to create high quality multilocus barcodes from insect collections.

Key words: Insect collections, multilocus barcoding, MiSeq

Introduction

Accurate species identification and classification underpins all biological research. Despite considerable efforts, it is estimated that 86% terrestrial and 91% marine species are still undescribed (Mora, Tittensor, Adl, Simpson, & Worm, 2011). This is because until recently, the tasks of species description and classification have been shouldered by morphological taxonomists (Dunn, 2003), who are limited in numbers. This task is greater for highly diverse taxonomic groups such as insects, with 80% of the estimated number of species globally, yet to be discovered (Stork, 2018). Molecular barcoding has been used to fast track species description and classification with approximately 20% of known insect species having barcodes of the cytochrome c oxidase subunit 1 (*CO1 5'*) marker in the Barcode of Life Database (Ratnasingham & Hebert, 2007). In addition, advancements in technology have enabled the sequencing of type specimens which has enhanced accurate species identification for users of sequence databases globally (Prosser, deWaard, Miller, & Hebert, 2016; Hausmann et al., 2016; Hebert et al., 2013).

Mulitilocus barcoding can enhance accurate species identification especially in species groups where the CO1 marker (658 bp of 5' end of CO1) is problematic (Cruaud, Rasplus, Rodriguez, & Cruaud, 2017; Rach et al., 2017; Liu et al., 2017; Klopfstein, Kropf, & Baur, 2016; Paknia, Bergmann, & Hadrys, 2015). For example, some odonate and dipteran taxa lack a sufficient barcoding gap in the standard CO1 barcode sequence (Koroiva & Kvist, 2018; Meier, Shiyang, Vaidya, & Ng, 2006), the presence nuclear-mitochondrial DNA pseudo genes (Numts) in the CO1 sequences of several taxa in the order Orthoptera (Moulton, Song, & Whiting, 2010; Song, Buhay, Whiting, & Crandall, 2008) and some species of blue banded bees (*Amegilla*; Leijs, Batley, &

Hogendoorn, 2017) and the genus *Sitobion* (Hemiptera; Sunnucks & Hales, 1996). Distortion of mitochondrial inheritance patterns by bacterial endosymbionts was found in some species of the wasp genus *Diplazon* (Hymenoptera; Klopfstein, Kropf, & Baur, 2016). More importantly, multilocus barcoding can allow us to explore the relationships and evolutionary history of species through phylogenetic construction (Sonet et al., 2018; Mallo & Posada, 2016). This significantly increases the usefulness of public sequence databases for scientific research.

It is recommended that for purposes of multilocus barcoding, a conservative approach of sequencing genes that are in common use for species delineation and phylogenetics, be taken (Cruaud et al., 2017). This should enable compatibility between new sequence datasets and existing sequence databases. For this purpose, amplicon sequencing is ideal because it is a cost effective method for targeting a small number of informative genes from multiple samples (Burrell, Disotell, & Bergey, 2015). Advancements in sequencing technology have enabled high-throughput amplicon sequencing of insect specimens on a range of sequencing platforms, e.g. the 454 pyrosequencing technology (Shokralla et al., 2014; Bybee et al., 2011), the Illumina MiSeq platform (Sonet et al., 2018; Cruaud et al., 2017) and the single molecule real time (SMRT) sequencing method on a Sequel platform (Hebert et al., 2018). Owing to its comparatively high sequencing output (Shokralla et al., 2014), the MiSeq platform is ideal for multilocus barcoding of insect collections because it provides flexibility in the number of target loci and samples that can be processed in a single sequencing run (Sonet et al., 2018; Cruaud, 2017). However, owing to its sequence length limitation, currently 500 bp via paired end sequencing with a 600 cycle kit (Shokralla et al., 2014), previous attempts at multilocus barcoding using the MiSeq have targeted short sequence loci of less than

600 bp (Sonet et al., 2018). Loci with complete longer sequence lengths can be generated by amplifying and sequencing multiple short fragments, and then overlapping them using bioinformatics (Sonet et al., 2018; Shokralla et al., 2015). While this approach extends the sequence length achievable with the MiSeq platform, it is prone to creating chimeric sequences when short fragments of pseudo genes, paralogous genes or nuclear mitochondrial genes (numts) are simultaneously sequenced with target genes and merged using bioinformatics (Hebert et al., 2018; Song et al., 2008). This could mislead species identification and phylogenetic analysis.

When sequencing mitochondrial genes such as CO1 using short read next generation sequencing platforms, there is a considerable risk of generating chimeras due to the prevalence of NUMTS that are readily amplified using PCR (Hebert et al., 2018; Song et al., 2008). Additionally, several genes currently used for phylogenetic studies of insects including eukaryotic translation elongation factor 1 alpha (*EF-1* α), wingless (*wg*) and long-wavelength rhodopsin have multiple copies (Almeida & Danforth, 2009; Spaethe & Briscoe, 2004; Danforth & Ji, 2001; Gibbs, Brady, Kanda, & Danforth, 2012; Mardulyn & Cameron, 1999), that could be co-amplified, sequenced and erroneously combined to form chimeras. The formation of such chimeras can be avoided by directly amplifying near full-length amplicons for sequencing, using locus specific primers. The amplicons generated could then be sequenced using methods that are capable of directly sequencing long length amplicons (over 600 bp) such as Sanger (Shokralla et al., 2014) or SMRT sequencing (Hebert et al., 2018). However, these methods have only been used for sequencing long length amplicons from freshly collected or well preserved samples and not samples with degraded DNA, which form the bulk of insect collections in museums. While the MiSeq platform has previously

been used to sequence loci with long sequence lengths, from samples with degraded DNA, the sequences generated were incomplete (Cruaud et al., 2017).

Therefore, the aim of this study was to increase the sequence length of loci achievable on the MiSeq platform using a modified nested PCR approach (Mitchell, 2015), in order to develop a high-throughput amplicon sequencing method capable of generating high quality multiple near full-length gene sequences from pinned insect specimens. The modification exists in using a three-step nested PCR approach targeting near full-length loci in the initial PCR, which ensures the specificity of a target amplicon, and then subsequently amplifying short fragments of the target amplicon for next generation sequencing using Illumina chemistry. So far, multilocus barcoding approaches on the MiSeq have been trialled with fresh or ethanol preserved samples (Sonet et al., 2018; Cruaud et al., 2017) and have not been tested on pinned specimens. In this study, we use both ethanol and pinned specimens to test the specimen age limits of the sequencing method developed. We targeted some of the most commonly used loci for insect species delineation and phylogenetics including; eukaryotic translation elongation factor 1 alpha (*EF-1* α), wingless (*wg*) and cytochrome c oxidase subunit 1 (*CO1*) (Gibbs, Brady, Kanda, & Danforth, 2012; Almeida & Danforth, 2009; Mardulyn & Cameron, 1999; Danforth & Ji, 1998). By choosing to amplify the near full-length loci in the initial PCR, we tried to overcome some of the challenges that could be faced in multilocus barcoding projects. Examples of these challenges include introns in nuclear genes, which are variable and difficult to amplify (e.g. *EF-1* α; Gibbs et al., 2012), presence of multiple copies of genes (Gibbs et al., 2012; Almeida & Danforth, 2009; Danforth & Ji, 2001; Mardulyn & Cameron, 1999) and nuclear paralogues (numts), which can mislead phylogenetic analysis (Leijs et al., 2017).

In this study, we used insect collections of *Lasioglossum* (*Chilalictus*: Hymenoptera: Halictidae), of which most species have been described (Walker, 1995), while the group contains species that are quite difficult to identify using morphological characters (Walker pers. com). In addition, representatives of the taxon are relatively commonly found on introduced plants and crops (Cunningham, Schellhorn, Marcora, & Batley, 2013; Arthur, Li, Henry, & Cunningham, 2010; Michener, 1965). Hence, molecular markers developed from reliably identified specimens could assist in species identification of crop pollinators, currently a rapidly emerging area of interest and importance, by researchers that lack taxonomic experience.

Materials and Methods

Sample Collection

The specimens of the *Lasioglossum* subgenus *Chilalictus* used in this study were sampled from the pinned and ethanol collections curated by Hogendoorn and Leijs and the Museums Victoria collection. The identities of the specimens were checked by K. Walker, an expert in Australian *Lasioglossum* (Walker, 1995). Voucher specimens were given a DNA extraction number (Table S2.1). Voucher specimens from Hogendoorn and Leijs have been donated to the SA Museum while vouchers of samples collected from K. Walker are kept in Museum Victoria collection. Two to three legs were sampled per specimen, using triplicate specimens for each species when available, resulting in a total of 138 specimens representing 52 species. Of these, 18 specimens had been preserved in ethanol while 120 specimens had been pinned. The oldest ethanol and pinned specimens were 14 and 49 years respectively (Table S2.1). Of the pinned specimens, 110 were between 0 – 5 years old and 103 of these had at least two replicates (Table S2.1), to maximise the probability of generating barcodes. A further range of specimens aged over 5 years were included for both ethanol preserved and pinned specimens to determine the specimen age limit for the sequencing method being developed. All molecular work prior to the sequencing run was conducted at the South Australian Regional Facility for Molecular Ecology and Evolution (SARFMEE).

DNA extraction

Leg tissues from ethanol samples were dried for 10 seconds at 50 °C and all tissue samples disrupted using a Tissue Lyser II (Qiagen) following the manufacturer's instructions. DNA was extracted using an in-house Gentra salt precipitation protocol (Qiagen) modified for maximum DNA yield by using overnight incubation steps

(Appendix 2.1). DNA concentrations were measured using Promega™ QuantiFluor™ dsDNA System.

Selection of target gene regions and primer design for PCR

Published sequences of three gene regions commonly used for molecular phylogenetics in bees namely; CO1 (two gene regions), EF-1α, and wg (Gibbs et al., 2012) were downloaded from GenBank and aligned using MAFFT v 1.3.3 (Katoh, Misawa, Kuma, & Miyata, 2002) in Geneious v 8.1.3 (Kearse et al., 2012). Published primers for the CO1 5' (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994) and CO1 3' gene regions (Simon et al., 1994) and primers for single copy gene regions of EF-1α - F2 copy (Danforth, Brady, Sipes, & Pearson, 2004) and wg (Almeida & Danforth, 2009; Brower & DeSalle, 1998) were checked *in silico* for degeneracy against *Lasioglossum* (*Chilalictus*) spp. alignments and their sequences revised appropriately. The degenerate primers were quality-checked using OligoAnalyzer 3 following recommended primer properties for successful PCR amplification (PREMIER Biosoft, 2018). They were then tested in the laboratory for their ability to amplify near-full length genes, following the 1st PCR step in the three-step-PCR amplification protocol detailed below. Finally, the resulting PCR products were analysed by agarose gel electrophoresis (1.5 % agarose in 1x Tris Borate EDTA buffer at 100 V/cm) and classical Sanger sequencing (Macrogen Inc, Korea) to ensure the identity of the amplified gene regions. Primer sequences of near full-length genes that were used in this study are shown in Table 2.1.

Nested primers were designed in Geneious to amplify 300 – 350 bp overlapping fragments from the CO1 5', CO1 3', EF-1α (F2 copy) and wg gene regions. These were tested for degeneracy and quality checked as described above. Primers with

recommended properties and sufficient overlap between gene fragments were selected (Table 2.1) and

Table 2.1. Degenerate primers for *Lasioglossum* used to amplify CO1, EF-1α and WNT with references for those modified from published primers. Modifications include; adopting degenerate bases and adding M13 tails and Illumina sequencing adapters for next generation sequencing. For each gene region in the table, the first and last primers amplify near-full length loci and the middle primers are nested within each gene sequence to amplify 300-350 bp gene fragments.

Primer Name	Sequence (3' & 5')	Published Primer Name. Reference
mCO15f	m-GGTCAACAAATCATAAAGATATTGG	LCO1490. Folmer et al., 1994.
iCO15_1f	i-GGTCAACAAATCATAAAGATATTGG	LCO1490. Folmer et al., 1994.
iCO15_1r	i-GATARDGGDGGRTAWAYTGTTC	AMbc5r1m. Mitchell, A. 2015
TAiCO15_2f	i-GAYWWTAWMWWCWGGWTCWGG	This study
iCO15_2r	i-TAWACTTCWGGRTGDCCAAAAAATCA	HC02198. Folmer et al., 1994
mCO15r	m-TAWACTTCWGGRTGDCCAAAAAATCA	HC02198. Folmer et al., 1994
mCO13f	m-CAACAYYTATTTTGATTTTTTGG	C1-J-2183 Simon, C. et al., 1994
iCO13_1f	i-CAACAYYTATTTTGATTTTTTGG	C1-J-2183 Simon, C. et al., 1994
TAiCO13_1r	i-ATRATWCCWGTWADNCCHCC	This study
TAiCO13_2f	i-TGAYTWGCWACWTAYTGTGG	This study
TAiCO13_2r	i-GGRTARTCWGARTAWCGWCGWGG	This study
iCO13_3f	i-GTWAAYWTAACHTTYTTYCCHCAACA	CO1 9 UEA9. Lunt, D.H. et al., 1996
iCO13_3r	i-TCCAATCGACTAATCTGCCATATTA	TL2-N-3014. Simon et al., 1994.
mCO13r	m-TCCAATCGACTAATCTGCCATATTA	TL2-N-3014. Simon et al., 1994.
mWNTf	m-TGCACNGTSAAGACCTGYTGGATGAG	beewgFor. Almeida, E & Danforth, B. N., 2009.
iWNT_1f	i-TGCACNGTSAAGACCTGYTGGATGAG	beewgFor. Almeida, E & Danforth, B. N., 2009.
TAiWNT_1r	i-TCGTTRCACTGTCTRCCGTG	This study
TAiWNT_2f	i-GCRATAGCCAGCAAYTCGGC	This study
iWNT_2r	i-ACTICGCARCACCARTGGAATGTRCA	ModLEPWG2. Brower, A &DeSalle, R. 1998.
mWNTr	m-ACTICGCARCACCARTGGAATGTRCA	ModLEPWG2. Brower, A &DeSalle, R. 1998.

mEFf	m-GGGYAAAGGWTCCTTCAARTATGC	HaF2for1. Danforth, B.N et al., 2004.
iEF_1f	i-GGGYAAAGGWTCCTTCAARTATGC	HaF2for1. Danforth, B.N et al., 2004.
TAiEF_1r	i-CTTCRAATCGRGCTTCGGAG	This study
TAiEF_2f	i-GGACAAACYCGTGAGCATGC	This study
TAiEF_2r	i-CCTGAAGRGGAAGACGGAGAGCC	This study
TAiEF_3f	i-CATTCTTCCACCTTCGAGAC	This study
TAiEF_3r	i-AGTTTCGACACGACCAACGGG	This study
TAiEF_4f	i-TGTTGATGGCCAGAGGTGTC	This study
iEF_4r	i-AATCAGCAGCACCTTTAGGTGG	F2-Rev1. Danforth, B.N et al., 2004
mEFr	m-AATCAGCAGCACCTTTAGGTGG	F2-Rev1. Danforth, B.N et al., 2004

Naming convention: 'TA' denotes author; CO15 & CO13=Cytochrome Oxidase 1 5' & 3', EF=Elongation Factor 1 alpha 2nd copy, WNT=Wingless gene region (Wnt-1). Numbers (1-4) are gene fragments for each locus numbered in the 5' to 3' direction, f=Forward primer, r=Reverse primer, m=M13tail, i=Illumina sequencing primer. For forward primers (f); m=GTAAAACGACGGCCAGT, and i=TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, for reverse primers (r); m=CAGGAAACAGCTATGAC and i=GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG.

their PCR products were Sanger sequenced to ensure that they amplified the expected gene fragments. Finally, primer pairs were adapted to contain M13 tails (near full-length gene primers) or Illumina sequencing tags (nested primers) following Illumina guidelines (Bell, 2011). For two specimens of *L. lanarium*, we sequenced all of the gene fragments using classical Sanger sequencing and later compared them with their MiSeq sequences to test for consistency between the two methods.

Library preparation for paired end amplicon sequencing

To allow the use of NGS platforms, modifications to the nested PCR method (Mitchell, 2015), were implemented by combining a nested PCR protocol with the tailed amplicon sequencing approach (Bell, 2011) and the Illumina 16S Metagenomic Sequencing Library Preparation protocol (Illumina, 2013).

Three-step PCR amplification

A three-step-PCR approach was used to prepare the amplicon-sequencing library. The 1st PCR targeted near full-length loci followed by a 2nd PCR that used nested primers

(containing a tail complementary to the Illumina sequencing primers) to amplify 300 – 350 bp fragments and a final indexing PCR for multiplexing the samples (Fig. S2.1 and Fig. S2.2). Sterile techniques recommended for handling museum material were followed to avoid environmental contamination and three negative controls were included for each gene region and maintained throughout the library to identify sources of environmental contamination (Mitchell, 2015). DNA samples were transferred to 96-well-plates arranged by concentration to allow pipetting different DNA volumes for PCR using a multichannel pipette. The annealing temperature optimised for each gene fragment was; CO1 5': 48 °C, CO1 3': 46 °C, EF-1α: 54 °C and wg: 58 °C.

The 1st PCR was conducted in 15 µl reaction volumes containing 7.5 µL. KAPA HiFi HotStart Ready-mix, 3 µl sterile water, 1.2 µl of 5 µM forward and reverse primers (Table 2.1), 1 ng of template DNA. The PCR conditions used were; 95 °C/10 min, 35 cycles of 95 °C/40 s, gene specific Annealing temperature (i.e. CO1 5' = 48 °C, CO1 3' = 46 °C, EF-1α =54 °C and wg = 58°C see above)/ 40 s, 72 °C/60 s, and a final 6 min step at 72 °C. PCR products were then diluted 1:10 using Nanopure water.

The 2nd PCR reactions were prepared using Robotic Liquid Handlers – epMotion 5075 to reduce pipetting variability and cross-contamination. The PCR reactions were conducted in 15µl reaction volumes containing 8.54 µl sterile water, 3 µl 5x Immolase buffer, 0.06 µl (5 u/µl) Immolase enzyme, 1.2 µl of 5 µM forward and reverse primers (Table 2.1), 1 µl of 1:10 dilution of the 1st PCR reaction volume as template. The PCR conditions used were; 95 °C/10 min, 25 cycles of 95 °C/40 s, fragment specific Annealing temperature (see above)/ 40 s, 72 °C/60 s and a final 6 min step at 72 °C.

Optional Ampure clean up and library quality sampling

The PCR reactions from the 2nd PCR were cleaned with Agencourt Ampure XP system (Beckman Coulter) following the manufacturer's instructions using the Robotic Liquid Handlers – epMotion 5075. In the next step, the library was sampled for quality assessment by analysing the number of successful PCR amplification, specificity of the primers used, and quantity of the amplicons generated using the Experion automated electrophoresis system (Bio-Rad).

Dual Indexing PCR

The 3rd PCR was conducted in 12.5 µl reaction volumes containing 6.45 µl sterile water, 2.5 µl 5 x Immolase buffer, 2 µl (5 µM) premix of custom designed Illumina i7 and i5 primers (Table S2.2), 0.05 µl (5 u/µl) Immolase enzyme and 1 µl of the 2nd PCR reaction volume. The PCR conditions used were; 95 °C/10 min, 10 cycles of 95 °C/40 s, 55°C / 40 s, 72 °C/60 s, and a final 6 min step at 72°C. PCR success was checked by running randomly selected indexed samples alongside non-indexed PCR 2 samples on 2 % agarose gels. PCR products were then pooled by gene fragment into 11 pools. Individual pools were checked for fragment size and concentration using the Agilent 2200 Tape Station System and then pooled equimolarly into a single library. The final library was diluted to 20nM and submitted to the Australian Genomics Research Facility (AGRF) Adelaide node for sequencing using a MiSeq 600 bp pair-end Reagent Kit v3 (Illumina).

Analysis of sequencing data

The raw sequence data were uploaded onto the Phoenix high performance computing facility of the University of Adelaide. Paired end fastq files were merged using BBMerge (BBMap/35.92-GCC-5.3.0-binutils-2.25-Java-1.8.0_71) with the strictness

settings adjusted to Xloose following the authors' recommendations for amplicon sequence data of long reads ("BBMerge Guide - DOE Joint Genome Institute," 2017). The fastQC results were visualised with the ngsReports tool (Ward, To, & Pederson, 2017). Samples that passed the quality check were de-multiplexed by gene fragment into separate fastq files using the gene-fragment primers as inline barcodes. A fastQC check was performed on individual gene fragments by pooling all samples together based on their gene fragment. De-multiplexing was achieved by trimming the forward primer using cutadapt (version 1.9.1) and only retaining the trimmed reads. The reverse primers were then trimmed from de-multiplexed fastq files and retaining the trimmed reads only to filter out possible sample to sample contaminants during the library preparation or sequencing. A custom bioinformatics script for the raw data analysis pipeline can be found in Appendix 2.2 with annotations for each step.

Consensus sequence assembly

For assembling gene fragment sequences into a consensus sequence we combined two pipelines; the Geneious mapper built into a desktop version of the Geneious software program (v 8.1.3) and a custom bioinformatics pipeline trained on the former (see supplementary material). The custom bioinformatics pipeline, written with a combination of R v 3.5.2 (R Core Team 2018) and bash scripts, was developed to overcome the memory and storage space limitations of the desktop Geneious version and exploit the computation resources available on the Phoenix high performance computing facility for high-throughput analysis. The chosen method of consensus sequence assembly was by reference sequence mapping. This approach was executed in two stages, first filtering out possible environmental contaminants, using a minimum sequence similarity threshold of 84% to a reference sequence of *Lasioglossum*

(*Chilalictus*) previously generated by Sanger sequencing (Fig. S2.3) and then removing sequences that were 250 bp shorter than the expected minimum fragment length.

As the most represented sequence (major cluster) usually matches the target sequence (Cruaud et al., 2017), we combined sequences of major clusters from each gene fragment to form a specimen specific reference sequence for each locus. Sequence reads from each sample were then remapped to the sample specific reference sequence with a minimum threshold of 98% sequence similarity to generate a consensus sequence for each sample and gene region. The minimum sequence read coverage for generating a consensus sequence was set to 3. Species names were added to the specimen identifiers of the consensus sequences assembled in order to identify replicates. Consensus sequences of the negative controls were assembled *de novo* in Geneious (v 8.1.3) to determine all the possible sources of library contamination if any. A custom bioinformatics script for the sequence assembly pipeline can be found in Appendix 2.2 with comments to explain each step.

Quality assessment of assembled consensus sequences

The consensus sequences were uploaded into Geneious (v 8.1.3) and sorted by length to remove sequences less than half of the full length of the targeted gene region (500 bp for EF-1α and CO1 3' gene regions, 300 bp for CO1 5' and 200 bp for wg gene regions). Sequences with appropriate length were blasted against the NCBI GenBank database accessed through Geneious and only sequences with a minimum of two top congeneric matches to *Lasioglossum* were passed (Meier et al., 2006). Allowance was made for GenBank matches to the family Halictidae for CO1 5' consensus sequences due to the scarcity of CO1 5' barcode sequences of *Lasioglossum* (*Chilalictus*). Sequences that passed the above-described filtering procedure were aligned with

MAFFT v 1.3.3 (Katoh et al., 2002) to detect any obvious sequence errors (mostly base inserts occurring in a single sequence and not in its replicates) that could be edited manually. Sequence alignments were also converted to amino acid sequences to check for premature stop codons as a sign of insertions or deletions introduced during library preparation and sequencing (Cruaud et al., 2017).

Maximum likelihood (ML) trees were generated for individual gene regions using RAxML v 8.2.10 (Stamatakis, 2014) with a GTRGAMMAI model and no bootstrap replicates. The gene trees were compared to identify inconsistently clustered replicates and mislabelled vouchers. Only samples with well clustered replicates or with a read coverage above the negative controls were retained (Sproul & Maddison, 2017). Additionally, the percentage pairwise sequence similarity and number of identical sites between consensus sequences of replicates were calculated in Geneious and only samples with a 95% score for both parameters were passed. An attempt was made to recover samples with lower replicate pairwise similarity scores by modifying the consensus sequence assembly script to retrieve sequence reads shared with their conspecific replicates once the replicates were confirmed to be accurate (see Appendix 2.2 for modified script).

We compared a CO1 phylogeny (Fig. S2.4) to a multilocus phylogeny constructed with concatenated gene sequences in order to check for congruence and to determine the added phylogenetic value of the extra loci sequenced. For this analysis, we used species with consensus sequences across all three loci to allow comparison between phylogenetic trees. The sequences were aligned using MAFFT v 1.3.3 (Katoh et al., 2002) in Geneious v 8.1.3 (Kearse et al., 2012). Sequences were concatenated manually and separated into ten partitions by gene, codon and intron, and appropriate models of

evolution for each partition were assigned using Partition finder v 2.1.1 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017). Phylogenetic trees were generated using MrBayes v 3.2.6 (Ronquist et al., 2012), hosted on the Cipress getway portal (Miller, Pfeiffer, & Schwartz, 2010). In addition, the sequences of the two *L. lanarium* specimens that had been sequenced using the classical Sanger sequencing method as well as MiSeq for the three different loci were aligned to determine their sequence similarity as a measure of consistency in the results generated by the two methods (Table S2.4).

Statistical analysis

Using SPSS statistical software programme (version 25) we estimated the correlation between the age of the specimens and (a) the average number of reads passing the filtering pipeline (b) the DNA concentration. In the latter analysis, we used a partial correlation to control for the effect of the amount of tissue sampled, using the average body length of the species as proxy for the amount of tissue sampled (Walker, 1995).

Results

For the pinned specimens, there was a significant negative partial correlation between specimen age and log transformed DNA concentration when controlling for average species' size ($r = -0.41$; $p<0.01$; treating the concentration of the specimen aged 33 as an outlier because its DNA was heavily contaminated). For specimens preserved in ethanol, the correlation between age and DNA concentration was not significant when controlling for average species' size ($r = 0.07$; $p= 0.78$; Fig. 2.1).

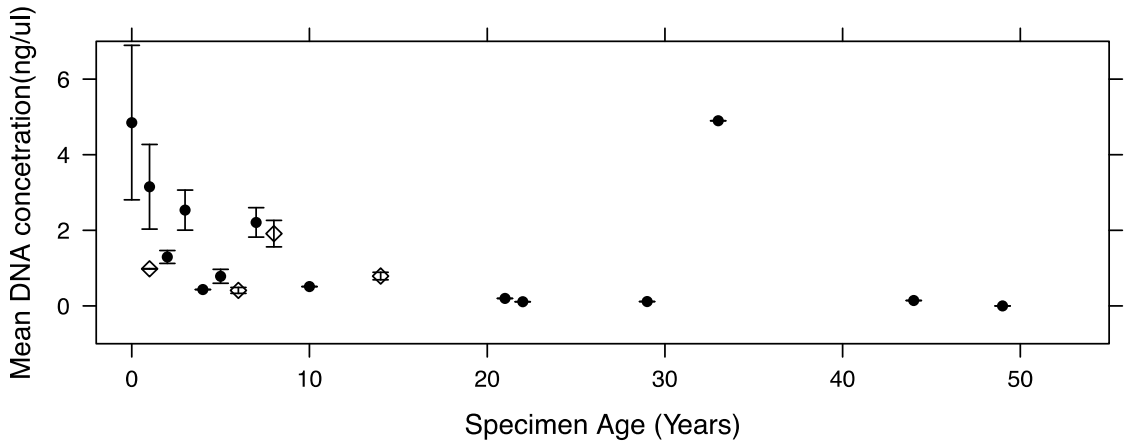

Figure 2.1. The average DNA concentration (± s. e.) of samples grouped by specimen age and method of storage. Closed circles: pinned specimens; Rhombus: ethanol specimens.

For the pinned specimens, there was a significant negative correlation between specimen age and the number of reads passing the filtering pipeline ($r = -0.53$; $p < 0.01$) while for ethanol specimens there was no significant correlation between specimen age and the number of reads passing the filtering pipeline ($r = 0.25$; $p = 0.31$; Fig. 2.2).

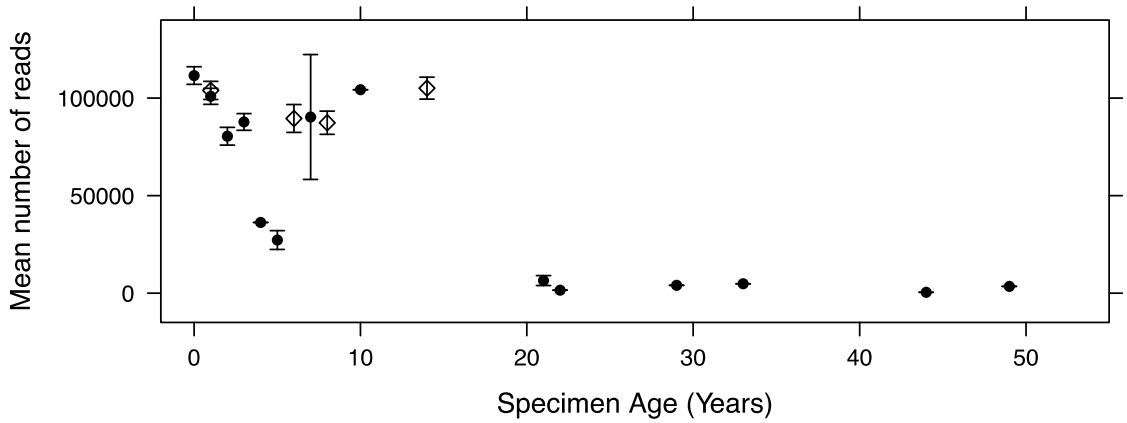

Figure 2.2. The average number of reads (± s. e.) passing the filtering pipeline with samples grouped by specimen age and method of storage. Closed circles: pinned specimens; Rhombus: ethanol specimens.

As an approximate indicator of sequence accuracy, we used the proportion of sequences that made up the most represented sequence ('the major cluster' after the filtering procedure) per sample and gene fragment (Cruaud et al., 2017). If the major cluster makes up a small proportion of the sequences (below 0.5), this implies that there is a high diversity of sequence reads in a sample, which is indicative of either PCR error, contamination across samples during library preparation, or multiple copies of the target sequence (Cruaud et al., 2017). To visually compare the sequence quality between gene fragments, a frequency distribution of the proportion of the major cluster was plotted for each gene fragment (Fig. 2.3). The wg gene fragments had the highest number of specimens with a proportion of the major cluster above 0.5 (96%) followed by EF-1α; 83-92%, CO1 3'; 82-91% and CO1 5'; 78-88% (Fig. 2.3).

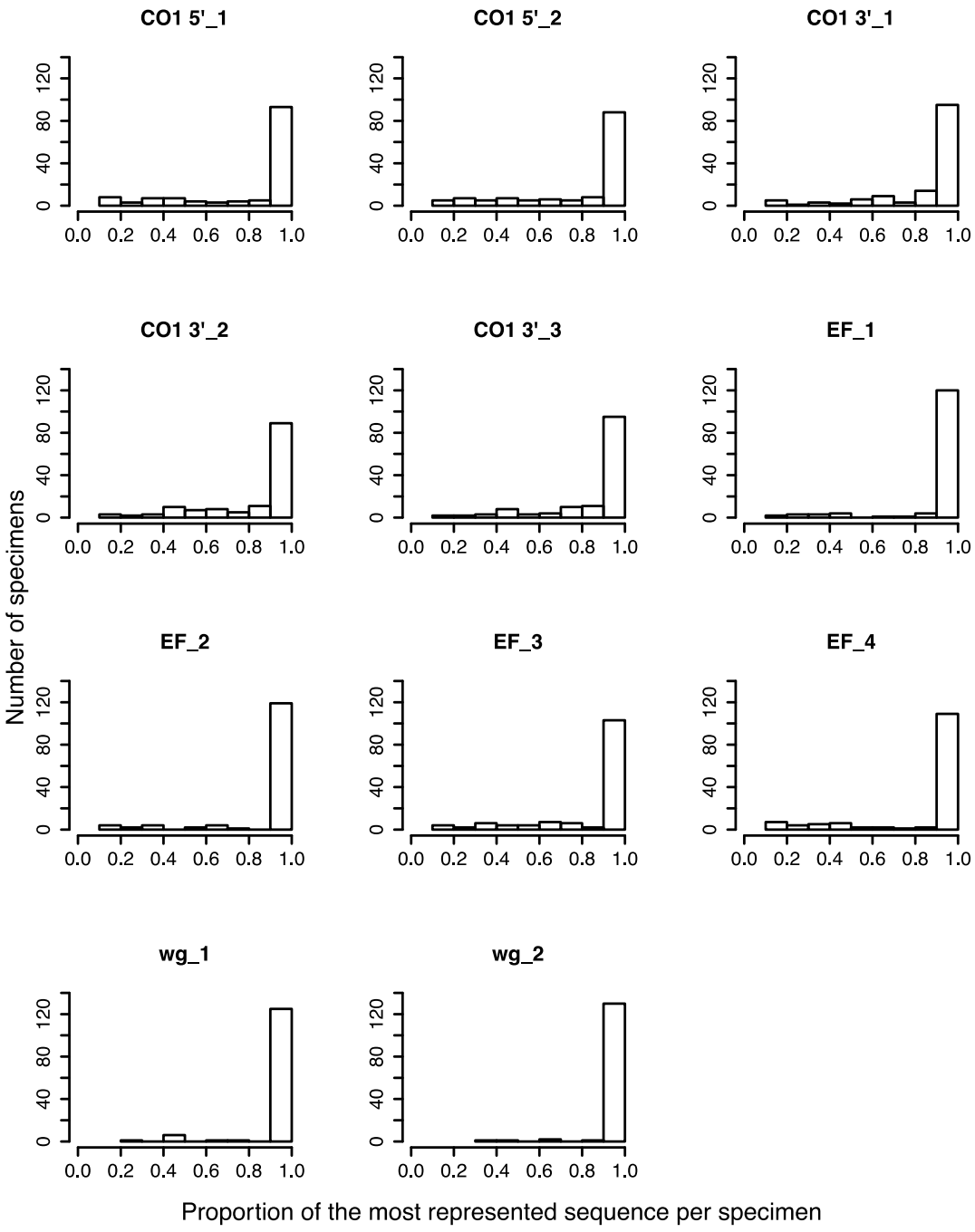
CO1 5'_1
CO1 5'_2
CO1 3'_1
CO1 3'_2
CO1 3'_3
EF_1
EF_2
EF_3
EF_4
wg_1
wg_2
Number of specimens
Proportion of the most represented sequence per specimen

Figure 2.3. Frequency distributions of the proportion of the most represented sequence per specimen and gene fragment for 11 gene fragments across 138 specimens. Naming convention: CO1 5' & CO1 3' = cytochrome c oxidase subunit 1 5' & 3', EF = eukaryotic translation elongation factor 1 alpha, wg = wingless gene region. Numbers (1-4) are gene fragments for each locus numbered in the 5' to 3' direction.

From 86% of the specimens sequenced in this study, 407 barcode-compliant sequences from three loci were generated, which have been deposited in GenBank under the accession numbers MH320097 - MH320504. Out of 103 pinned bee specimens of replicated species aged between 0 - 5 years, 71% (81% of the species included) yielded high quality sequences from the three loci targeted. For all loci targeted, the percentage of samples that passed the final filter of 95% sequence similarity between replicates was higher for ethanol preserved specimens than for pinned specimens (Fig. 2.4). For pinned specimens, high quality long read sequences could not be generated for specimens older than 10 years for EF-1α and wg and older than 7 years for CO1 (5' and 3') loci. For ethanol preserved specimens, high quality sequences of the three gene regions targeted were obtained from all age groups represented in this study (1-14 years).

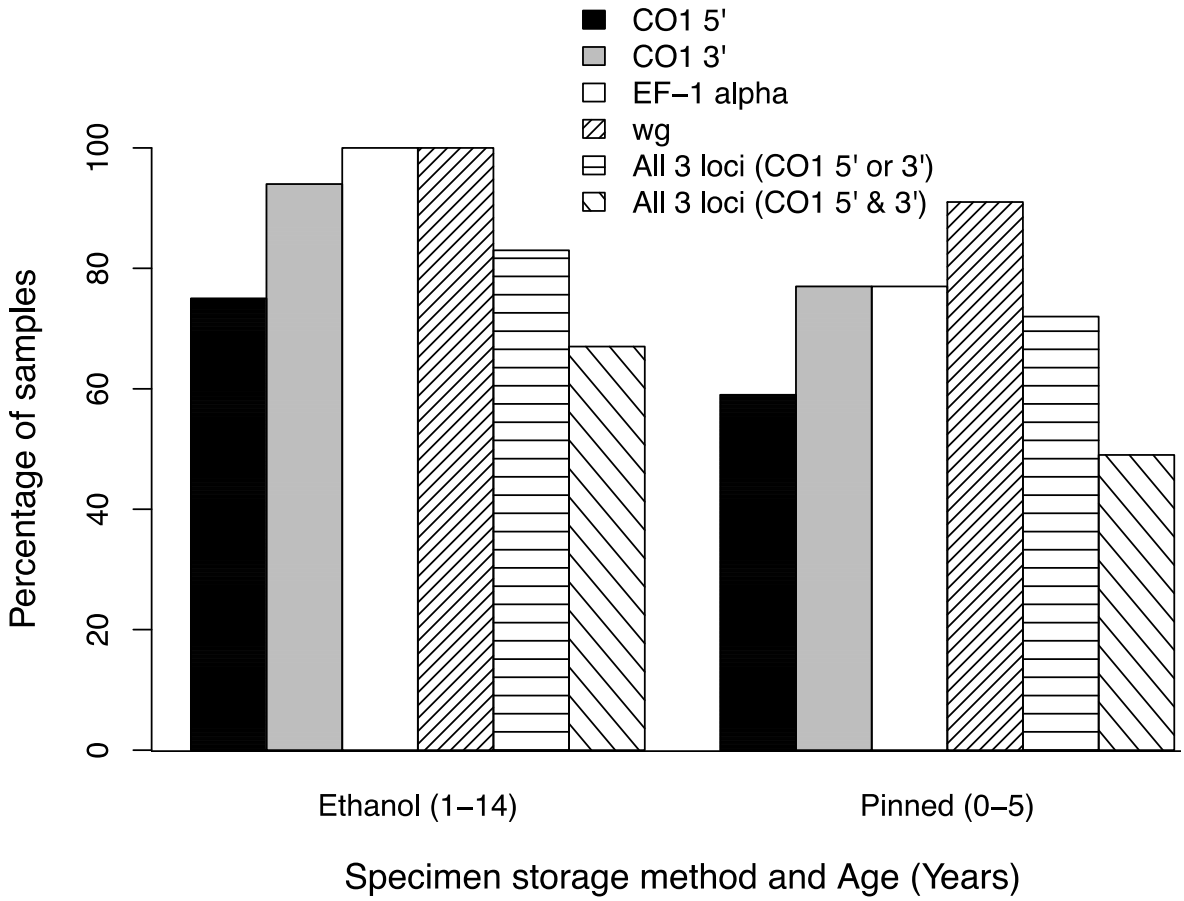

Figure 2.4. The percentage of samples that passed the final filter of 95% sequence similarity between replicates, grouped by method of storage. No data are given for pinned specimens older than five years as they were not sufficiently replicated. Naming convention: CO1 5' & CO1 3' = cytochrome c oxidase subunit 1 5' & 3', EF-1 alpha = eukaryotic translation elongation factor 1 alpha, wg = wingless gene region.

The multilocus and CO1 phylogenies were mostly congruent with species replicates grouping together but there was disagreement in the placement of 7 species (marked with an asterisk in both phylogenies). The multilocus phylogeny was better resolved with stronger posterior support for the internal nodes (Fig. 2.5 and Fig. S2.4).

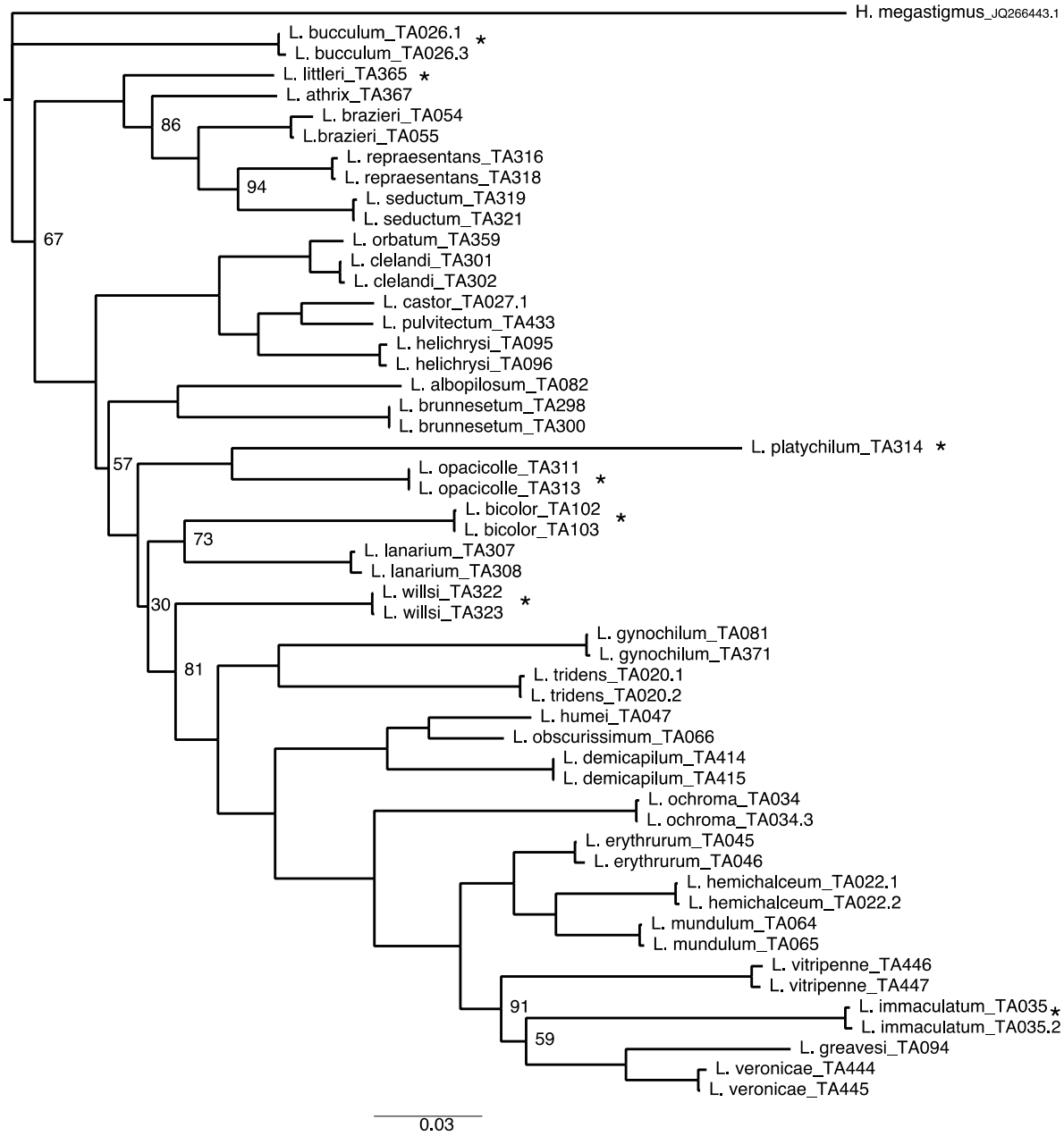

Figure 2.5. A Bayesian consensus multilocus phylogeny of *L.* (*Chilalictus*) species constructed using concatenated sequences of elongation factor 1 alpha (EF-1α), wingless (wg) and cytochrome c oxidase (*CO1*). Posterior probabilities of < 99% are shown above the nodes. Asterisks mark disagreements in the placement of species between the phylogeny constructed with concatenated gene sequences and a *CO1* phylogeny (Fig. S2.4).

Discussion

We have developed a nested PCR approach using the MiSeq for next generation sequencing of multiple loci that are longer than 600 bp from pinned specimens of up to 10 years and from ethanol preserved specimens of at least 14 years depending on the target gene region. Using the nested amplicon sequencing approach, we generated 407 barcode compliant sequences of two nuclear and one mitochondrial loci. Comparison to previous multilocus amplicon sequencing efforts on the MiSeq (Cruaud et al., 2017; Sonet et al., 2018), shows that our nested amplicon sequencing approach doubled the sequence length for pinned as well as ethanol preserved specimens. The success rates for sequencing each of the three loci were comparable to previous multilocus sequencing methods on the MiSeq using specimens of the same age group (Cruaud et al., 2017; Sonet et al., 2018). While the developed method was only tested on specimens of *L.* (*Chilalictus*), it potentially can be used for generating sequences from fresh, ethanol preserved and pinned insect specimens of the same age group, from different taxonomic groups. Here, we discuss how we dealt with common limitations of amplicon sequencing and compare our method to alternative multilocus sequencing methods.

The amount of DNA recovered from pinned insect specimens declined with increasing specimen age. This is because in pinned specimens, the rate of DNA degradation is thought to occur rapidly after death due to autolytic and hydrolytic processes (Burrell et al., 2015). In contrast to our results, Andersen and Mills (2012) found no correlation between the age of pinned specimens and amount of DNA recovered. However, they noted that their results were exceptional and could be partially explained by their use of a Nanodrop spectrophotometer which does not distinguish between low and high molecular DNA (Andersen & Mills, 2012). On the other hand, there was no correlation

between the age of specimens preserved in ethanol and DNA concentration. DNA has been shown to preserve better in specimens stored in ethanol because ethanol facilitates rapid desiccation of the specimens and arrests the degradation process (Burrell et al., 2015; Stein, White, Mazor, Miller, & Pilgrim, 2013). However, Cruaud et al., (2017) demonstrated that DNA degrades over time in ethanol preserved specimens too and our results might be due to the small sample size of specimens preserved in ethanol. Regardless of the method of storage, the concentration of DNA extracted was low due to the small amounts of starting tissue (Andersen & Mills, 2012).

Low DNA quantities are a major limitation for barcoding specimens because they can reduce the chances of PCR amplification (Burrell et al., 2015). The nested PCR approach attempts to overcome this challenge by amplifying a near full-length target locus in the initial PCR and using aliquots of the PCR products as templates in the subsequent PCR steps. This approach differs from previous nested PCR methods used for amplifying samples with degraded DNA, which target short gene fragments of the target locus in the initial PCR (Sonet et al., 2018; Mitchell, 2015), and therefore would typically require higher DNA quantities for the same number of loci targeted in this study.

Similar to the trends shown regarding DNA concentration, the average number of reads passing the filtering pipeline was negatively correlated with the age of pinned specimens. Specimen age is a strong predictor of the length of PCR amplicons achievable from pinned collections (Andersen & Mills, 2012; Heintzman, Elias, Moore, Paszkiewicz, & Barnes, 2014; Mitchell, 2015), and hence fewer good quality amplicons and consequently fewer reads are expected from older pinned specimens. There was a sharp decline in the number of reads passing the filtering pipeline in pinned

specimens aged four and five years, which is in agreement with previous observations that full length amplicons might not be readily obtained from pinned specimens after 3 years (Mitchell, 2015). However, given the high number of good quality reads generated from specimens aged 7 and 10 in our collection, other factors apart from age, such as storage conditions and speed of desiccation might affect the quality of DNA in pinned specimens (reviewed by Burrell et al., 2015). With regard to the specimens preserved in ethanol, there was no correlation between specimen age and the number of reads passing the filtering pipeline. This implies that our nested PCR approach obtained good quality PCR amplicons across the range of ethanol preserved specimen age used in this study. Our results here differ from Cruaud et al., (2017), who reported a negative correlation between ethanol preserved specimen age and PCR success. This could be because the latter study targeted near full-length loci in one PCR reaction. If that is the case, our approach would have capacity to sequence ethanol preserved specimens that are older than 20 years as used by Cruaud et al. (2017).

Interestingly, only a few reads from the 33-year-old specimen sampled in this study passed the filtering pipeline, despite a high DNA concentration, as only 4.7% of the reads generated from this sample mapped to *Lasioglossum* (*Chilalictus*). This suggests that most of the DNA extracted from this sample resulted from contamination (Burrell et al., 2015). The chances of amplifying contaminant DNA from old specimens is increased when using multiple step PCR reactions (Arandjelovic et al., 2009), and several recommendations have been made to reduce and detect contamination that happens during sequence library preparation (Mitchell, 2015). In addition to the standard wet lab recommendations, the use of nested primers should further lower the risk of environmental DNA contamination by enhancing taxonomic specificity during the second PCR (Carr, Williams, & Hayden, 2010; Reyes & Zervos, 2010). Still,

primers of loci with universally conserved primer sites across taxa, such as the CO1 barcode fragment (Folmer et al., 1994; Sharma & Kobayashi, 2014), would be prone to amplifying non-target regions (Cruaud et al., 2017). Indeed, compared to other gene regions, CO1 had the highest number of environmental contaminant sequences filtered out (supplementary data, Fig. S2.3). Based on the negative controls, however, cross-sample contamination was the main source of external contamination (supplementary data S2.1).

Cross sample contamination could have occurred during removal of the legs, if forceps used were not appropriately cleaned, and at several stages of the library preparation, but is usually caused by pipetting errors and environmental DNA aerosol and can be detected using negative controls (Lee, Lee, Tang, Loh, & Koay, 2016; Mitchell, 2015). We ran three negative controls for each of the eleven gene fragments and although the total read count per fragment was low (maximum 30 reads after filtering for environmental contaminants), 90% of the assembled sequences from negative controls mapped to *L.* (*Chilalictus*) in GenBank (Data S2.1). This suggests that sample to sample contamination was prevalent during library preparation, as has been shown elsewhere for amplicon sequencing using the MiSeq (Cruaud et al., 2017; Lee et al., 2016), but occurred at low levels. Nonetheless, in order to filter out cross sample contamination, we used the most common sequence per specimen and gene fragment to assemble a specimen specific reference sequence for mapping and assembling consensus sequences for each locus. This approach was adopted because our results showed a consistently high proportion of the major sequence cluster (most common sequence read per sample and gene fragment) across all gene fragments sequenced. This is in agreement with previous research in which the major cluster of sequences

generated from most samples on the MiSeq platform, matched the target sequence (Cruaud et al., 2017).

Besides cross sample contamination, multiple step PCR reactions introduce PCR errors in the sequences due to the high number of PCR cycles used (Potapov & Ong, 2017). We followed recommended wet and dry lab practices to deal with PCR errors by using a high fidelity KAPA HiFi polymerase in the initial PCR due to the high number of PCR cycles (Potapov & Ong, 2017), assembling consensus sequences using the most common base following previous research (Cruaud et al., 2017; Ji et al., 2013; Shokralla et al., 2011), translating the assembled consensus sequences to identify premature stop codons if any and comparing the consensus sequences between replicates (Cruaud et al., 2017; Mitchell, 2015) using both gene trees and percentage pairwise similarity scores (Sproul & Maddison, 2017; Cruaud et al., 2017; Meier et al., 2006).

Our modified nested PCR approach has achieved comparable results to previous multilocus amplicon sequencing methods of insects using the MiSeq (Sonet et al., 2018; Cruaud et al., 2017), with 71% of all pinned specimens and 81% of ethanol preserved specimens yielding high quality sequences for each locus sequenced. We achieved complete sequences of target loci that were longer than 600 bp with minimal risks of generating chimeras, a criticism that has been levied on previous barcoding methods using NGS platforms (Hebert et al., 2018). The modified nested PCR approach reduces the risk of chimera formation by amplifying near full-length loci in the initial PCR using amplicon specific primers before subsequently amplifying shorter fragments of the target loci using nested primers. While this reduces chimeras that are formed due to the erroneous combination of sequences from target and non-target loci,

it does not reduce the chimeras that form due to incomplete primer extension or template switching during PCR (Judo, Wedel, & Wilson, 1998; Meyerhans, Vartanian, & Wain-Hobson, 1990; Odelberg, Weiss, Hata, & White, 1995). This is particularly problematic during the index PCR reactions as it can lead to tag jumping and subsequently, the misallocation of sequences to samples (Schnell, Bohmann, & Gilbert, 2015) thus reducing the quality of the consensus sequences assembled. In order to reduce the effects of chimeras on the quality of assembled sequences, we employed the following recommended approaches (Schnell et al. 2015): (a) dual indexing of amplicons to reduce the chances of misallocating sequences to samples; (b) robotic liquid handlers for pipetting to reduce cross contamination between samples; (c) reducing the number of index PCR cycles to 12; (d); only using sequences that matched their samples with both the 5' and 3' index; (e) following Cruaud et al. (2017), trimming the sample files further to remove small clusters (see Custom scripts in Appendix 2.2 for details), which are often associated with chimeras; (f) mapping the sequences to reference sequences of *L*. (*Chilalictus*) before assembling consensus sequences of individual samples.

The difference observed in the sequencing success achieved between ethanol and pinned specimens is likely due to the integrity of DNA that is better preserved in ethanol material (Burrell et al., 2015). Whereas the difference in the sequencing success between loci could be explained by the specificity of the primers used in the initial PCR of the nested amplicon sequencing approach. This is because the poor performance of the CO1 5' gene region in comparison to other gene regions is most likely due to the higher number of non-target sequences generated for CO1 during the library preparation. Indeed, the universal primers of the CO1 5' gene region (Folmer

et al., 1994), also used in this study, have been previously reported to amplify symbionts and parasites during amplicon sequencing of insect collections (Sonet et al., 2018; Bergmann et al., 2013; Cruaud et al., 2017). Therefore, taxon specific primers might be necessary to improve the sequencing success of the CO1 barcode fragment for the *L.* (*Chilalictus*) subgenus. Our results also support previous findings that the amplification and sequencing of the CO1 3' fragment may be more reproducible compared to the CO1 5' barcode fragment in some insect groups (Rach et al., 2017). This further supports the use of a multilocus barcoding approach for insect species identification.

The CO1 barcode fragment was sufficient to identify species of *L.* (*Chilalictus*) in this study, since all species replicates were correctly grouped with high posterior support. This is in agreement with previous studies suggesting that the CO1 barcode fragment is able to resolve most morphologically described halictid species (Schmidt, Schmid-Egger, Morinière, Haszprunar, & Hebert, 2015; Sonet et al., 2018). However, Gibbs (2018) found the CO1 barcode fragment to be insufficient for characterising species from the subgenus *Lasioglossum* (*Dialictus*) - Halictidae, and recommended the use of multilocus barcodes and integrated taxonomy to improve species description in this group. Additionally, the use of multiple loci for phylogenetic construction improved the resolution of species relationships and provided stronger posterior support for the internal nodes of the *L.* (*Chilalictus*) phylogeny.

The nested amplicon sequencing approach developed here, failed to generate barcodes from pinned specimens that were older than 10 years. Therefore, this method cannot be used for multilocus barcoding of pinned type specimens, which are the gold standard

for insect barcoding (Hebert et al., 2013). Several molecular methods that are capable of sequencing type specimens have been developed, including; the single stranded DNA library preparation techniques (Gansauge & Meyer 2019; Gansauge et al. 2017; Gansauge & Meyer 2013), single tube library preparation techniques (Carøe et al. 2017), restriction-associated DNA tags, low coverage shotgun sequencing (Tin, Economo, & Mikheyev, 2014) and hybrid capture (Blaimer, Lloyd, Guillory, & Brady, 2016). These methods have been optimised to produce genomic data from highly degraded samples. This is useful for separating closely related species, and could replace mitochondrial based DNA barcoding technology with time, as sequencing costs become cheaper (Burrell et al. 2015; Tin et al. 2014). However, the immediate goal of the global barcoding initiative is to assemble a complete and accurate DNA barcode reference sequence database for all multicellular species, that could be used for phylogenetic studies among other applications (Hebert, Hollingsworth, & Hajibabaei, 2016). To this end, it is preferable to focus the sequencing efforts on a selection of genes that have been verified for use in insect barcoding and phylogenetic analysis (Cruaud et al. 2017). The hybrid capture method can also be used for multilocus barcoding of type specimens because it is capable of targeting specific loci (Liu et al. 2017; Burrell et al. 2015), has been optimised to use low DNA quantities from highly degraded DNA samples (Sproul & Maddison, 2017) and allows multiplexing several samples (Burrell et al., 2015; McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013). However, it is technically demanding and has high upfront costs (Burrell et al., 2015), which make it less accessible to ordinary laboratories.

Therefore, amplicon-sequencing methods remain preferable for multilocus barcoding of animals because they can be used to target a small number of informative genes from a large number of samples, cost effectively (Burrell et al. 2015). Several amplicon

sequencing methods have been used to generate CO1 barcode sequences from type specimens (Hausmann et al., 2016; Hebert et al., 2013; Prosser et al., 2016; Shokralla et al., 2014), but might need to be modified in order to sequence multiple loci, including nuclear genes that have lower copy numbers in type specimens (Burrell et al., 2015). Multiple loci, including nuclear genes, have been previously obtained from thirteen type specimens of beetles using a combination of reference based and denovo sequence assembly of sequence data generated on the Illumina HiSeq 2000 platform (Kanda et al. 2015). However, this method required a minimum of 9.9 ng/µl per specimen. We could not obtain this amount of DNA from pinned specimens older than 10 years using only one or a few legs. Furthermore, the method might not be scalable due to the cost of sequencing associated with using the Illumina HiSeq platform (Kanda et al. 2015). Multilocus barcoding of ethanol preserved specimens has also been performed using pyrosequencing (Bybee et al., 2011), but this approach is yet to be tested on pinned specimens. Thus, all sequencing approaches have different advantages and limitations, and present trade-offs in sequencing capabilities, such as specimen age or number and length of target loci. Therefore, building a comprehensive multilocus database from insect collections might necessitate using a number of methods in order to leverage their strengths.

Multilocus barcoding promises to improve the accuracy of species identification in some insect groups where the standard CO1 barcode fragment is problematic and allows the use of sequence databases for phylogenetic studies. In order to achieve these objectives, it is necessary to keep improving the capacity of current sequencing technologies. Our nested amplicon sequencing approach has improved the capacity of the MiSeq platform to sequence multiple long-length loci with minimal risks of

generating chimeric sequences. This approach requires limited molecular expertise and could therefore be used in ordinary laboratories. Given the poor state of insect barcoding in developing nations, improving the capabilities of affordable sequencing platforms like the MiSeq, will equip ordinary laboratories to make significant contributions to the global barcoding initiative.

Acknowledgements

We thank Dr Ken Walker of Museums Victoria for identifying and providing samples of specimens, Tessa Bradford of the South Australian Regional Facility for Molecular Ecology and Evolution for providing guidance in the laboratory, and three anonymous reviewers for useful comments on the manuscript. This project was supported by funding from The Holsworth Wildlife Research Endowment Fund and the Australian Government Department of Agriculture and Water Resources through its Rural R&D for Profit program (Grant number 15-02-035).

References

Almeida, E. A. B., & Danforth, B. N. (2009). Phylogeny of colletid bees (Hymenoptera: Colletidae) inferred from four nuclear genes. *Molecular Phylogenetics and Evolution*, *50*(2), 290–309. doi:10.1016/j.ympev.2008.09.028

Andersen, J. C., & Mills, N. J. (2012). DNA extraction from museum specimens of parasitic Hymenoptera. *Plos One*, *7*(10), e45549. doi:10.1371/journaL. pone.0045549

Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T. R., Thalmann, O., Siedel, H., & Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from non-invasive and museum samples. *Molecular Ecology Resources*, *9*(1), 28–36. doi:10.1111/j.1755-0998.2008.02387.x

Arthur, A. D., Li, J., Henry, S., & Cunningham, S. A. (2010). Influence of woody vegetation on pollinator densities in oilseed *Brassica* fields in an Australian temperate landscape. *Basic and Applied Ecology*, *11*(5), 406–414. doi:10.1016/j.baae.2010.05.001

BBMerge Guide - DOE Joint Genome Institute. (2017). Retrieved October 1, 2017, from https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmerge-guide/

Bell, J. (2011). *Overview of tailed amplicon sequencing approach with MiSeq*. USA: Illumina.

Bergmann, T., Rach, J., Damm, S., Desalle, R., Schierwater, B., & Hadrys, H. (2013). The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. *Molecular*

Ecology Resources, *13*(6), 1069–1081. doi:10.1111/1755-0998.12125

Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *Plos One*, *11*(8), e0161531. doi:10.1371/journal.pone.0161531

Brower, A. V., & DeSalle, R. (1998). Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Molecular Biology*, *7*(1), 73–82. doi:10.1046/j.1365-2583.1998.71052.x

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, 35–44. doi:10.1016/j.jhevoL. 2014.10.015

Bybee, S. M., Bracken-Grissom, H., Haynes, B. D., Hermansen, R. A., Byers, R. L., Clement, M. J., Crandall, K. A. (2011). Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution*, *3*, 1312–1323. doi:10.1093/gbe/evr106

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Gilbert, M. T. P. (2017). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution / British Ecological Society*. doi:10.1111/2041-210X.12871

Carr, J., Williams, D. G., & Hayden, R. T. (2010). Molecular detection of multiple respiratory viruses. In *Molecular Diagnostics* (pp. 289–300). Elsevier. doi:10.1016/B978-0-12-369428-7.00024-0

Cruaud, P., Rasplus, J. Y., Rodriguez, L. J., & Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy.

Scientific Reports, *7*, 41948. doi:10.1038/srep41948

Cunningham, S. A., Schellhorn, N. A., Marcora, A., & Batley, M. (2013). Movement and phenology of bees in a subtropical Australian agricultural landscape. *Austral Ecology*, *38*(4), 456–464. doi:10.1111/j.1442-9993.2012.02432.x

Danforth, B. N., & Ji, S. (1998). Elongation factor-1α occurs as two copies in bees: Implications for phylogenetic analysis of EF-1α sequences in insects. *Molecular Biology and Evolution*, *15*(3), 225–235.

Danforth, B. N., Brady, S. G., Sipes, S. D., & Pearson, A. (2004). Single-copy nuclear genes recover cretaceous-age divergences in bees. *Systematic Biology*, *53*(2), 309–326. doi:10.1080/10635150490423737

Danforth, B. N., & Ji, S. (2001). Australian *Lasioglossum* + *Homalictus* form a monophyletic group: resolving the "Australian enigma". *Systematic Biology*, *50*(2), 268–283.

Dunn, C. P. (2003). Keeping taxonomy based in morphology. *Trends in Ecology & Evolution*, *18*(6), 270–271. doi:10.1016/S0169-5347(03)00094-6

Folmer, Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Gansauge, M.T., & Meyer, M. (2019). A Method for Single-Stranded Ancient DNA Library Preparation. *Methods in Molecular Biology*, *1963*, 75–83. doi:10.1007/978-1-4939-9176-1_9

Gansauge, M.T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research*, *45*(10), e79.

doi:10.1093/nar/gkx033

Gansauge, M.T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, *8*(4), 737–748. doi:10.1038/nprot.2013.038

Gibbs, J, Brady, S. G., Kanda, K., & Danforth, B. N. (2012). Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Molecular Phylogenetics and Evolution*, *65*(3), 926–939. doi:10.1016/j.ympev.2012.08.013

Gibbs, Jason. (2018). DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. *Genome*, *61*(1), 21–31. doi:10.1139/gen-2017-0096

Hausmann, A., Miller, S. E., Holloway, J. D., deWaard, J. R., Pollock, D., Prosser, S. W. J., & Hebert, P. D. N. (2016). Calibrating the taxonomy of a megadiverse insect family: 3000 DNA barcodes from geometrid type specimens (Lepidoptera, Geometridae). *Genome*, *59*(9), 671–684. doi:10.1139/gen-2015-0197

Hebert, P. D. N., Braukmann, T. W. A., Prosser, S. W. J., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., Zakharov, E. V. (2018). A sequel to sanger: amplicon sequencing that scales. *BMC Genomics*, *19*(1), 219. doi:10.1186/s12864-018-4611-3

Hebert, P. D. N., Hollingsworth, P. M., & Hajibabaei, M. (2016). From writing to reading the encyclopedia of life. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1702). doi:10.1098/rstb.2015.0321

Hebert, P. D. N., Dewaard, J. R., Zakharov, E. V., Prosser, S. W. J., Sones, J. E., McKeown, J. T. A., La Salle, J. (2013). A DNA "barcode blitz": rapid digitization and sequencing of a natural history collection. *Plos One*, *8*(7), e68535. doi:10.1371/journaL. pone.0068535

Heintzman, P. D., Elias, S. A., Moore, K., Paszkiewicz, K., & Barnes, I. (2014). Characterizing DNA preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera). *Molecular Ecology Resources*, *14*(3), 606–615. doi:10.1111/1755-0998.12205

Illumina. (2013). Preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system. Retrieved May 30, 2017, from https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. doi:10.1111/ele.12162

Judo, M. S., Wedel, A. B., & Wilson, C. (1998). Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research*, *26*(7), 1819–1825. doi:10.1093/nar/26.7.1819

Kanda, K., Pflug, J. M., Sproul, J. S., Dasenko, M. A., & Maddison, D. R. (2015). Successful Recovery of Nuclear Protein-Coding Genes from Small Insects in Museums Using Illumina Sequencing. *Plos One*, *10*(12), e0143929. doi:10.1371/journal.pone.0143929

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic*

Acids Research, *30*(14), 3059–3066. doi:10.1093/nar/gkf436

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. doi:10.1093/bioinformatics/bts199

Klopfstein, Seraina, Kropf, C., & Baur, H. (2016). Wolbachia endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society*, *177*(3), n/a-n/a. doi:10.1111/zoj.12380

Koroiva, R., & Kvist, S. (2018). Estimating the barcoding gap in a global dataset of cox1 sequences for Odonata: close, but no cigar. *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis*, *29*(5), 765–771. doi:10.1080/24701394.2017.1357709

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. doi:10.1093/molbev/msw260

Lee, H. K., Lee, C. K., Tang, J. W.T., Loh, T. P., & Koay, E. S.-C. (2016). Contamination-controlled high-throughput whole genome sequencing for influenza A viruses using the MiSeq sequencer. *Scientific Reports*, *6*, 33318. doi:10.1038/srep33318

Leijs, R., Batley, M., & Hogendoorn, K. (2017). The genus *Amegilla* (Hymenoptera, Apidae, Anthophorini) in Australia: A revision of the subgenera *Notomegilla* and *Zonamegilla*. *ZooKeys*, (653), 79–140. doi:10.3897/zookeys.653.11177

Liu, J., Jiang, J., Song, S., Tornabene, L., Chabarria, R., Naylor, G. J. P., & Li, C.

(2017). Multilocus DNA barcoding - species identification with multilocus data. *Scientific Reports*, *7*(1), 16601. doi:10.1038/s41598-017-16920-2

Mallo, D., & Posada, D. (2016). Multilocus inference of species trees and DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1702). doi:10.1098/rstb.2015.0335

Mardulyn, P., & Cameron, S. A. (1999). The major opsin in bees (Insecta: Hymenoptera): A promising nuclear gene for higher level phylogenetics. *Molecular Phylogenetics and Evolution*, *12*(2), 168–176. doi:10.1006/mpev.1998.0606

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, *66*(2), 526–538. doi:10.1016/j.ympev.2011.12.007

Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. L. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*(5), 715–728. doi:10.1080/10635150600969864

Meyerhans, A., Vartanian, J. P., & Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Research*, *18*(7), 1687–1691. doi:10.1093/nar/18.7.1687

Michener, C. D., (1965). A classification of the bees of the Australian and South Pacific regions. Bulletin of the AMNH ; v. 130.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)* (pp. 1–8). IEEE.

doi:10.1109/GCE.2010.5676129

Mitchell, A. (2015). Collecting in collections: a PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. *Molecular Ecology Resources*, *15*(5), 1102–1111. doi:10.1111/1755-0998.12380

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How many species are there on Earth and in the ocean? *PLoS Biology*, *9*(8), e1001127. doi:10.1371/journaL. pbio.1001127

Moulton, M. J., Song, H., & Whiting, M. F. (2010). Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources*, *10*(4), 615–627. doi:10.1111/j.1755-0998.2009.02823.x

Odelberg, S. J., Weiss, R. B., Hata, A., & White, R. (1995). Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. *Nucleic Acids Research*, *23*(11), 2049–2057. doi:10.1093/nar/23.11.2049

Paknia, O., Bergmann, T., & Hadrys, H. (2015). Some 'ant'swers: Application of a layered barcode approach to problems in ant taxonomy. *Molecular Ecology Resources*, *15*(6), 1262–1274. doi:10.1111/1755-0998.12395

Potapov, V., & Ong, J. L. (2017). Examining sources of error in PCR by single-molecule sequencing. *Plos One*, *12*(1), e0169774. doi:10.1371/journaL. pone.0169774

PREMIER Biosoft. (2018). Primer design guide for PCR : learn designing primers for PCR. Retrieved June 1, 2017, from http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html

Prosser, S. W. J., deWaard, J. R., Miller, S. E., & Hebert, P. D. N. (2016). DNA barcodes from century-old type specimens using next-generation sequencing.

Molecular Ecology Resources, *16*(2), 487–497. doi:10.1111/1755-0998.12474

Rach, J., Bergmann, T., Paknia, O., DeSalle, R., Schierwater, B., & Hadrys, H. (2017). The marker choice: Unexpected resolving power of an unexplored CO1 region for layered DNA barcoding approaches. *Plos One*, *12*(4), e0174842. doi:10.1371/journaL. pone.0174842

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355–364. doi:10.1111/j.1471-8286.2007.01678.x

Reyes, K., & Zervos, M. (2010). Role of molecular techniques in the epidemiologic investigation of hospital-acquired infections. In *Molecular Diagnostics* (pp. 351–364). Elsevier. doi:10.1016/B978-0-12-369428-7.00029-X

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. doi:10.1093/sysbio/sys029

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated--reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303. doi:10.1111/1755-0998.12402

Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., & Hebert, P. D. N. (2015). DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). *Molecular Ecology Resources*, *15*(4), 985–1000. doi:10.1111/1755-0998.12363

Sharma, P., & Kobayashi, T. (2014). Are “universal” DNA primers really universal?

Journal of Applied Genetics, *55*(4), 485–496. doi:10.1007/s13353-014-0218-9

Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources*, *14*(5), 892–901. doi:10.1111/1755-0998.12236

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, *5*, 9687. doi:10.1038/srep09687

Shokralla, S., Zhou, X., Janzen, D. H., Hallwachs, W., Landry, J.F., Jacobus, L. M., & Hajibabaei, M. (2011). Pyrosequencing for mini-barcoding of fresh and old museum specimens. *Plos One*, *6*(7), e21252. doi:10.1371/journaL. pone.0021252

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, *87*(6), 651–701. doi:10.1093/aesa/87.6.651

Sonet, G., Pauly, A., Nagy, Z. T., Virgilio, M., Jordaens, K., Van Houdt, J., Backeljau, T. (2018). Using next-generation sequencing to improve DNA barcoding: lessons from a small-scale study of wild bee species (Hymenoptera, Halictidae). *Apidologie*, *49*(5), 671–685. doi:10.1007/s13592-018-0594-y

Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13486–13491.

doi:10.1073/pnas.0803076105

Spaethe, J., & Briscoe, A. D. (2004). Early duplication and functional diversification of the opsin gene family in insects. *Molecular Biology and Evolution*, *21*(8), 1583–1594. doi:10.1093/molbev/msh162

Sproul, J. S., & Maddison, D. R. (2017). Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources*, *17*(6), 1183–1201. doi:10.1111/1755-0998.12660

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. doi:10.1093/bioinformatics/btu033

Stein, E. D., White, B. P., Mazor, R. D., Miller, P. E., & Pilgrim, E. M. (2013). Evaluating ethanol-based sample preservation to facilitate use of DNA barcoding in routine freshwater biomonitoring programs using benthic macroinvertebrates. *Plos One*, *8*(1), e51273. doi:10.1371/journaL. pone.0051273

Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, *63*, 31–45. doi:10.1146/annurev-ento-020117-043348

Sunnucks, P., & Hales, D. F. (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, *13*(3), 510–524.

Tin, M. M.Y., Economo, E. P., & Mikheyev, A. S. (2014). Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *Plos One*, *9*(5), e96793. doi:10.1371/journal.pone.0096793

Walker, K. L. (1995). Revision of the Australian native bee subgenus *Lasioglossum* (*Chilalictus*). *Memoirs of Museum Victoria*, *55*(1&2), 1–423.

Ward, C., To, H., & Pederson, S. (2017, November 17). An introduction to ngsReports. Retrieved October 1, 2017, from https://uofabioinformaticshub.github.io/ngsReports/vignettes/ngsReportsIntroduction

Data Accessibility

Raw data from MiSeq: GenBank and can be accessed using the link; https://www.ncbi.nlm.nih.gov/sra/SRP151814.

DNA Sequences: GenBank accession numbers MH320097 - MH320504

Custom Scripts used for analysing raw data and assembling consensus sequences: Deposited on Github and can be accessed via this link;

https://github.com/UofABioinformaticsHub/Trace_NestedAmpliconSeqAnalysis.

Also deposited in Dyrad under the DOI: https://doi.org/10.5061/dryad.g05g074.

Data S2.1: GenBank blast search results for the consensus sequences generated in this study: Deposited in Dyrad under the DOI: https://doi.org/10.5061/dryad.g05g074.

The files archived with Dyrad are currently accessible via this link;

https://datadryad.org/stash/share/T1wl7rMCpHoekuEACE31xSDGFycyjFr2ola39kVf1tg.

Appendices

Appendix 2.1. In-house modified Gentra protocol used for extracting DNA from pinned specimens.

Appendix 2.2. Custom Scripts used for analysing raw data and assembling consensus sequences.

Author Contributions

Trace Akankunda: designed the project, performed the laboratory work, analysed data, wrote the paper. Katja Hogendoorn: designed the project, reviewed the data analysis and the paper. Carlos Rodriguez Lopez: designed the project, provided advice on molecular methods, reviewed the data analysis and the paper. Remko Leijs: provided advice on molecular methods, reviewed the data analysis and the paper. Hien To: assisted with the bioinformatics and wrote scripts for the analysis.

Supporting information

Table S2.1. Metadata of the specimens used in this study. Average bee sizes are extracted from Ken Walker's revision of *Lasioglossum* (*Chilalictus*) (Memoirs of the Museum of Victoria 55(1&2): 1-423 (1995).

Species Name	DNA extraction number	Voucher ID	Datum	Age	Storage method	Ave. Size (mm)	DNA ng/µl (30 µl)
L. plebeium	TA037	RL1883	31/8/16	1	ethanol	5.52	0.99
L. plebeium	TA037.2	RL1883	31/8/16	1	ethanol	5.52	0.98
L. ochroma	TA034	RL1777	28/6/11	6	ethanol	5.07	0.37
L. ochroma	TA034.2	RL1777	28/6/11	6	ethanol	5.07	0.67
L. ochroma	TA034.3	RL1777	28/6/11	6	ethanol	5.07	0.46
L. immaculatum	TA035	RL1810	26/6/11	6	ethanol	4.71	0.38
L. immaculatum	TA035.2	RL1810	26/6/11	6	ethanol	4.71	0.19
L. bucculum	TA026.1	RL1256	16/9/09	8	ethanol	7.97	1.9
L. bucculum	TA026.2	RL1256	16/9/09	8	ethanol	7.97	1.72
L. bucculum	TA026.3	RL1256	16/9/09	8	ethanol	7.97	2.32
L. castor	TA027.1	RL1265	16/9/09	8	ethanol	6.77	2.87
L. cephalochilum	TA028.1	RL1281	17/9/09	8	ethanol	5.12	0.76
L. tridens	TA020.1	RL0345	8/11/03	14	ethanol	5.42	0.51
L. tridens	TA020.2	RL0345	8/11/03	14	ethanol	5.42	1.03
L. tridens	TA021	RL0362	30/10/03	14	ethanol	5.42	1.12
L. hemicalceum	TA022.1	RL0346	15/11/03	14	ethanol	5.04	0.74
L. hemicalceum	TA022.2	RL0346	15/11/03	14	ethanol	5.04	0.78
L. hemicalceum	TA022.3	RL0346	15/11/03	14	ethanol	5.04	0.56
L. clelandi	TA301	KR05647	11/1/17	0	pinned	8.13	1.43
L. lanarium	TA306	KR05378	11/1/17	0	pinned	10.03	*3.01*
L. platychilum	TA314	KR05690	10/1/17	0	pinned	6.73	3.38
L. platychilum	TA315	KR05691	10/1/17	0	pinned	6.73	3.35
L. willsi	TA322	KR05398	11/1/17	0	pinned	6.44	3.6
L. willsi	TA323	KR05386	11/1/17	0	pinned	6.44	3.42
L. vitripenne	TA446	KR05956	10/2/17	0	pinned	5.81	19
L. vitripenne	TA447	KR05262	8/1/17	0	pinned	5.81	1.61
L. bicolor	TA102	KR04427	17/2/16	1	pinned	5.81	3.83
L. bicolor	TA103	KR04426	17/2/16	1	pinned	5.81	1.43
L. pulvitectum	TA108	KR04347	18/2/16	1	pinned	6.86	4.43

L. brunnesetum	TA298	KR05180	12/12/16	1	pinned	5.39	1.48
L. brunnesetum	TA299	KR05181	12/12/16	1	pinned	5.39	1.21
L. brunnesetum	TA300	KR05182	12/12/16	1	pinned	5.39	1.06
L. clelandi	TA302	KR05348	13/12/16	1	pinned	8.13	2.48
L. clelandi	TA303	KR05600	28/10/16	1	pinned	8.13	1.83
L. littleri	TA304	KR05582	28/10/16	1	pinned	8.6	1.94
L. littleri	TA305	KR05583	28/10/16	1	pinned	8.6	3.63
L. lanarium	TA307	KR05375	15/12/16	1	pinned	9.02	*2.04*
L. lanarium	TA308	KR05342	13/12/16	1	pinned	10.03	*2.04*
L. littleri	TA309	KR05557	28/10/16	1	pinned	8.6	1.68
L. littleri	TA310	KR05608	28/10/16	1	pinned	8.6	0.11
L. opacicolle	TA311	KR05591	28/10/16	1	pinned	7.5	0.05
L. opacicolle	TA312	KR05592	28/10/16	1	pinned	7.5	1.93
L. opacicolle	TA313	KR05595	28/10/16	1	pinned	7.5	3.14
L. repraesentans	TA316	KR05160	12/12/16	1	pinned	6.85	1.45
L. repraesentans	TA317	KR05163	12/12/16	1	pinned	6.85	1.51
L. repraesentans	TA318	KR05162	12/12/16	1	pinned	6.85	2.06
L. seductum	TA319	KR05627	28/10/16	1	pinned	8.47	1.54
L. seductum	TA320	KR05618	28/10/16	1	pinned	8.47	1.96
L. seductum	TA321	KR05601	28/10/16	1	pinned	8.47	4.14
L. willsi	TA324	KR05490	4/11/16	1	pinned	6.44	2.99
L. demicapilum	TA414	KR04959	9/10/16	1	pinned	4.53	3.38
L. demicapilum	TA415	KR04960	9/10/16	1	pinned	4.53	34
L. demicapilum	TA416	KR04962	9/10/16	1	pinned	4.53	1.87
L. plebeium	TA430	KR04614	31/8/16	1	pinned	5.52	1.35
L. sororculum	TA438	KR04646	3/9/16	1	pinned	4.77	0.83
L. erythrurum	TA045	KR02775	13/9/15	2	pinned	4.5	1.85
L. erythrurum	TA046	KR03655	31/10/15	2	pinned	4.5	0.84
L. humei	TA047	KR03768	12/11/15	2	pinned	4.94	1.21
L. quadratum	TA059	KR02190	15/3/15	2	pinned	4.24	0.4
L. mundulum	TA063	KR04234	23/12/15	2	pinned	4.54	0.44
L. mundulum	TA064	KR04231	23/12/15	2	pinned	4.54	1.1
L. mundulum	TA065	KR04230	23/12/15	2	pinned	4.54	0.15
L. sororculum	TA070	KR02124	13/3/15	2	pinned	4.77	0.74
L. sororculum	TA071	KR02123	13/3/15	2	pinned	4.77	0.34
L. cf erythrurm	TA087	KR03250	2/10/15	2	pinned	4.5	0.1
L. cf erythrurm	TA088	KR03380	25/9/15	2	pinned	4.5	0.36
L. cf erythrurm	TA089	KR03182	24/9/15	2	pinned	4.5	0.49
L. globosum	TA090	KR03531	8/10/15	2	pinned	6.99	2.61
L. erythrurum	TA093	KR02560	15/3/15	2	pinned	4.5	0.5
L. helichrysi	TA095	KR03581	22/10/15	2	pinned	6.86	2.86

L. helichrysi	TA096	KR03404	30/9/15	2	pinned	6.86	1.95
L. chapmani	TA100	KR02126	13/3/15	2	pinned	7.05	2.66
L. platychilum	TA106	KR03703	31/10/15	2	pinned	7.49	0.2
L. fasciatum	TA361	KR02119	13/3/15	2	pinned	4.33	0.16
L. littleri	TA364	KR03427	9/10/15	2	pinned	8.6	3.18
L. littleri	TA365	KR03442	9/10/15	2	pinned	8.6	3.91
L. athrix	TA367	KR03517	3/10/15	2	pinned	9.36	1.52
L. cognatum	TA376	KR03705	31/10/15	2	pinned	6.39	0.74
L. cognatum	TA377	KR03750	1/11/15	2	pinned	6.39	0.95
L. cognatum	TA391	KR03669	31/10/15	2	pinned	6.39	2.6
L. erythrurm	TA392	KR03926	4/12/15	2	pinned	4.5	0.89
L. erythrurm	TA396	KR04196	3/12/15	2	pinned	4.5	1.09
L. erythrurm	TA397	KR03815	12/11/15	2	pinned	4.5	2.36
L. instabilis	TA398	KR03611	23/10/15	2	pinned	9.25	1.97
L. fasciatum	TA400	KR02122	13/3/15	2	pinned	4.33	0.52
L. fasciatum	TA401	KR02120	13/3/15	2	pinned	4.33	0.22
L. globosum	TA404	KR03191	24/9/15	2	pinned	6.99	2.39
L. globosum	TA405	KR03592	9/10/15	2	pinned	6.99	2.5
L. pulvitectum	TA432	KR04019	4/12/15	2	pinned	6.86	2.41
L. pulvitectum	TA433	KR03836	4/12/15	2	pinned	6.86	2.65
L. quadratum	TA434	KR02188	15/3/15	2	pinned	4.24	0.08
L. quadratum	TA435	KR02189	15/3/15	2	pinned	4.24	0.46
L. veronicae	TA444	KR03873	4/12/15	2	pinned	5.2	0.47
L. veronicae	TA445	KR03925	4/12/15	2	pinned	5.2	0.68
L. expansifrons	TA048	KR01260	20/9/14	3	pinned	5.56	3.92
L. expansifrons	TA050	KR01193	13/9/14	3	pinned	5.56	0.75
L. chapmani	TA052	KR01737	25/10/14	3	pinned	7.05	1.53
L. brazieri	TA053	KR01831	9/11/14	3	pinned	10.31	7.1
L. brazieri	TA054	KR01856	9/11/14	3	pinned	10.31	7
L. brazieri	TA055	KR01830	9/11/14	3	pinned	10.31	3.39
L. obscurissimum	TA066	KR01694	13/10/14	3	pinned	4.83	1.86
L. speculatum	TA073	KR01739	25/10/14	3	pinned	8	4.55
L. speculatum	TA074	KR01984	20/12/14	3	pinned	8	3.36
L. vitripenne	TA078	KR02077	21/12/14	3	pinned	5.81	0.22
L. gynochilum	TA081	KR01602	11/10/14	3	pinned	5.34	0.71
L. albopilosum	TA082	KR01210	13/9/14	3	pinned	6.07	0.66
L. greavesi	TA094	KR01308	25/9/14	3	pinned	4.07	0.02
L. helichrysi	TA097	KR01904	9/11/14	3	pinned	6.86	2.46
L. littleri	TA366	KR01171	13/9/14	3	pinned	8.6	3.45
L. gynochilum	TA370	KR01611	11/10/14	3	pinned	5.34	0.95
L. gynochilum	TA371	KR01612	11/10/14	3	pinned	5.34	1.18

L. albopilosum	TA083	KR00536	18/5/13	4	pinned	4.45	0.43
L. chapmani	TA057	KR00423	22/10/12	5	pinned	7.05	0.86
L. oblitum	TA062	KR00027	29/9/12	5	pinned	7.42	0.05
L. gilesi	TA072	KR00086	29/9/12	5	pinned	8.72	0.75
L. veronicae	TA076	KR00319	29/9/12	5	pinned	5.2	0.20
L. impunctatum	TA084	KR00301	22/10/12	5	pinned	7.01	0.39
L. impunctatum	TA085	KR00391	22/10/12	5	pinned	7.01	0.02
L. impunctatum	TA086	KR00394	22/10/12	5	pinned	7.01	0.06
L. conspicuum	TA360	KR00180	29/10/12	5	pinned	8.26	1.64
L. athrix	TA374	KR00296	22/9/12	5	pinned	9.36	1.37
L. athrix	TA375	KR00009	22/9/12	5	pinned	9.36	1.49
L. eremaean	TA393	KR00300	22/10/12	5	pinned	6.86	0.40
L. eremaean	TA394	KR00406	22/10/12	5	pinned	6.86	1.05
L. eremaean	TA395	KR00422	22/10/12	5	pinned	6.86	0.9
L. mediopolitum	TA402	KR00199	22/10/12	5	pinned	4.74	2.75
L. impunctatum	TA412	KR00278	22/10/12	5	pinned	7.01	0.37
L. impunctatum	TA413	KR00302	22/10/12	5	pinned	7.01	0.261
L. orbatum	TA358		30/12/10	7	pinned	8.64	*2.6*
L. orbatum	TA359		31/12/10	7	pinned	8.64	1.82
L. orbatum	TA357		16/1/07	10	pinned	8.64	0.51
L. albopilosum	TA119	HYM38760	26/10/96	21	pinned	6.07	0.2
L. chapmani	TA289	HYM535	27/10/96	21	pinned	7.05	0.2
L. amplexum	TA121	HYM39034	10/10/95	22	pinned	6.24	0.11
L. erythrurum	TA278	HYM1089	2/12/88	29	pinned	4.5	0.12
L. littleri	TA180	HYM7367	16/11/84	33	pinned	8.6	4.9
L. asperithorax	TA127	HYM5759	24/10/73	44	pinned	8	0.15
L. brunnesetum	TA137	HYM4308	12/11/68	49	pinned	5.39	0

Table S2.2. Illumina P5 and P7 Indexed primers used in this study.

Illumina Index	Illumina Primer Sequences
N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG
N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG
N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG
N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG
N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG
N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG
N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG
N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG
N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG
N714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGG
N715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTCGG
N716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTCGG
N718	CAAGCAGAAGACGGCATACGAGATGTAGCTCCGTCTCGTGGGCTCGG
N719	CAAGCAGAAGACGGCATACGAGATTACTACGCGTCTCGTGGGCTCGG
N720	CAAGCAGAAGACGGCATACGAGATAGGCTCCGGTCTCGTGGGCTCGG
N721	CAAGCAGAAGACGGCATACGAGATGCAGCGTAGTCTCGTGGGCTCGG
N722	CAAGCAGAAGACGGCATACGAGATCTGCGCATGTCTCGTGGGCTCGG
N723	CAAGCAGAAGACGGCATACGAGATGAGCGCTAGTCTCGTGGGCTCGG
N724	CAAGCAGAAGACGGCATACGAGATCGCTCAGTGTCTCGTGGGCTCGG
S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
S503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC
S506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC
S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC
S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC
S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTC
S511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTC

Table S2.3. Summary of the raw data generated in this study.

Sample Name	Paired Reads	Data Yield (bp)	Merged Reads	Read Count after filtering out environmental contaminants
Neg1	582	0.00 Gb	116	41
Neg2	672	0.00 Gb	190	110
Neg3	802	0.00 Gb	105	94
TA020_1	264,010	0.16 Gb	112720	130812
TA020_2	210,004	0.13 Gb	100024	96118
TA021	189,186	0.11 Gb	116477	92645
TA022_1	212,486	0.13 Gb	119573	109256
TA022_2	198,700	0.12 Gb	114979	98730
TA022_3	211,653	0.13 Gb	74888	103159
TA026_1	205,292	0.12 Gb	115810	101136
TA026_2	188,874	0.11 Gb	103977	88942
TA026_3	191,447	0.11 Gb	96718	98463
TA027_1	200,064	0.12 Gb	99632	78800
TA028_1	147,630	0.09 Gb	104988	69484
TA034	221,732	0.13 Gb	75237	90448
TA034_2	232,208	0.14 Gb	151381	87846
TA034_3	213,708	0.13 Gb	111827	80631
TA035	234,232	0.14 Gb	97276	115470
TA035_2	140,107	0.08 Gb	111808	73336
TA037	193,504	0.12 Gb	102520	99287
TA037_2	215,527	0.13 Gb	106811	108625
TA045	257,500	0.15 Gb	117338	78840
TA046	201,876	0.12 Gb	125374	86438
TA047	154,105	0.09 Gb	95231	68130
TA048	232,927	0.14 Gb	98759	119115
TA050	128,773	0.08 Gb	119406	64996
TA052	172,323	0.10 Gb	116391	81779
TA053	221,820	0.13 Gb	124314	112843
TA054	220,439	0.13 Gb	124405	105010
TA055	199,450	0.12 Gb	97448	94799
TA057	94,834	0.06 Gb	92147	42836
TA059	175,571	0.11 Gb	101159	69183
TA062	75,632	0.05 Gb	107041	38310
TA063	209,015	0.13 Gb	112083	112977
TA064	214,255	0.13 Gb	98801	106751
TA065	127,350	0.08 Gb	135958	64935
TA066	220,229	0.13 Gb	156589	91326
TA070	139,871	0.08 Gb	91065	60829
TA071	203,267	0.12 Gb	125417	91827

TA072	16,571	0.01 Gb	118228	63
TA073	112,639	0.07 Gb	127334	50497
TA074	177,049	0.11 Gb	102843	70974
TA076	63,493	0.04 Gb	60680	17365
TA078	148,348	0.09 Gb	82818	81552
TA081	167,105	0.10 Gb	118079	82903
TA082	272,700	0.16 Gb	92088	109692
TA083	128,768	0.08 Gb	119973	36227
TA084	148,171	0.09 Gb	91772	64813
TA085	9,585	0.01 Gb	100697	332
TA086	43,186	0.03 Gb	114826	17440
TA087	114,341	0.07 Gb	129584	55908
TA088	213,113	0.13 Gb	127421	89474
TA089	134,715	0.08 Gb	126349	41794
TA090	267,025	0.16 Gb	91177	100173
TA093	221,605	0.13 Gb	105008	113877
TA094	179,832	0.11 Gb	115136	87458
TA095	172,064	0.10 Gb	108706	67541
TA096	153,629	0.09 Gb	110971	67739
TA097	204,190	0.12 Gb	136441	84292
TA100	205,141	0.12 Gb	107130	113662
TA102	183,886	0.11 Gb	76396	94327
TA103	182,102	0.11 Gb	79100	85933
TA106	115,593	0.07 Gb	114269	48761
TA108	193,888	0.12 Gb	109216	75129
TA119	27,990	0.02 Gb	66085	3967
TA121	30,459	0.02 Gb	69853	1494
TA127	89,234	0.05 Gb	101141	428
TA137	26,995	0.02 Gb	56152	3476
TA180	31,040	0.02 Gb	103798	4840
TA278	111,662	0.07 Gb	49148	4044
TA289	85,010	0.05 Gb	147902	8996
TA298	210,940	0.13 Gb	119366	105881
TA299	213,985	0.13 Gb	87394	108369
TA300	189,135	0.11 Gb	80134	92733
TA301	218,417	0.13 Gb	115524	106560
TA302	253,602	0.15 Gb	56882	139973
TA303	278,636	0.17 Gb	49620	147708
TA304	176,557	0.11 Gb	40425	85712
TA305	222,729	0.13 Gb	95403	118366
TA306	238,744	0.14 Gb	112492	124741
TA307	219,182	0.13 Gb	81478	113150
TA308	235,418	0.14 Gb	102935	126371
TA309	202,892	0.12 Gb	98711	98658
TA310	117,695	0.07 Gb	133147	56058

TA311	175,504	0.11 Gb	112601	82264
TA312	241,598	0.14 Gb	145267	116971
TA313	195,606	0.12 Gb	109514	90990
TA314	187,599	0.11 Gb	107880	89418
TA315	187,784	0.11 Gb	41451	96595
TA316	231,469	0.14 Gb	181438	107122
TA317	190,705	0.11 Gb	119238	81696
TA318	197,923	0.12 Gb	169264	89219
TA319	222,873	0.13 Gb	119178	102172
TA320	234,512	0.14 Gb	44739	127192
TA321	244,937	0.15 Gb	20940	113259
TA322	225,147	0.14 Gb	123068	118730
TA323	215,381	0.13 Gb	137464	115899
TA324	237,906	0.14 Gb	123762	125940
TA357	187,311	0.11 Gb	65590	104333
TA358	297,085	0.18 Gb	82440	122341
TA359	189,231	0.11 Gb	124794	58303
TA360	99,369	0.06 Gb	109774	41704
TA361	82,308	0.05 Gb	97300	25190
TA364	119,543	0.07 Gb	118035	38343
TA365	188,820	0.11 Gb	54647	86981
TA366	167,274	0.10 Gb	88334	74041
TA367	217,761	0.13 Gb	83397	94547
TA370	178,111	0.11 Gb	85412	90965
TA371	184,977	0.11 Gb	142510	89838
TA374	57,742	0.03 Gb	96882	21285
TA375	9,474	0.01 Gb	111433	5002
TA376	169,679	0.10 Gb	81752	73453
TA377	201,674	0.12 Gb	91991	93617
TA391	194,563	0.12 Gb	92077	89711
TA392	242,717	0.15 Gb	63707	91221
TA393	81,079	0.05 Gb	57582	40419
TA394	50,262	0.03 Gb	46716	22158
TA395	20,333	0.01 Gb	7767	9103
TA396	220,555	0.13 Gb	31901	105911
TA397	264,685	0.16 Gb	78753	127711
TA398	197,877	0.12 Gb	4297	109114
TA400	199,046	0.12 Gb	24366	69875
TA401	74,518	0.04 Gb	57382	37168
TA402	101,807	0.06 Gb	27636	33322
TA404	317,959	0.19 Gb	5212	73660
TA405	232,552	0.14 Gb	47921	74572
TA412	42,793	0.03 Gb	28083	26377
TA413	103,452	0.06 Gb	9515	55358
TA414	182,523	0.11 Gb	52367	89528

TA415	198,092	0.12 Gb	26609	103634
TA416	213,620	0.13 Gb	61780	114211
TA430	201,887	0.12 Gb	168130	60885
TA432	295,181	0.18 Gb	91152	137321
TA433	222,079	0.13 Gb	113360	111870
TA434	90,086	0.05 Gb	15241	43383
TA435	86,715	0.05 Gb	52205	18060
TA438	205,895	0.12 Gb	15231	73058
TA444	236,125	0.14 Gb	57954	83099
TA445	262,306	0.16 Gb	15062	114644
TA446	228,080	0.14 Gb	52897	122002
TA447	243,722	0.15 Gb	13080	118508
Total	24,280,530	14.57 Gb	12697893	10957101

Table S2.4. The percentage pairwise sequence similarity between sequences of two samples (TA307 and TA308) generated by both Sanger and the novel Next generation sequencing pipeline.

Sequence method	**%Pairwise sequence similarity**
CO15_Sanger vs. NextGen	99.8
CO13_Sanger vs. NextGen	99.9
EF_Sanger vs. NextGen	99.6
WNT_Sanger vs. NextGen	99.4

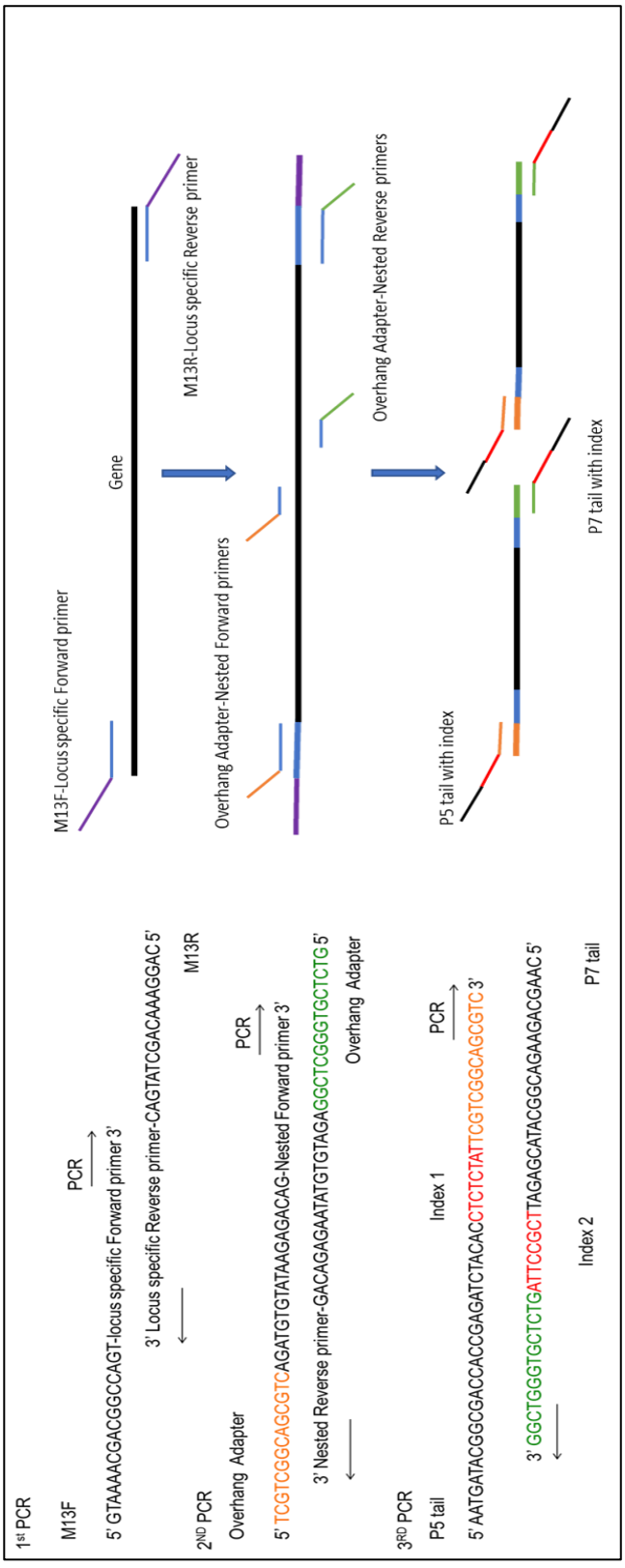

Figure S2.1. A schematic representation of the three-step PCR approach and the primer design used to amplify all the gene regions in this study.

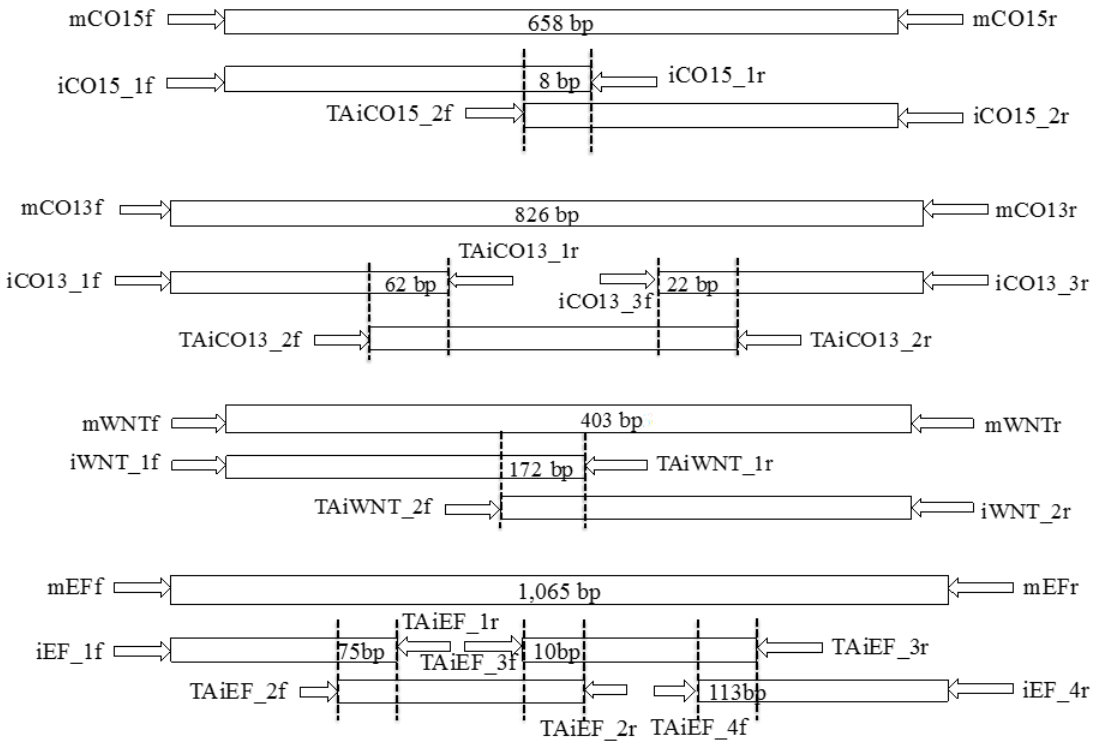

Figure S2.2. A detailed primer map showing the length of overlaps between PCR fragments of the four gene regions amplified. Primers are labelled according to Table 2.1.

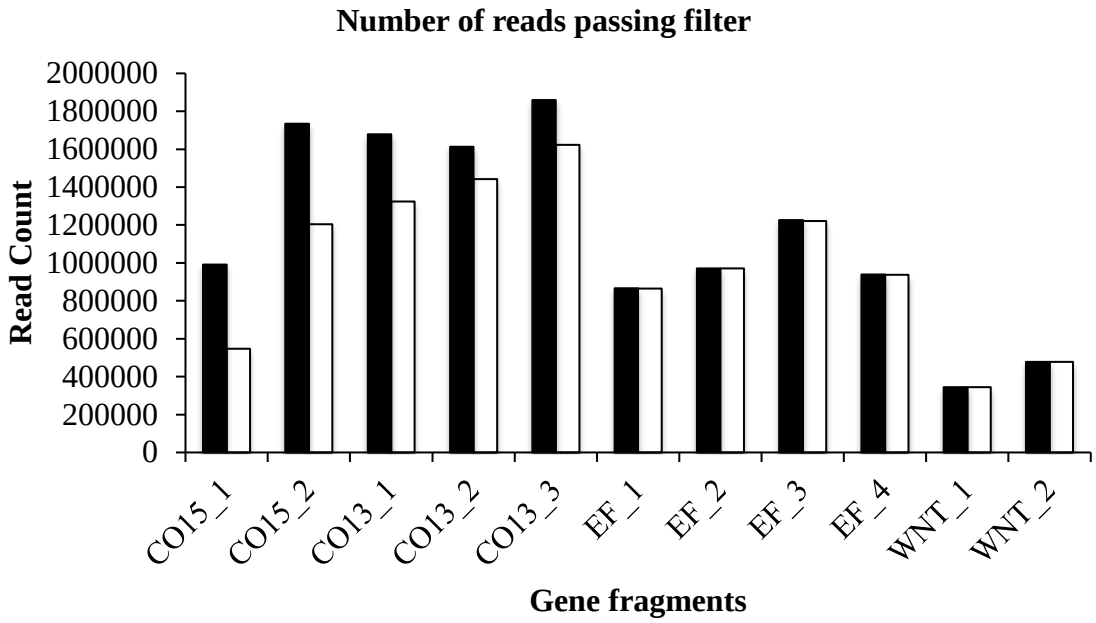

Figure S2.3. The number of reads passing the filter of environmental contaminants for each gene fragment. Black bars: Number of reads passing fastQC; White bars: Number of reads passing the filter of environmental contaminants.

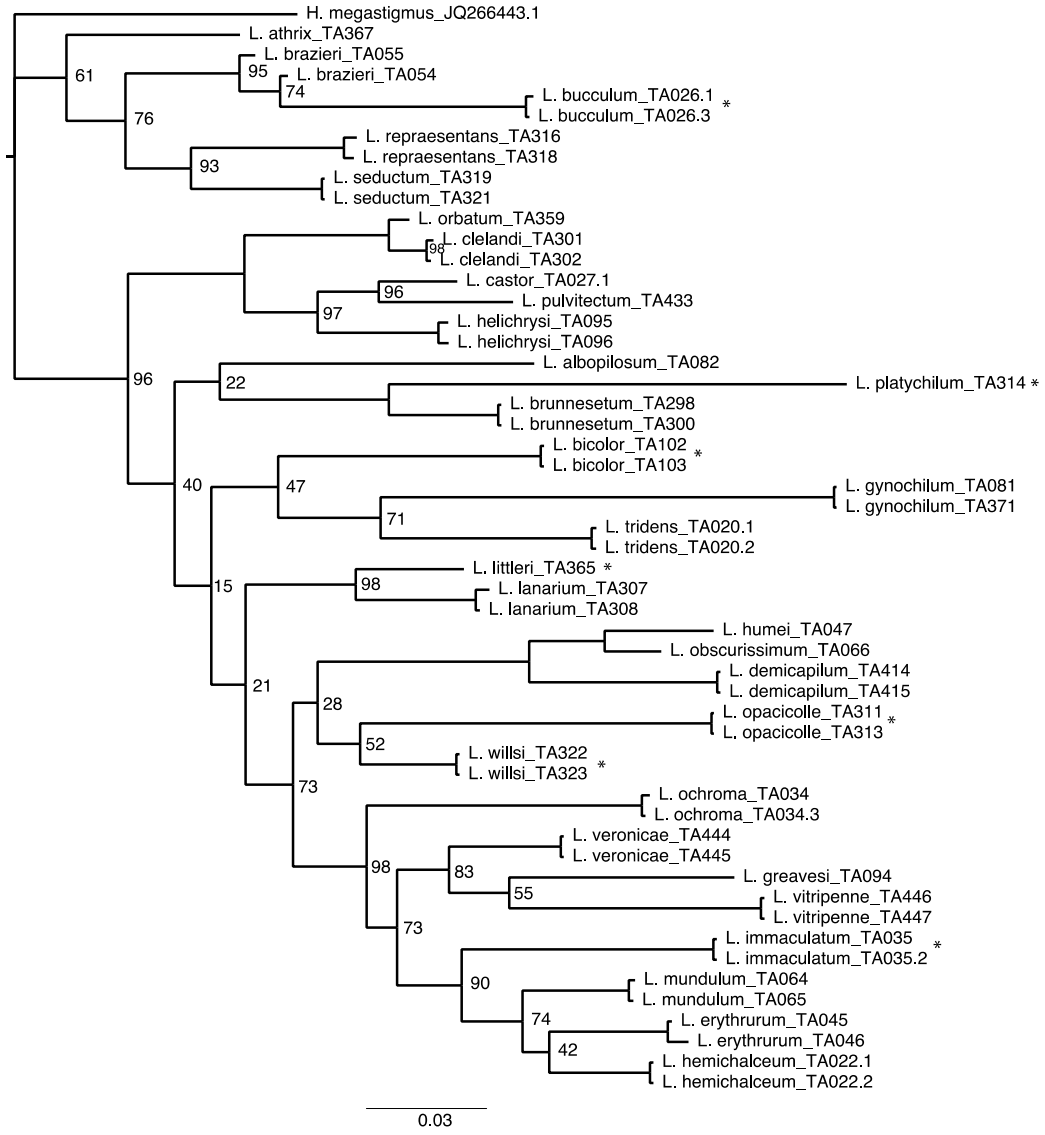

Figure S2.4. A Bayesian consensus phylogenetic tree of *L*. (*Chilalictus*) species constructed using CO1 sequences. Posterior probabilities of < 99% are shown above the nodes. Asterisks mark disagreements in the placement of species between the phylogeny constructed with concatenated gene sequences (Fig. 2.5) and the *CO1* phylogeny.

Appendix 2.1

In house Gentra protocol (modified from the Qiagen - Puregene Gentra Tissue Kit protocol)

Step 1

- Add 300µl of Cell lysis buffer (100mM Tris, 2%SDS, 0.1 M EDTA, PH8) into tube containing tissue (If tissue is in alcohol, let it dry for a few minutes before adding the cell lysis buffer).
- Add 1.5µl of 20mg/ µl Proteinase K, vortex briefly.
- Incubate at 55^0 overnight.

Step 2

- Place samples into ice for 1 min
- Add 100µl of 7.5M Ammonium Acetate (Protein precipitate solution)
- Vortex at a maximum speed of 20 sec
- Incubate on ice for 5 min
- Centrifuge at 18,407 rcf for 5 min and place the tube on ice
- Prepare new screw tube and put the supernatant in it (Use a 1 ml pipette)
- Add 300 µl of 100 % isopropanol (kept at 4^0 C) and 0.8 µl of Glycogen (stored at -20^0 C).
- Mix by inverting 50 times and incubate at -20^0C overnight.

Step 3

- Centrifuge at 13,523 rcf for 20 min. Mark where the pellet should be because sometimes it is not visible.
- Decant the supernatant (save the pellet) and drain tubes on clean absorbent. paper.
- Add 300 µl of 70% Ethanol (kept at 4^0 C) and vortex at low speed

- Centrifuge at 13,523 rcf for 8 mins.
- Pour off Ethanol and let tubes dry for 3 hours or more (a tiny pellet should be visible at the bottom of the tube)
- Add 30µl of TLE (10mM Tris, 0.1mM EDTA, PH8)
- Incubate at room temperature overnight
- Vortex for 10 sec.
- Quick spin.
- Store at -20^0 C.

Statement of Authorship

Title of Paper	The evolution of broadly polylectic behaviour in *Lasioglossum* (*Chilalictus*) – Apoidea Halictidae
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	NA

Principal Author

Name of Principal Author (Candidate)	Trace Akankunda
Contribution to the Paper	Developed the concept, assembled visitation records and sequence data for analysis, conducted 90% of the data analysis, wrote and revised the manuscript
Overall percentage (%)	85%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 9/04/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Katja Hogendoorn
Contribution to the Paper	Contributed to concept development, provided part of the data on visitation records, conducted 10% of the data analysis, revised the manuscript
Signature	Date 4/4/2019

Name of Co-Author	Carlos Rodriguez Lopez
Contribution to the Paper	Contributed to concept development, assisted with data analysis, revised the manuscript
Signature	Date 3/4/2019

Name of Co-Author	Remko Leijs		
Contribution to the Paper	Contributed to concept development, assisted with data analysis, revised the manuscript		
Signature		Date	4/4/2019

Name of Co-Author	Ken Walker		
Contribution to the Paper	Provided some of the data on visitation records and the morphological characters used for constructing a phylogeny, revised the manuscript		
Signature		Date	1/04/2019

Chapter 3: The evolution of broadly polylectic behaviour in *Lasioglossum* (*Chilalictus*) (Halictidae, Apoidea)

Trace Akankunda[1], Carlos Rodriguez Lopez[2], Remko Leijs[3], Ken Walker[4], Katja Hogendoorn[1, 5]

[1]The University of Adelaide, School of Agriculture, Food and Wine Adelaide SA 5005, [2]Environmental Epigenetics and Genetics Group; Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington, KY 40546, [3]South Australian Museum, North Terrace Adelaide SA 5000, [4]Museums Victoria, Melbourne VIC 3001, [5]Corresponding author

Abstract

Based on the number of pollen hosts utilised, bees have been categorised as generalists (polylectic) or specialists (oligolectic). Polylectic bees forage on a wide variety of plant families while oligolectic bees forage on a few plant families and at times a single plant genus or species. Faced with a changing habitat, polylectic bees can diversify their pollen 'portfolio', while oligolectic bees cannot and therefore may go locally extinct. Research into the evolution and maintenance of broad polylecty is scant. Instead, research has mainly focussed on the factors that constrain oligolectic species to a narrow diet. Here, we developed a molecular phylogeny of a native Australian subgenus *Lasioglossum* (*Chilalictus*), (Halictidae), to study the evolution of pollen host breadth within the group. We find that broad polylecty has evolved independently at least four times in *L.* (*Chilalictus*) and did not result in subsequent speciation. Oligolecty has evolved once and is found in at least three related species. In addition, broadly polylectic species have significantly larger areas of occurrence than oligolectic and polylectic species. Taken together, these results suggest that there is less opportunity for speciation in broadly polylectic than in oligolectic species. As broad polylecty is uncommon in bees, we hypothesize the existence of genetic constraints to its evolution. Future studies on the evolution of broad polylecty should examine both the existence of constraints and selective advantages for host broadening in polylectic species.

Key words: diet width, broad polylecty, phylogenetics, evolution

Introduction

The resilience of species depends on the capacity to adapt to environmental change, e.g. to changes in climate, composition of air, soil, shelter and food sources. Adaptation to a changing environment is more likely in those species that have generalist rather than specific requirements (Slatyer et al. 2013; Lavergne et al. 2013; Thuiller et al. 2005). This is particularly the case for bee species, as species that closely rely on a small selection of plants are more likely to be affected by a change or reduction in the availability of floral resources than those that are more versatile in their diet requirements (Burkle et al. 2013; Packer et al. 2005).

Bees have been categorised as generalists or specialists based on the number of pollen hosts utilised (Waser 2006). Generalist or polylectic species utilise multiple, while specialist or oligolectic species utilise a restricted number of plant species as pollen hosts. Further distinctions, using the number of plant families, genera and species visited, include monolecty, narrow oligolecty, eclecty, oligolecty, mesolecty, polylecty and broad polylecty (Cane & Sipes 2006; Table S3.1). Polylectic bees utilise 4 - 25%, and broadly polylectic bees utilise more than 25% of the locally available plant families as pollen hosts (Cane & Sipes 2006). While these definitions do not completely resolve the ambiguity in categorising bees based on their pollen host breadth (Vossler 2018), the classification provides meaningful boundaries in the context of pollen host breadth and conservation, and underscore the existence of broad polylecty in bees.

Pollen host breadth is an important attribute of bee species, as it affects a number of life-history traits. For example, pollen host breadth is strongly correlated with the phenology and longevity of species and is therefore a major determinant of the potential for the evolution of sociality (e.g. Velthuis & Hogendoorn 2017; Danforth 2002). In addition, as mentioned above, pollen host breadth is a predictor of the conservation

status of bee species in the face of environmental changes. Furthermore, from a utilitarian point of view, pollen host breadth determines the suitability of species as crop pollinators (Klein et al. 2007).

In this context, broadly polylectic species are of specific interest. Firstly, this is the most versatile group of pollinators of both crops and native plants (Klein et al. 2007; Hung et al. 2018). With the loss of oligolectic species, pollination of both crops and wild flora is increasingly dependent on broadly polylectic bees, and this expands the significance of broadly polylectic bee species in pollination networks (Burkle et al. 2013). Secondly, the diet choices made by broadly polylectic bees can give an indication of the floral resources available in any given landscape, and hence can provide a method for monitoring and quantifying floral resources (Colwell et al. 2017; Requier et al. 2015; Cane & Sipes 2006). Thirdly, the versatile behavioural, neurological, sensory and digestive systems of broadly polylectic species can be studied to unravel the adaptations used to locate, harvest and digest the wide range of floral resources (Cane & Sipes 2006).

Despite the importance of broadly polylectic species, research into the evolution and maintenance of broadly polylectic behaviour in bees is limited. Instead, research has mainly focussed on the factors that constrain oligolectic species to a narrow diet (Minckley & Roulston 2006; Wcislo & Cane 1996). The current lack of attention for broad polylecty may be a consequence of the fact that broadly polylectic species are often wide spread and reach high local abundance as do many food generalists (Inkinen 1994), and hence do not require our attention from a conservationist point of view. In addition, broad polylecty may seem nothing special, as making use of all pollen supplies available in the landscape makes evolutionary sense. Despite this latter notion, broadly polylectic species appear to be rare across the bee families, with most of the

reported species occurring in two out of seven bee families of Apidae and Halictidae (Vossler 2018; Dalmazzo & Vossler 2015; Cane & Sipes 2006), although there are some broadly polylectic species in the family Megachilidae (Haider et al. 2014). This implies that there may be constraints to the evolution of broad polylecty. The need for attention to the factors involved in the evolution of and constraints to broad polylecty are further underscored by the recent insights that broad polylecty is a derived trait in bees. By contrast, until 1996, oligolecty was mostly thought to be a derived trait, and therefore in need of an explanation (reviewed in 1996 by Waser et al., and by Muller). However, oligolecty has now become well established as the ancestral character state of pollen host breadth in bees (Litman et al. 2011; Larkin et al. 2008; Michez et al. 2008; Patiny et al. 2008; Muller 1996).

Here we studied the occurrence and evolution of broad polylecty in halictine bees. This taxon was chosen because its species represent the full spectrum of pollen host breadth from oligolecty to broad polylecty (Cane and Sipes 2006; Danforth 2001). In Australia, the largest and most diverse halictine genus is *Lasioglossum*, comprising eight subgenera. Of these, the subgenus *Lasioglossum* (*Chilalictus*) has the best-documented floral visitation records (Atlas of Living Australia website at http://www.ala.org.au. Accessed June 2018). Apart from analysing the evolution of broad polylecty, our study also explores potential links between pollen host breadth, species distribution and speciation. Because broadly polylectic species can find sustenance in many types of habitats, it is expected that they have a wider distribution than polylectic and oligolectic species (Slatyer et al. 2013; Slove & Janz 2011). We investigated this by mapping the area of occurrence of known species against flower visitation data. In addition, as a broad diet and a larger area of possible occurrence could lead to larger population sizes, this could reduce the isolation between populations and hence the opportunities for

speciation. The opposite would hold for oligolectic species. In summary, this study explores the phylogenetic ancestry of pollen host breadth in *Lasioglossum* (*Chilalictus*) and its association with speciation and species distribution in Australia. This was done by constructing a molecular phylogeny of *Lasioglossum* (*Chilalictus*) species, mapping their pollen host breadth onto the phylogeny and reconstructing the ancestral states of pollen host breadth for nodes in the phylogeny using a Bayesian inference method (Ronquist 2004). The association between pollen host breadth and speciation and geographical range size was determined by examining the molecular phylogeny and records of the species' area of occupancy in Australia, respectively.

Methods

Phylogenetic Analysis

Taxon sampling and gene selection

Using the single gene-tree performance method (Ai and Kang et al. , 2015), we selected to use three of four gene regions previously used to construct a molecular phylogeny of the subfamily Halictinae, which included representatives of *L.* (*Chilalictus*; Gibbs et al. 2012). The three gene regions selected were eukaryotic translation elongation factor 1 alpha (*EF-1α*), wingless (*wg*) and cytochrome c oxidase subunit 1 (*CO1*). We assembled a sequence dataset for 51 species of *Lasioglossum* (*Chilalictus*) with 14 species from publicly available sequences (Gibbs et al. 2012) and 37 species sequenced using a nested amplicon sequencing approach (Chapter 2). All the major species clades of the *L.* (*Chilalictus*) morphological phylogeny (Walker K, unpublished) were represented in the sequence dataset. Five species were included as an out-group with two *Homalictus* species and three *Lasioglossum* species from the subgenera *Australictus* and *Parasphecodes* (Table 3.1).

Partitions and model selection

The concatenated sequence dataset was split into 10 data blocks by gene, codon and introns and we used PartitionFinder 2 algorithm (Guindon et al. 2010; Lanfear et al. 2017) to select the best partitioning scheme and nucleotide evolutionary models for each partition. Model selection was done using the corrected Akaike Information Criterion metric and a greedy search algorithm (Guindon et al. 2010; Lanfear et al. 2017) with linked branch lengths. Only models that could be implemented in MrBayes (v3.2) were compared.

Table 3.1. Species of *Lasioglossum* (*Chilalictus*) included in the sequence dataset for phylogenetic analysis with GenBank accession numbers. Sequences whose accession numbers start with MH were produced using the amplicon sequencing method developed in chapter 2, while the rest were sourced from GenBank archives.

Taxon	Diet width	**Sample name**	**CO1**	***EF-1 α***	**wg**
L. tridens	Polylectic	TA020.2	MH320179	MH320399	MH320282
L. hemichalceum	Polylectic	TA022.2	MH320181	MH320401	MH320284
L. bucculum	Polylectic	TA026.3	MH320185	MH320405	MH320288
L. castor	Polylectic	TA027.1	MH320186	MH320406	MH320289
L. ochroma	Polylectic	TA034.3	MH320189	MH320409	MH320292
L. immaculatum	Polylectic	TA035.2	MH320191	MH320411	MH320294
L. plebeium	Polylectic	TA037.2	MH320193	MH320413	MH320296
L. humei	Oligolectic	TA047	MH320196	MH320416	MH320299
L. expansifrons	Polylectic	TA050		MH320418	MH320301
L. brazieri	Polylectic	TA055	MH320199	MH320421	MH320305
L. mundulum	Polylectic	TA065	MH320204	MH320425	MH320310
L. obscurissimum	Oligolectic	TA066	MH320205	MH320426	MH320311
L. speculatum	Polylectic	TA074	MH320208	MH320430	MH320315
L. albopilosum	Polylectic	TA082	MH320210	MH320432	MH320318
L. erythrurum	Broadly polylectic	TA087	MH320212	MH320433	MH320320
L. greavesi	Polylectic	TA094	MH320216	MH320438	MH320325
L. helichrysi	Polylectic	TA096	MH320218	MH320440	MH320327
L. bicolor	Polylectic	TA103	MH320222	MH320444	MH320331
L. brunnesetum	Polylectic	TA300	MH320226	MH320450	MH320336
L. lanarium	Broadly polylectic	TA308	MH320234	MH320458	MH320344
L. opacicolle	Polylectic	TA313	MH320239	MH320462	MH320349
L. platychilum	Polylectic	TA315	MH320241	MH320465	MH320351
L. repraesentans	Polylectic	TA318	MH320244	MH320468	MH320354
L. seductum	Polylectic	TA321	MH320247	MH320471	MH320357
L. willsi	Polylectic	TA323	MH320249	MH320473	MH320359
L. orbatum	Polylectic	TA359	MH320252	MH320477	MH320363
L. littleri	Polylectic	TA366	MH320256	MH320480	MH320368
L. athrix	Polylectic	TA367	MH320257	MH320481	MH320369
L. gynochilum	-	TA371	MH320259	MH320483	MH320371
L. cognatum	Broadly polylectic	TA376	MH320260	MH320484	MH320374
L. globosum	Polylectic	TA405	MH320268	MH320504	MH320384
L. demicapillum	Oligolectic	TA415	MH320270	MH320493	MH320386
L. pulvitectum	Polylectic	TA433	MH320273	MH320496	MH320390
L. quadratum	Polylectic	TA434	MH320274	MH320497	MH320391
L. sororculum	Polylectic	TA438	MH320276	MH320499	MH320393

L. veronicae	Polylectic	TA445	MH320278	MH320501	MH320395
L. vitripenne	Polylectic	TA447	MH320280	MH320503	MH320397
L. conspicuum	Polylectic	na	AF103952.1	AF264789.1	JQ266627
Homalictus megastigmus	Polylectic	na	JQ266443.1	AF264839.1	AY222600.1
Homalictus punctatum	Polylectic	na	JQ266444.1	AF264840.1	AY222601.1
L. (Australictus) lithuscum	Polylectic	na	JQ266378.1	AF435372.1	AY222598.1
L. (Parasphecodes) lacthium	Polylectic	na	JQ266460.1	JQ266527.1	JQ266735.1
L. (Australictus) plorator	Polylectic	na	JQ266381.1	JQ266494.1	JQ266622.1
L. baudini	Polylectic	na	JQ266386.1	JQ266497.1	JQ266626.1
L. calophyllae	Polylectic	na	JQ266390.1	JQ266501.1	JQ266632.1
L. chapmani	Broadly polylectic	na	JQ266387.1	JQ266499.1	JQ266630.1
L. clelandi	Broadly polylectic	na	JQ266388.1	JQ266500.1	JQ266631.1
L. convexum	Polylectic	na	AF103951.1	AF264790.1	JQ266625.1
L. fasciatum	Polylectic	na	JQ266391.1	JQ266502.1	JQ266634.1
L. florale	Polylectic	na	AF103955.1	AF264792.1	AY222602.1
L. gilesi	Polylectic	na	JQ266392.1	JQ266503.1	JQ266635.1
L. supralucens	-	na	JQ266402.1	AF264797.1	JQ266644.1
L. tamburinei	-	na	JQ266403.1	JQ266510.1	JQ266645.1
L. mediopolitum	Polylectic	na	AF103957.1	AF264794.1	JQ266636.1
L. mirandum	Mesolectic	na	AF103958.1	AF264795.1	JQ266637.1
L. parasphecodum	-	na	AF103959.1	AF264796.1	JQ266640.1

Phylogenetic methods

Phylogenetic analysis was conducted using MrBayes v3.22 (Ronquist et al. 2012), hosted on the CIPRES Science Gateway web portal (Miller et al. 2010). We performed four independent analyses each comprising 8 simultaneous chain runs for 50,000,000 generations, sampling trees after every 5000 generations. Analyses were performed using an unconstrained branch length prior with an exponential distribution parameter set to 100. All model parameters were unlinked and the rate prior set to be variable across partitions. The topology prior was constrained to have the *L.* (*Chilalictus*) clade as a monophyletic group based on previous research (Gibbs et al. 2012). Analyses were run with and without introns and the dataset trimmed to minimise the effect of missing data on tree topology. After the analyses, parameter files of the two independent runs

per analysis were analysed simultaneously in Tracer v1.6 (Rambaut et al. 2013) to determine if model parameters had stabilised.

Following the MrBayes v3.2 manual, we determined that the tree topologies had reached convergence by making sure that the average standard deviation of split frequencies between the two independent runs for each MrBayes analysis was below 0.01. We also examined the plots of log probabilities per generation for each run after discarding the trees sampled within the burn-in region and observed for a "white noise" pattern in the likelihood estimates, which indicates convergence. We discarded a total of 5,000 trees per analysis to remove all trees generated before model parameters had stabilised. The trees generated from each of the two runs per analysis were summed up. The consensus phylogenetic trees from each of the four analyses were merged in a text editor and combined to form a maximum clade credibility tree using TreeAnnotator v1.10.1 (Suchard et al. 2018). Preview v9.0 (Apple Inc 2016) was used to perform final editing of the phylogenetic tree for purposes of presentation.

Reconstruction of the morphological phylogeny

A morphological data matrix (Data S3.1) comprising 80 characters from 107 *L.* (*Chilalictus*) species was kindly provided by Walker (Museums Victoria). Two taxa from the genus *Ceylalictus* and *Homalictus* were included as out-groups. The data matrix was treated as a single data partition and the phylogenetic analysis run in MrBayes (v3.2) using the standard discrete model with a gamma distribution for estimating rate heterogeneity among sites. Subsequent analysis was conducted as described above and the molecular and morphological phylogenetic trees of *L.* (*Chilalictus*) were compared for congruence.

Analysis of floral visitation records

We compiled 8,486 floral visitation records for 107 *L.* (*Chilalictus*) species by combining records from the PaDIL database at http://www.padiL. gov.au, the database from the Western Australian Museum and a range of recent publications in our own collections. We studied the correlation between the number of visitation records and the number of pollen hosts per species in order to determine the effect of sampling bias in classifying species according to their diet width. Based on the results, we modified the diet width classification system suggested by (Cane & Sipes 2006); Table S3.1, to take the sampling bias into account (Table 3.2).

Table 3.2. A modified diet width classification system based on suggestions by Cane & Sipes (2006), that takes sampling bias into account.

Diet Width	No. of host families	No. of host Genera	Condition
Broadly polylectic	2 * Avg. # Polylecty	2 * Avg. # Polylecty	None
Polylectic	≥ 4	≥ 4	None
Mesolectic	1 - 3	> 4	≥ 30 visitation records
Oligolectic	1	1 - 4	Supported by Literature

Area of occupancy

For the species in the phylogeny, we calculated the area of occupancy in the Australian region using the tools in the Atlas of Living Australia (2019). We chose a grid-size of 2 km (resolution 0.02), to conform to the recommendations by (IUCN 2012). Spatially suspect data were excluded (https://www.ala.org.au/blogs-news/data-processing/). Partial regression analysis was used to test the relationship between the number of plant families visited and the area of occupancy, controlling for the number of observations per species. The association between pollen host breadth and area of occupancy was tested using a general linear model in SPSS.

Mapping diet width onto the molecular phylogeny and ancestral state reconstruction

To map diet width onto the molecular phylogeny, we assigned character states to the classes of diet width as follows; oligolecty = 0, mesolecty = 1, polylecty = 2, and broad polylecty = 3. These character states were then added to the molecular data matrix of *L.* (*Chilalictus*) as a separate character partition in a nexus file with a symbol for missing characters used where the species' diet width was unknown. We reanalysed the molecular phylogeny of *L.* (*Chilalictus*) and simultaneously reconstructed ancestral states for 12 well-supported nodes of interest (with posterior probability ≥ 87). We used the full hierarchical Bayesian approach in MrBayes (v3.2) with a symmetric evolutionary model for the diet width character and without ordering the diet width character states. The analyses were repeated as described above for the construction of the molecular phylogeny but with the topology prior constrained to fix the 12 well-supported nodes in every sampled tree during the analysis, so that their reconstructed ancestral states could be reported. Parameter and log files were analysed for convergence as described above. We analysed the posterior probability of the four

character states at the reconstructed ancestral nodes as reported in the parameter tables of the log files in the four analyses performed. Consensus phylogenetic trees summed up from the analyses were combined to form one maximum clade credibility tree as described above and visualised using FigTree v 1.2.3 (Rambaut, 2016). Preview (v9.0) was used to edit the phylogenetic tree and add posterior probabilities of the ancestral character states for the reconstructed nodes. We repeated the analysis with ordered character states of the diet width and restricted the rate parameters for the evolution of oligolecty to broad polylecty and vice versa to zero as described below by the instantaneous rate matrix (Q). The restriction is justified because host broadening in oligolectic species is often limited to plant species that are related to their primary hosts (Sedivy et al. 2008), making the likelihood of transitioning directly to broad polylecty extremely low. We implemented this analysis in MrBayes using a symmetric three state evolutionary model having merged the mesolectic and oligolectic categories to simplify the model. The mesolectic and oligolectic categories of diet width were not statistically significantly different based on the average number of plant families and genera visited and hence could be merged.

$$Q = \begin{bmatrix} & [a] & [b] & [c] \\ [a] & - & 1 & 0 \\ [b] & 1 & - & 1 \\ [c] & 0 & 1 & - \end{bmatrix}$$

Where a = oligolectic state b = polylectic state, c = broadly polylectic state.

Results

Assembled sequence dataset

The concatenated sequence dataset used for phylogenetic analysis comprised of sequences from 155 specimens which included 63 species and 3,579 characters from three gene regions namely; CO1 (1,495 bp), EF-1 α 2nd copy (1,627 bp), and wg (457 bp). In order to reduce missing data, the sequence dataset was trimmed to 91 specimens representing 56 species and the number of characters reduced to 2,357 including CO1 (826 bp), EF-1 α (1,128 bp) and wg (403 bp).

Bayesian analyses

Based on the AICc metric, the best data-partitioning scheme for the molecular dataset was gene/codon/intron with the 3rd codon position of the wg gene region combined with the 2nd codon position of the EF-1 α (Table S3.2). The best evolutionary models selected for each of the 9 partitions in the molecular dataset are presented in Table S3.2. The molecular phylogeny generated using a Bayesian approach recovered five well-supported clades of species (pp > 85%, marked with an asterisk in Fig. 3.1) but with poorly supported relationships between the clades. The topology of the molecular phylogeny was mostly consistent with the morphological phylogeny, with three out five clades equally well supported by both methods (marked with an asterisk in Fig. 3.1 and Fig. S3.1) and 40 out of the 51 species grouping in similar patterns (highlighted with the same colour in Fig. S3.1. However, the relationships between species within the clades of the morphological phylogeny were not consistent with the molecular phylogeny. In addition, a well-supported species clade in the morphological phylogeny, comprising of *L. mirandum, L. calophyllae, L. tamburinei and L. supralucens* (marked with an arrow in Fig. S3.1), was not recovered in our molecular phylogeny. Instead, our analysis put *L. tamburinei and L. supralucens* outside the main

L. (*Chilalictus*) clade. A complete molecular phylogeny with replicated species (sequences from 91 specimens) is shown in supporting documents Fig. S3.2.

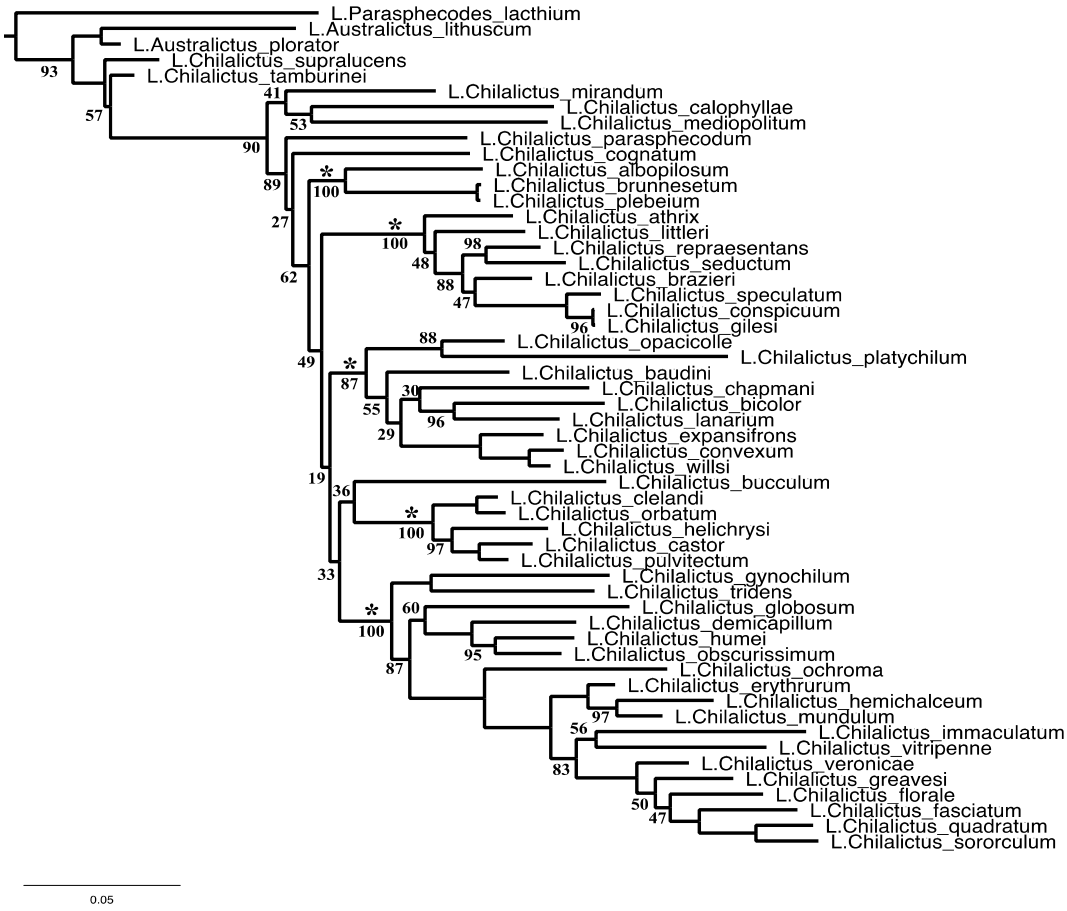

Figure 3.1. A maximum clade credibility tree of the Bayesian phylogenetic analysis of 51 *L.* (*Chilalictus*) species based on 2 nuclear gene regions and 1 mitochondrial gene region. *Homalictus* outgroups not shown to improve resolution. Posterior probabilities of < 99% are shown above the nodes. Species clades that were consistent with morphological results are marked with an asterisk (*) and the posterior probability support of the clade. Species replicates were collapsed to a single randomly chosen specimen per species for ancestral state reconstruction.

Analysis of floral visitation records

There was a significant positive correlation between the log-transformed number of plant families and genera visited and the log-transformed number of floral visitation records per species (r^2 = 0.76, 0.82 respectively, $p < 0.05$; Fig. 3.2). The average number of plant families and genera visited by polylectic species was 12 and 19 respectively, and polylectic species were reclassified under the broadly polylectic category if they visited a minimum of 24 plant families and 38 plant genera (Table S3.3). While *L. humei* was caught on 5 plant families and 6 plant genera (Table S3.3) and would as such classify as polylectic, 90% of the visitations occurred on two plant families and genera and hence it was reclassified into the oligolectic clade, consistent with the classification suggested by Walker (1995) for this species. Table 3.3 is a summary of the categories of diet width determined from floral visitation records of 77 *L.* (*Chilalictus*) species. Based on the Welch Two Sample t-test, there was a significant difference in the mean number of plant families and genera visited by the broadly polylectic and polylectic species ($p < 0.01$) and the polylectic and mesolectic species ($p < 0.01$).

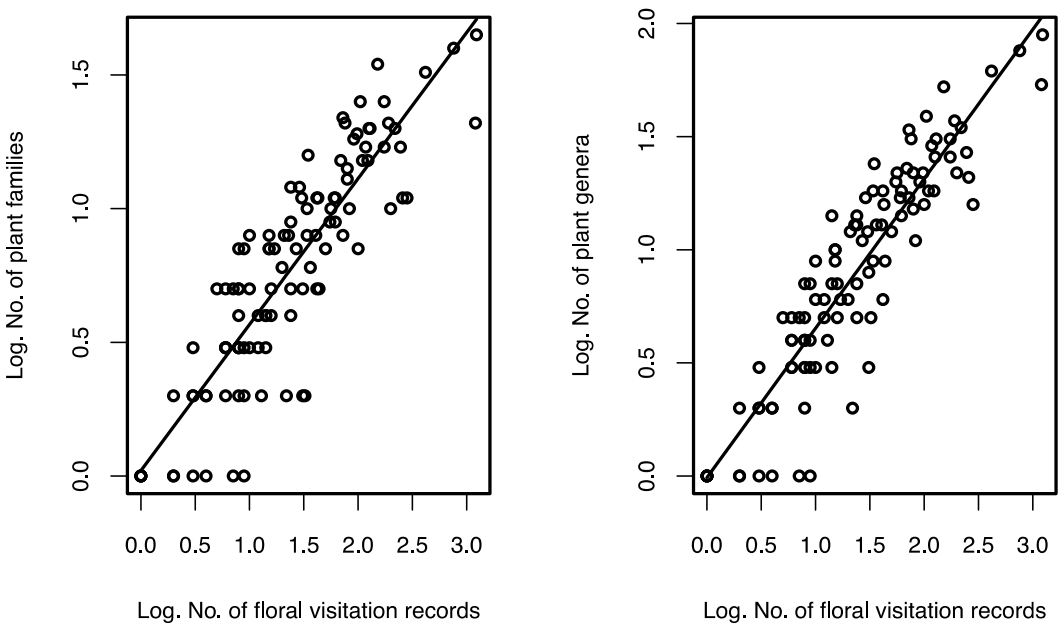

Figure 3.2. Correlation between the log-transformed number of plant families and genera visited against the log-transformed number of floral visitation records for 107 *L.* (*Chilalictus*) species. Circles represent visitation records for individual species.

Table 3.3. Summary of the categories of diet width for 77 *Lasioglossum* (*Chilalictus*) species.

Diet width category	**Number of Species**	**Average number of plant families**	**Average number of plant genera**
Broadly polylectic	6	33	62
Polylectic	66	10	15
Mesolectic	1	3	6
Oligolectic	4	2	3

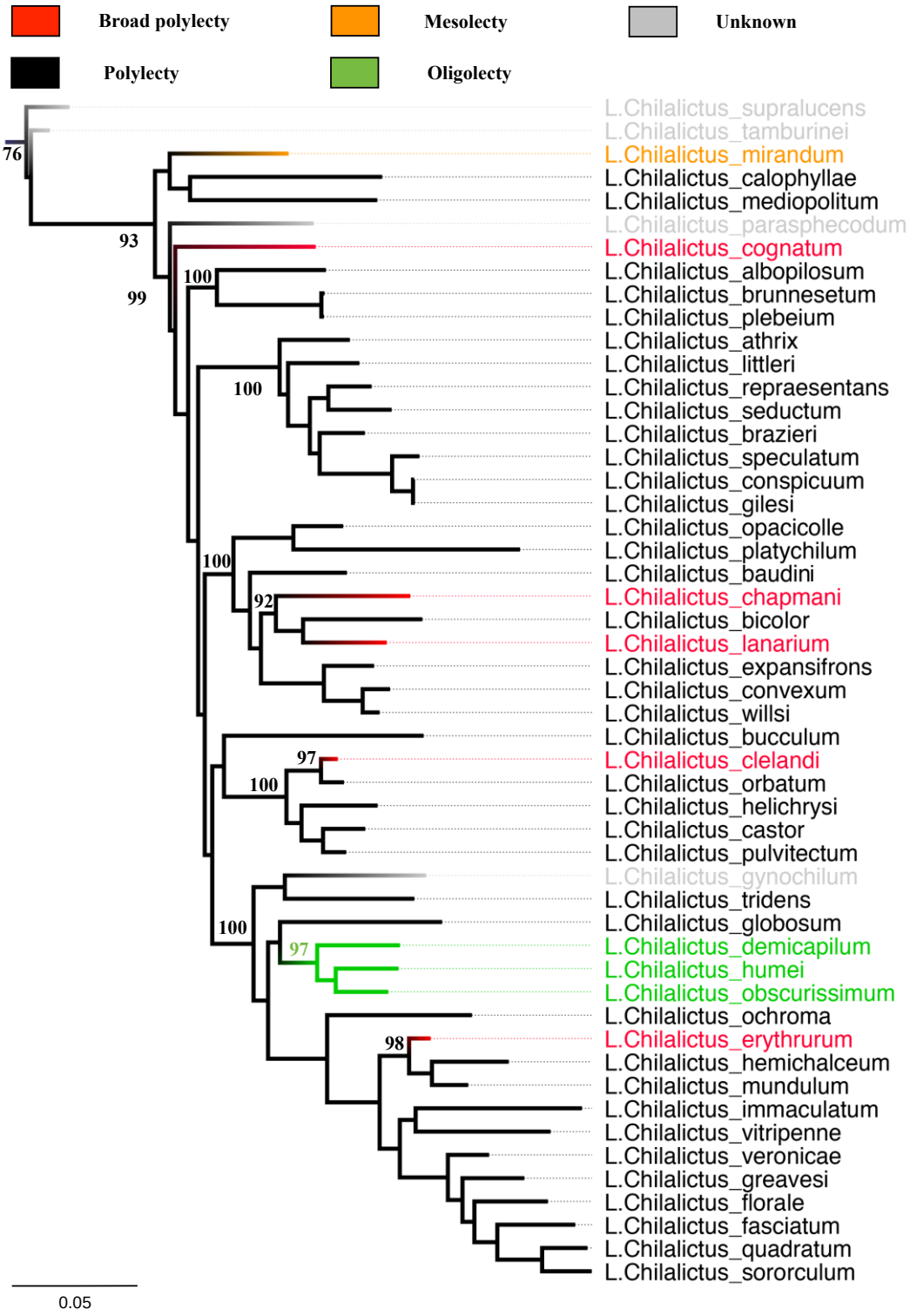
Broad polylecty
Mesolecty
Unknown
Polylecty
Oligolecty
L.Chilalictus_supralucens
L.Chilalictus_tamburinei
L.Chilalictus_mirandum
L.Chilalictus_calophyllae
L.Chilalictus_mediopolitum
L.Chilalictus_parasphecodum
L.Chilalictus_cognatum
L.Chilalictus_albopilosum
L.Chilalictus_brunnesetum
L.Chilalictus_plebeium
L.Chilalictus_athrix
L.Chilalictus_littleri
L.Chilalictus_repraesentans
L.Chilalictus_seductum
L.Chilalictus_brazieri
L.Chilalictus_speculatum
L.Chilalictus_conspicuum
L.Chilalictus_gilesi
L.Chilalictus_opacicolle
L.Chilalictus_platychilum
L.Chilalictus_baudini
L.Chilalictus_chapmani
L.Chilalictus_bicolor
L.Chilalictus_lanarium
L.Chilalictus_expansifrons
L.Chilalictus_convexum
L.Chilalictus_willsi
L.Chilalictus_bucculum
L.Chilalictus_clelandi
L.Chilalictus_orbatum
L.Chilalictus_helichrysi
L.Chilalictus_castor
L.Chilalictus_pulvitectum
L.Chilalictus_gynochilum
L.Chilalictus_tridens
L.Chilalictus_globosum
L.Chilalictus_demicapilum
L.Chilalictus_humei
L.Chilalictus_obscurissimum
L.Chilalictus_ochroma
L.Chilalictus_erythrurum
L.Chilalictus_hemichalceum
L.Chilalictus_mundulum
L.Chilalictus_immaculatum
L.Chilalictus_vitripenne
L.Chilalictus_veronicae
L.Chilalictus_greavesi
L.Chilalictus_florale
L.Chilalictus_fasciatum
L.Chilalictus_quadratum
L.Chilalictus_sororculum
76
93
100
99
100
100
92
97
100
100
97
98
0.05

Figure 3.3. Diet width of *L.* (*Chilalictus*) species mapped onto a molecular phylogeny with reconstructed ancestral character states for 12 well-supported nodes. Species are coloured by the category of their diet width as shown in the key. Reconstructed nodes are labelled by the colour of the ancestral states as shown by the key and support for the reconstructed ancestral state is indicated by the posterior probabilities.

Mapping diet width onto the molecular phylogeny and ancestral state reconstruction

The most likely ancestral state for the subgenus of *L.* (*Chilalictus*) was polylecty (Fig. 3.3). In addition, polylecty was also the best supported ancestral state for all internal ancestral nodes, except for the oligolectic clade, which contained *L. humei*, *L. obscurissimum* and *L. demicapillum*. There were at least four independent origins of broad polylecty i.e in *L. cognatum*, *L. clelandi*, *L. erythrurum* and at least one in the clade containing *L. chapmani* and *L. lanarium*, with broadly polylectic behaviour evolving from a polylectic ancestor in each case (Fig. 3.3). The ancestral states of two nodes giving rise to broadly polylectic species *L. cognatum* and *L. chapmani* were also most likely to be polylectic (pp = 99%, 85% respectively, results not shown) but could not be confirmed because the nodes are weakly supported (Fig. S3.2). The results from ancestral state reconstruction run without any evolutionary restrictions imposed (Data S3.2) were not different from the analyses repeated with transitions from oligolecty to broad polylecty restricted to zero (Data S3.3). In addition, broad polylecty evolved in terminal branches and did not lead to further speciation. Speciation did take place after oligolecty evolved.

Diet width and area of occupancy

There was a strong positive correlation between pollen host breadth and area of occupancy ($r^2 = 0.74$, $P < 0.001$; Fig. 3.4), which remained significant after controlling for the number of host records per species ($r^2 = 0.46$, $P < 0.001$). Controlling for the number of host records, there was a significant effect of pollen host breadth on area of occupation ($F = 9.93$, $P < 0.001$) and a significant interaction between pollen host breadth and number of observations ($F = 7.73$, $p = 0.001$). Broadly polylectic species had a higher area of occupancy than both polylectic and oligolectic species (Tukey's tests: $P < 0.001$), but oligolectic and polylectic species did not differ significantly in their area of occupancy (Tukey's test; $p = 0.99$).

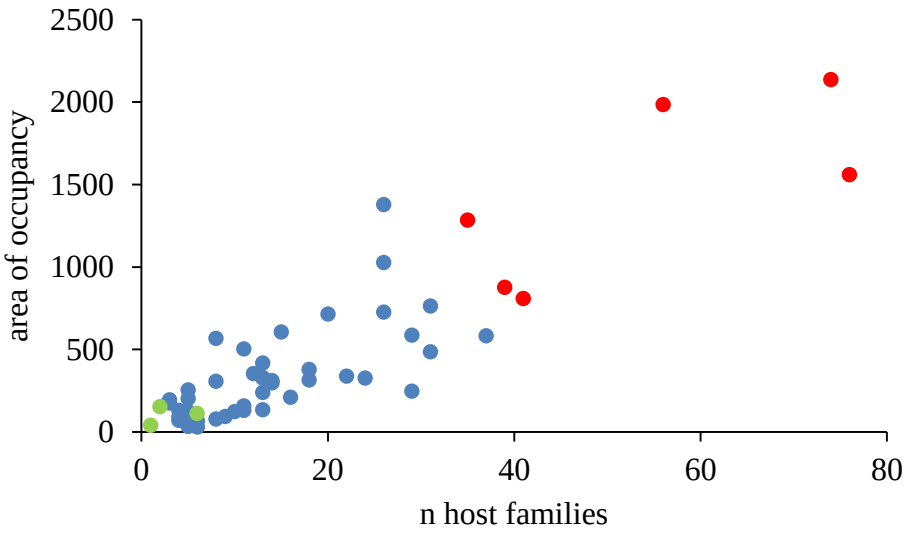

Figure 3.4. The number of host families visited by *L.* (*Chilalictus*) species plotted against the area of occupancy. red = broadly polylectic species, blue = polylectic species, green = oligolectic species.

Discussion

There are at least four well supported independent origins of broad polylecty among the 51 species of *Lasioglossum* (*Chilalictus*) included in this study, namely *L. cognatum*, *L. clelandi*, *L. erythrurum* and at least once in the clade containing *L. chapmani* and *L. lanarium*. In contrast, there was a single origin of oligolectic behaviour for three species associated with the plant genus *Wahlenbergia*. Both broad polylecty and oligolecty evolved from polylectic ancestors, which was the ancestral state of the clade.

The classification of species into categories of diet width was based on floral visitation records. The floral visitation records used here do not accurately represent the pollen hosts of bees because some of the plants could have been nectar rather than pollen hosts (Cane & Sipes 2006). A more accurate approach would be to identify pollen carried on the body of female bees or in their nest provisions (Cane & Sipes 2006; Muller 1996). However, this approach is limited by cost in time and money and therefore the use of floral visitation records remains the most accessible method of determining the diet width of pollinators (Cane & Sipes 2006). Another challenge of using floral visitation records, is that the observer may be biased to capturing records for the most commonly occurring and attractive species within a given landscape, which are usually polylectic or broadly polylectic bees (Minckley & Roulston 2006; Waser et al. 1996). This might explain the strong positive correlation between the number of plant families and genera visited and the number of floral visitation records of species in our database. Therefore, to avoid erroneously classifying a species as oligolectic because of meagre visitation records, we relied on expert literature (Walker 1995) for the classification of species as oligolectic. Floral visitation records of bees suggest that monolecty is rare in most bee families (Minckley & Roulston 2006; Waser et al. 1996), including in the Halictidae

(Lando et al. 2018;Danforth & Ji 2001). In addition broad polylecty might be rare in bees because it has been mostly observed in two of seven bee families, namely Apidae and Halictidae (Vossler 2018; Dalmazzo & Vossler 2015; Cane & Sipes 2006). With six broadly polylectic species and four oligolectic species, among 66 polylectic species, our results support this suggestion.

The molecular phylogeny recovered well-supported clades of species and the topology was mostly consistent with a previous molecular phylogeny that included 24 species (Gibbs et al. 2012) and a morphological phylogeny including 107 species of *L.* (*Chilalictus*; Walker –unpublished; Fig. S3.1). However, the relationships among clades, as well as relationships of species within clades remain largely unresolved as evidenced by the poor posterior support of the backbone and nodes within well-supported clades. In addition, there were several inconsistencies in the placement of species between the three phylogenies. For example, the strength of posterior support and composition of the species clade comprising *L. supralucens, L. tamburinei, L. mirandum, L. calophyllae* and *L. mediopolitum* varied considerably between the three phylogenies, with the first two species placed outside the major clade of *L.* (*Chilalictus*) in our results but occurring in well supported clades of *L.* (*Chilalictus*) in the other two phylogenies (Gibbs et al. 2012, Fig. S3.1). Furthermore, the species clade composed of *L. baudini, L. chapmani, L. bicolor, L. opacicolle, L. convexum, L. willsi* and *L. expansifrons* was well supported in both molecular phylogenies but split into two poorly supported clades within the morphological phylogeny.

The type and number of characters and the density of taxon sampling used in phylogenetic analysis could explain some of the inconsistencies observed between the three phylogenies analysed. Molecular datasets provide stronger posterior support and resolution for phylogenetic relationships than morphological datasets, in part because

they provide more phylogenetically informative characters for analysis (Wortley & Scotland 2006). In addition, an increase in the number of characters and density of taxon sampling, can increase the posterior support and taxonomic resolution of phylogenies by improving the accuracy of evolutionary model parameters estimated and reducing the effect of conflicting phylogenetic signals from different loci (Nabhan & Sarkar 2012; Heath et al. 2008; Hedtke et al. 2006). Indeed, the inclusion of two extra introns of the Elongation factor 1 alpha gene and the Long wavelength rhodopsin gene by Gibbs et al., (2012), improved the strength of the posterior support for nodes within their phylogeny as compared to ours. The Long wavelength rhodopsin gene was not used in our study because it has multiple copies in bees (Spaethe & Briscoe, 2004) and there were no copy specific primers available to amplify it in *L*. (*Chilalictus*). It was therefore dropped to avoid generating chimeras by amplifying, sequencing and combining sequences from its different copies. These results suggest that future studies will need to increase both the number of genes and species sampled in order to obtain more accurate phylogenetic relationships of species in *L*. (*Chilalictus*).

The reconstruction of the ancestral character states for 12 well-supported ancestral nodes of interest support polylecty as the ancestral state for *L*. (*Chilalictus*) with both broad polylecty and oligolecty as derived states. While oligolecty is currently the most accepted ancestral state of the bees in general (Danforth et al. 2006; Danforth et al. 2013), reversals from polylectic to oligolectic behaviour are commonly observed in Halictidae (Danforth et al. 2003; Wcislo & Cane 1996). Such reversals could be driven by the presence of a stable and highly abundant floral resource that occurs all year round (Wcislo & Cane 1996; Minckley & Roulston 2006; Sedivy et al. 2008). In support of this, *Wahlenbergia,* the plant genus associated with our oligolectic clade (Walker 1995), has a high spatial and temporal abundance in Australia (Atlas of Living

Australia website at http://www.ala.org.au. Accessed June 2018), and could therefore locally provide a stable pollen source for oligolectic species. The specialisation within a single clade supports the hypothesis that these oligolectic species might be genetically constrained to their narrow diet (Sedivy et al. 2008). This hypothesis is supported by several experimental studies that have reported evidence for neurological and digestive physiological constraints in oligolectic bee species (Praz et al. 2008a; Praz et al. 2008b) and other herbivorous insects including beetles (Futuyma et al. 1995) and butterflies (Janz et al. 2001).

We propose that there may be similar constraints to the evolution of polylecty to broadly polylectic behaviour, which could explain why broadly polylectic behaviour is rare in *L.* (*Chilalictus*). According to our molecular phylogeny, broadly polylectic behaviour has evolved by convergent evolution in at least four species and in each case from a polylectic ancestor. While stochastic processes can drive the convergent evolution of traits in unrelated species (Stayton 2015), there is evidence to suggest that the loss of stable pollen resources could select for bee species capable of foraging on multiple pollen hosts overtime (Burkle et al. 2013). Under such selection pressures, bee species expand their diet width by incorporating pollen hosts that they can exploit without further adaptations (Minckley & Roulston 2006; Sedivy et al. 2008). However, the capacity to exploit unrelated pollen hosts might necessitate genetic adaptations (Futuyma et al. 1995; Sedivy et al. 2008), in the absence of which, broadly polylectic behaviour might not evolve in a species. The finding that broadly polylectic species have much wider distributions than polylectic and oligolectic species suggest that the former are able to find food in places where their congeners with a more narrow diet fail to establish.

Our results also lend support to the hypothesis that diet width is positively correlated with geographical range size (Slatyer et al. 2013; Slove & Janz 2011). Moreover, a species geographical range size is positively correlated with its population size (Gaston & Lawton 1988). Therefore, unlike broadly polylectic species, oligolectic species are likely to live in small and isolated populations with low genetic variation (Zayed et al. 2006; Packer et al. 2005). This makes them vulnerable to extinction due to stochastic or genetic factors (Slatyer et al. 2013; Zayed et al. 2006; Packer et al. 2005). Our results suggest that this might be the case with the oligolectic species of *L.* (*Chilalictus*) and therefore urgent steps need to be taken to protect them from becoming extinct.

While the dataset is small, we find that speciation had taken place in the oligolectic and polylectic lineages, but not in the four broadly polylectic lineages. In addition, broadly polylectic species occupy a significantly larger area than both oligolectic and polylectic species. We interpret this finding as follows: A wide geographic distribution can facilitate speciation by exposing subpopulations of a species to diversifying selection pressures that might foster local adaptation and subsequent population fragmentation and isolation (Slove & Janz 2011; Janz & Nylin 2008; Thompson 1994). However, due to their large diet width, a wide geographical range size and high local abundance, broadly polylectic species are able to maintain sufficient gene flow and avoid isolation between subpopulations of species (Hardy & Otto 2014). In contrast, both polylectic and oligolectic species can exhibit strong but not necessarily exclusive preferences for specific pollen hosts (Haider et al. 2014; Danforth & Ji 2001; Muller 1996; Tasei & Picart 1973). Additionally, in comparison with broadly polylectic species, polylectic and oligolectic species are more likely to occur with a narrower geographic distribution and achieve a lower local abundance. Therefore it is plausible that sub populations of polylectic and oligolectic species might have become isolated and thus facilitated the

process of speciation in *L.* (*Chilalictus*). Support for this hypothesis exists in the mechanisms suggested for speciation in herbivorous insects, where host specialisation is thought to enable reproductive isolation and subsequently lead to sympatric, parapatric or allopatric speciation (Janz & Nylin 2008; Drès & Mallet 2002). A similar mechanism of speciation via host specialisation has been previously suggested to explain the diversity of species in xeric habitats (Danforth et al. 2003). However, we note that ecological speciation via host specialisation has been contested (Hardy & Otto 2014) and that our study lacked a sufficient sample size of species within different pollen host breadth categories to statistically test the association between pollen host breadth and speciation in *L.* (*Chilalictus*). Therefore, further studies will need to be conducted to determine the specific mechanism of speciation in *L.* (*Chilalictus*).

To further explore the constraints involved in the evolution of broad polylecty, we propose that future studies should examine both the existence of genetic constraints and selective advantages for pollen host broadening in polylectic species. Such studies would require studying closely related species that differ in pollen host breadth and we recommend *L.* (*Chilalictus*) as one such a taxonomic group. Results from such studies would improve our understanding of the effect of genetic constraints on evolution of pollen host breadth in bees. These studies would also have implications for the global bee conservation efforts since they could provide insight into the selection pressures that act against broadening pollen host breadth and hence increase species extinction risk. This would facilitate targeted conservation of the most vulnerable bee species.

References

Atlas of Living Australia at https://www.ala.org.au/. Accessed 19 January 2019.

Burkle, L. A., Marlin, J.C. & Knight, T.M., 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science*, 339(6127), pp. 1611–1615.

Cane, J.H. & Sipes, S., 2006. Characterizing floral specialization by bees: analytical methods and a revised lexicon for oligolecty. In N. M. Waser, Ollerton Jeff, ed. *Plant-pollinator interactions: from specialization to generalization*. Chicago University of Chicago Press, pp. 99–122.

Colwell, M.J. et al. , 2017. Honey bee-collected pollen in agro-ecosystems reveals diet diversity, diet quality, and pesticide exposure. *Ecology and evolution*, 7(18), pp. 7243–7253.

Dalmazzo, M. & Vossler, F.G., 2015. Pollen host selection by a broadly polylectic halictid bee in relation to resource availability. *Arthropod-plant interactions*, 9(3), pp. 253–262.

Danforth, B.N., 2002. Evolution of sociality in a primitively eusocial lineage of bees. *Proceedings of the National Academy of Sciences of the United States of America*, 99(1), pp. 286–290.

Danforth, B.N. et al. , 2006. The history of early bee diversification based on five genes plus morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 103(41), pp. 15118–15123.

Danforth, B.N. et al. , 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual Review of Entomology*, 58(1), pp. 57–78.

Danforth, B.N., Conway, L. & Ji, S., 2003. Phylogeny of eusocial *Lasioglossum* reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). *Systematic Biology*, 52(1), pp. 23–36.

Danforth, B. N., & Ji, S., 2001. Australian *Lasioglossum* + *Homalictus* form a monophyletic group: resolving the "Australian enigma". *Systematic Biology*, 50(2), pp. 268–283.

Drès, M. & Mallet, J., 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 357(1420), pp. 471–492.

Futuyma, D.J., Keese, M.C. & Funk, D.J., 1995. Genetic constraints on macroevolution: the evolution of host affiliation in the leaf beetle genus *Ophraella*. *Evolution*, 49(5),p. 797.

Gaston, K.J. & Lawton, J.H., 1988. Patterns in the distribution and abundance of insect populations. *Nature*, 331(6158), pp. 709–712.

Gibbs, J. et al. , 2012. Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Molecular Phylogenetics and Evolution*, 65(3), pp. 926–939.

Guindon, S. et al. , 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), pp. 307–321.

Haider, M. et al. , 2014. Phylogeny and floral hosts of a predominantly pollen generalist group of mason bees (Megachilidae: Osmiini). *Biological journal of the Linnean Society. Linnean Society of London*, 111(1), pp. 78–91.

Hardy, N.B. & Otto, S.P., 2014. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proceedings. Biological Sciences / the Royal Society*, 281(1795).

Heath, T.A., Hedtke, S.M. & Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46 (3), pp. 239–257.

Hedtke, S.M., Townsend, T.M. & Hillis, D.M., 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology*, 55(3), pp. 522–529.

Hung, K.L.J. et al. , 2018. The worldwide importance of honey bees as pollinators in natural habitats. *Proceedings of the Royal Society: Biological Sciences*, 10; 285, 20172140.

Inkinen, P., 1994. Distribution and abundance in British noctuid moths revisited. *Annales Zoologici Fennici*, 31(2), pp. 235–243.

IUCN, 2012. *IUCN Red List Categories and Criteria* 2nd ed., Gland, Switzerland and Cambridge, UK: IUCN.

Janz, N., Nyblom, K. & Nylin, S., 2001. Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. *Evolution*, 55(4), pp. 783–796.

Janz, N. & Nylin, S., 2008. The oscillation hypothesis of host-plant range and speciation. In K. Tilmon, ed. *Specialization, speciation, and radiation. The evolutionary biology of herbivorous insects*. University California Press, pp. 203–215.

Kéry, M., Matthies, D. & Fischer, M., 2001. The effect of plant population size on the interactions between the rare plant *Gentiana cruciata* and its specialized herbivore *Maculinea rebeli*. *Journal of Ecology*, 89(3), pp. 418–427.

Klein, A.M. et al. , 2007. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society*: *Biological Sciences*, 274(1608), pp. 303–313.

Lando, F. et al. , 2018. Bee diversity and *Solanum didymum* (Solanaceae) flower–visitor network in an atlantic forest fragment in southern Brazil. *Diversity*, 10(1),p. 3.

Lanfear, R. et al. , 2017. Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), pp. 772–773.

Larkin, L. L. , Neff, J.L. & Simpson, B.B., 2008. The evolution of a pollen diet: Host choice and diet breadth of *Andrena* bees (Hymenoptera: Andrenidae). *Apidologie*, 39(1), pp. 133–145.

Lavergne, S. et al. , 2013. Are species' responses to global change predicted by past niche evolution? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1610),p. 20120091.

Litman, J.R. et al. , 2011. Why do leafcutter bees cut leaves? New insights into the early evolution of bees. *Proceedings. Biological Sciences / the Royal Society*, 278(1724), pp. 3593–3600.

Michez, D. et al. , 2008. Phylogeny and host-plant evolution in Melittidaes.*L*. (Hymenoptera: Apoidea). *Apidologie*, 39(1), pp. 146–162.

Miller, M.A., Pfeiffer, W. & Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*. 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8.

Minckley, R.L. & Roulston, T.H., 2006. Incidental mutualism and pollen specialization among bees. Plant-pollinator interactions. In N. M. Waser, Ollerton Jeff, ed. *From specialization to generalization*. Chicago: The University of Chicago Press, pp. 69–98.

Muller, A., 1996. Host-plant specialization in western palearctic anthidine bees (Hymenoptera: Apoidea: Megachilidae). *Ecological monographs*, 66(2), pp. 235–257.

Nabhan, A.R. & Sarkar, I.N., 2012. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Briefings in Bioinformatics*, 13(1), pp.122–134.

Packer, L. et al. , 2005. Conservation genetics of potentially endangered mutualisms: reduced levels of genetic variation in specialist versus generalist bees. *Conservation Biology*, 19(1), pp. 195–202.

Patiny, S., Michez, D. & Danforth, B. N., 2008. Phylogenetic relationships and host-plant evolution within the basal clade of Halictidae (Hymenoptera, Apoidea). *Cladistics : The international journal of the Willi Hennig Society*, 24(3), pp. 255–269.

Praz, C.J., Müller, A. & Dorn, S., 2008a. Host recognition in a pollen-specialist bee: evidence for a genetic basis. *Apidologie*, 39(5), pp. 547–557.

Praz, C.J., Müller, A. & Dorn, S., 2008b. Specialized bees fail to develop on non-host pollen: do plants chemically protect their pollen? *Ecology*, 89(3), pp. 795–804.

Rambaut, A. et al. , 2013. *Tracer*, Institute of Evolutionary Biology: University of Edinburgh.

Requier, F. et al. , 2015. Honey bee diet in intensive farmland habitats reveals an unexpectedly high flower richness and a major role of weeds. *Ecological Applications*, 25(4), pp. 881–890.

Ronquist, F. et al. , 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), pp. 539–542.

Ronquist, F., 2004. Bayesian inference of character evolution. *Trends in Ecology & Evolution*, 19(9), pp. 475–481.

Sedivy, C. et al. , 2008. Patterns of host-plant choice in bees of the genus *Chelostoma*: the constraint hypothesis of host-range evolution in bees. *Evolution*, 62(10), pp. 2487–2507.

Spaethe, J., & Briscoe, A. D. (2004). Early duplication and functional diversification of the opsin gene family in insects. *Molecular Biology and Evolution*, *21*(8), 1583–1594. doi:10.1093/molbev/msh162

Slove, J. & Janz, N., 2011. The relationship between diet breadth and geographic range size in the butterfly subfamily Nymphalinae-a study of global scale. *Plos One*, 6(1), p.e16057.

Stayton, C.T., 2015. What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface*

focus, 5(6),p. 20150039.

Suchard, M.A. et al. , 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus evolution*, 4(1), p.vey016.

Tasei, J.N. & Picart, M., 1973. Le comportement de nidification chez *Osmia* (*Osmia*) *cornuta* Latr. et *Osmia* (*Osmia*) *rufa* L. (Hymenoptera Megachilidae). *Apidologie*, 4, pp.195 –225 .

Thompson, J.N., 1994. Diversifying coevolution. In *The coevolutionary process*. Chicago, USA: The University of Chicago press, pp. 239–252.

Thuiller, W., Lavorel, S. & Araujo, M.B., 2005. Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, 14(4), pp. 347–357.

Velthuis, H.H.W. & Hogendoorn, K., 2017. The various origins of social behaviour in bees. *Bee World*, 94(1), pp. 2–7.

Vossler, F.G., 2018. Are stingless bees a broadly polylectic group? an empirical study of the adjustments required for an improved assessment of pollen diet in bees. In P. Vit, S. R. M. Pedro, & D. W. Roubik, eds. *Pot-Pollen in Stingless Bee Melittology*. Cham: Springer International Publishing, pp. 17–28.

Walker, K.L., 1995. Revision of the Australian native bee subgenus *Lasioglossum* (*Chilalictus*). *Memoirs of Museum Victoria*, 55(1 & 2), pp. 1–423.

Waser, N.M., Chittka, L., Price, M.V., Williams, N.M. and Ollerton, J., 1996. Generalization in pollination systems and why it matters. *Ecology*, 77(4),p. 1043 -1060.

Waser, N.M., 2006. Specialisation and generalisation in plant-pollinator interraction:

A historical perspective Plant-Pollinator interructions. In N. M. Waser, Ollerton Jeff, ed. *From Specialisation to Generalisation*. Chicago: University of Chicago Press, pp. 3–17.

Wcislo, W.T. & Cane, J.H., 1996. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. *Annual Review of Entomology*, 41, pp. 257–286.

Wortley, A. & Scotland, R., 2006. The Effect of Combining Molecular and Morphological Data in Published Phylogenetic Analyses. *Systematic Biology*, 55(4), pp.677–685.

Zayed, A. et al. , 2006. Increased genetic differentiation in a specialist versus a generalist bee: implications for conservation. *Conservation genetics*, 6(6), pp. 1017–1026.

Supporting information

Table S3.1. Lexicon and criteria for classifying species according to their diet width as suggested by Cane and Sipes (2006).

Class of specialisation	Number of plant taxa		
	Species	Genera	Tribes or families
Broad Polylecty	Many	Many	> 25% of available families
Polylecty	Many	Many	4 to < 25% of available families
Mesolecty	Many	> 4	1 - 3 families or big tribes
Oligolecty	> 1	1 - 4	1
- Eclectic	> 1	2 - 4	2 - 3
- Narrow	> 1	1	1
Monolecty	1	1	1

Table S3.2. The best evolutionary models selected for 9 partitions of the sequence dataset based on the corrected Akaike Information Criterion metric in PartitionFinder 2. CO1=Cytochrome Oxidase 1, EF=Elongation Factor 1 alpha (2nd copy), WNT = Wingless gene (wnt-1).

Partitions	AICc selected models
1stpos CO1	GTR + Γ
2ndpos CO1	GTR + I + Γ
3rdpos CO1	GTR + I + Γ
1stpos EF Exon	SYM + I + Γ
2ndpos EF Exon, 3rdpos WNT	GTR + I + Γ
3rdpos EF Exon	HKY + I + Γ
1stpos WNT	GTR + Γ
2ndpos WNT	SYM + I
EF Introns	GTR + I + Γ

Table S3.3. The number of visitation records, plant families and genera of 107 *Lasioglossum* (*Chilalictus*) species and their inferred diet width category based on our classification system. A dash indicates species that could not be classified into any diet width category because of low visitation records.

Species	Visitation Records	Plant Families	Plant Genera	Diet width
L. abrophilum	1	1	1	-
L. carpobrotum	1	1	1	-
L. copleyense	1	1	1	-
L. macrops	1	1	1	-
L. bibrochum	2	1	1	-
L. xerophilum	2	1	1	-
L. anforticornum	2	2	2	-
L. nefrens	3	1	1	-
L. bidens	3	2	2	-
L. lamellosum	3	2	2	-
L. cephalochilum	3	3	3	-
L. lineatum	4	1	1	-
L. frankenia	4	2	2	-
L. moreense	4	2	2	-
L. clypeatum	5	5	5	-
L. mesostenoideum	6	2	3	-
L. roddi	6	3	3	-
L. asperithorax	6	3	4	-
L. seriatum	6	3	4	-
L. bucculum	6	5	5	Polylectic
L. argopilatum	7	1	1	Oligolectic
L. metallicum	7	5	5	Polylectic
L. mu	8	2	2	-
L. inflatum	8	3	3	-
L. platycephalum	8	3	3	-
L. melanopterum	8	5	4	Polylectic
L. tridens	8	4	4	Polylectic
L. smaragdinum	8	3	4	-
L. pollux	8	5	5	Polylectic
L. subplebeium	8	7	7	Polylectic
L. demicapillum	9	1	1	Oligolectic
L. bubrachium	9	3	3	-
L. tamburinei	9	2	4	-
L. hapsidum	9	7	7	Polylectic

L. biceps	10	3	3	-
L. baudini	10	5	6	Polylectic
L. athrix	10	8	9	Polylectic
L. teltiri	12	4	5	Polylectic
L. mesembryanthemi	12	3	6	-
L. parasphecodum	13	2	4	-
L. gynochilum	14	3	3	-
L. impunctatum	14	4	7	Polylectic
L. litovillum	14	4	14	Polylectic
L. plebeium	15	7	9	Polylectic
L. alacarinatum	15	7	10	Polylectic
L. albopilosum	15	8	10	Polylectic
L. conspicuum	16	4	5	Polylectic
L. seminitens	16	5	7	Polylectic
L. ochroma	17	7	6	Polylectic
L. eurycephalum	20	6	6	Polylectic
L. brunnesetum	21	8	12	Polylectic
L. obscurissimum	22	2	2	Oligolectic
L. cambagei	23	8	13	Polylectic
L. aspratulum	24	4	5	Polylectic
L. ochrochilum	24	5	7	Polylectic
L. speculatum	24	9	13	Polylectic
L. seductum	24	12	14	Polylectic
L. bicolor	27	7	11	Polylectic
L. littleri	29	12	17	Polylectic
L. sexsetum	30	11	12	Polylectic
L. supralucens	31	2	3	-
L. quadratum	31	5	8	Polylectic
L. mirandum	32	2	5	Mesolectic
L. pachycephalum	34	8	9	Polylectic
L. convexum	34	10	18	Polylectic
L. brazieri	35	16	24	Polylectic
L. helichrysi	36	6	13	Polylectic
L. globosum	41	8	13	Polylectic
L. humei	42	5	6	Oligolectic
L. opacicolle	42	11	18	Polylectic
L. repraesentans	43	11	16	Polylectic
L. calophyllae	44	5	9	Polylectic
L. gunbowerense	50	7	12	Polylectic
L. victoriellum	55	9	20	Polylectic
L. sculpturatum	56	10	22	Polylectic
L. platychilum	60	11	17	Polylectic
L. fasciatum	61	11	14	Polylectic
L. imitans	62	9	18	Polylectic
L. expansifrons	69	15	23	Polylectic
L. pulvitectum	72	22	34	Polylectic

L. occiduum	73	8	17	Polylectic
L. gilesi	75	21	31	Polylectic
L. willsi	79	13	15	Polylectic
L. amplexum	79	14	22	Polylectic
L. mediopolitum	84	10	11	Polylectic
L. ebeneum	91	18	20	Polylectic
L. triangulatum	97	19	22	Polylectic
L. bicingulatum	101	7	16	Polylectic
L. chapmani	105	25	39	Broad Polylectic
L. adustum	109	15	18	Polylectic
L. instabilis	118	17	29	Polylectic
L. veronicae	123	15	18	Polylectic
L. soror	127	20	26	Polylectic
L. greavesi	128	20	31	Polylectic
L. eremaean	153	35	53	Broad Polylectic
L. mundulum	172	17	26	Polylectic
L. immaculatum	175	25	31	Polylectic
L. castor	190	21	37	Polylectic
L. orbatum	199	10	22	Polylectic
L. hemichalceum	220	20	35	Polylectic
L. vitripenne	243	17	27	Polylectic
L. appositum	255	11	21	Polylectic
L. florale	281	11	16	Polylectic
L. erythrurum	420	32	61	Broad Polylectic
L. cognatum	750	40	76	Broad Polylectic
L. clelandi	1205	21	54	Broad Polylectic
L. lanarium	1242	45	89	Broad Polylectic

Statement of Authorship

Title of Paper	The association between broad polylecty and olfactory sensory morphology in *L.* (*Chilalictus*) - Apoidea: Halictidae
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	NA

Principal Author

Name of Principal Author (Candidate)	Trace Akankunda		
Contribution to the Paper	Developed the concept, conducted scanning electron microscopy, conducted 95% of the data analysis, interpreted the results, wrote and revised the manuscript		
Overall percentage (%)	85%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	9/04/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Katja Hogendoorn		
Contribution to the Paper	Contributed to concept development, edited and revised the manuscript		
Signature		Date	1/4/2019

Name of Co-Author	Carlos Rodriguez Lopez		
Contribution to the Paper	Contributed to concept development, edited and revised the manuscript		
Signature		Date	3/4/2019

Name of Co-Author	Remko Leijs		
Contribution to the Paper	Contributed to concept development, assisted with data analysis, revised and editted the manuscript		
Signature		Date	4/4/2019

Name of Co-Author	Ken Walker		
Contribution to the Paper	Provided data on body measurements of the species analysed, revised the manuscript		
Signature		Date	2/04/2019

Chapter 4: The association between broad polylecty and olfactory sensory morphology in *Lasioglossum* (*Chilalictus*) (Halictidae, Apoidea)

Trace Akankunda[1], Carlos Rodriguez Lopez[2], Remko Leijs[3], Ken Walker[4], Katja Hogendoorn[1,5]

[1]The University of Adelaide, School of Agriculture, Food and Wine Adelaide SA 5005, [2]Environmental Epigenetics and Genetics Group; Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington, KY 40546, [3]South Australian Museum, North Terrace Adelaide SA 5000, [4]Museums Victoria, Melbourne VIC 3001, [5]Corresponding author

Abstract

Bees are classified as generalists (polylectic) or specialists (oligolectic) depending on the number of pollen hosts they visit. Species with a broad diet are expected to respond to a large range of sensory cues to identify their pollen hosts in the field, while specialists could afford to respond only to a few cues when searching for pollen. Therefore, we expect that polylectic species have a broader sensory sensitivity in comparison with oligolectic species. We investigated the association between broad polylecty and olfactory sensory morphological traits in the Australian native bees of *Lasioglossum* (*Chilalictus*), i.e. length of antennae and density of olfactory sensilla. We hypothesised that broadly polylectic species have relatively longer antennae and a higher density of olfactory sensillae in comparison with species in lower pollen host breadth categories. We found no significant differences in relative antennal length between broadly polylectic species and species in lower pollen host breadth categories. In addition, there was no correlation between the density of sensilla and body size, and the density of olfactory sensilla was not significantly different between broadly polylectic and polylectic species. Our results suggest that peripheral allometric growth in olfactory sensory organs is not associated with the evolution of broad polylecty in bees and support the hypothesis that behavioural traits are more evolutionarily labile than morphological traits. Further studies need to be conducted to determine if there are structural differences at higher olfactory sensory levels between broadly polylectic species and species with narrower pollen host breadths.

Key words: evolution, broad polylecty, olfactory sensory morphology

Introduction

Bee species that forage on multiple pollen hosts are referred to as 'polylectic', while those that forage on a limited number of pollen hosts are called oligolectic species (Cane & Sipes 2006). Using the number and taxonomic diversity of the pollen hosts visited, bees can be further categorised into monolectic, eclectic, oligolectic, polylectic and broadly polylectic groups (Cane & Sipes 2006), although this classification is still contentious (Vossler 2018; Müller & Kuhlmann 2008). Previous research into the evolution of pollen host breadth in bees has largely focussed on the evolution and maintenance of oligolectic behaviour. This is in part because oligolecty seems to be in need of an evolutionary explanation (Futuyma & Moreno 1988; Minckley & Roulston 2006), and also because it was assumed that oligolectic behaviour was a derived character state, while polylectic behaviour was thought to be ancestral (Waser et al. 1996). However, it has now been demonstrated that, within bees as a whole, oligolectic behaviour is considered ancestral and polylectic behaviour is the derived state (Danforth et al. 2013; Danforth et al. 2006), and that reversals are common (Haider et al. 2014; Danforth et al. 2003). Therefore, in addition to research into the evolution of oligolecty, the evolutionary drivers of polylectic behaviour and the factors that maintain polylectic and broadly polylectic behaviour in bees need to be explored.

Oligolectic species are subject to genetic constraints that might stop them from detecting, accessing or digesting pollen from alternative hosts (Sedivy et al. 2008). The existence of such constraints suggests that polylectic species have gained genetic adaptations that enable them to overcome some of the constraints presented by the variety of pollen hosts they visit (Sedivy et al. 2008). Such adaptations would perhaps be more pronounced in broadly polylectic species, which are capable of foraging on

over 25% of locally available pollen hosts (Cane & Sipes 2006). To date, no research has been done to determine the genetic adaptations that coincide with the evolution of broadly polylectic behaviour in bees. However, based on the constraint hypothesis suggested for the evolution of pollen host breadth in bees (Sedivy et al. 2008), the genetic adaptations that maintain broadly polylectic behaviour could be morphological, neurological or digestive physiological (Akankunda, in prep.).

In herbivorous insects, diet width is positively correlated with body size (Davis et al. 2013; Novotny & Basset 1999; Loder et al. 1998; Inkinen 1994; Wasserman & Mitter 1978), although this is not universal (Komonen et al. 2004; García-Barros 2000). However, for bees, a correlation between pollen host breadth and body size was not found to exist in a comparison of 13 pairs of closely related generalist and specialist bee species (Roulston & Cane 2000). Furthermore, in anthidiini bees, the evolution of polylecty from oligolecty was accompanied by a reduction instead of an increase in body size (Muller 1996). Combined, these results suggest that pollen host breadth in bees might be associated with other factors than body size.

The ability to utilise and perceive a wider range of pollen could lead to an association between pollen host breadth in bees and might be associated with the size and number of sensory organs (sensory allometry). The peripheral olfactory receptors in bees are found on the antenna and consist of sensilla placodea, trichodea type A, basiconica and coelonica. Sensilla placodea and trichodea are predominant in numbers (Frasnelli & Vallortigara 2017; Frasnelli et al. 2010). Riveros & Gronenberg (2010) found a positive correlation between the number of placodea and the efficiency of pollen collection in honeybees. This correlation was attributed to an increased sensory sensitivity in bees

with more sensilla placodea (Riveros & Gronenberg 2010). In addition, species of bumble bees with larger and more olfactory sensory organs exhibit a higher sensory sensitivity (Spaethe et al. 2007), and collect more life time pollen bouts (Russell, Morrison, et al. 2017). Given that olfactory sensory sensitivity is positively correlated with resource exploitation (Riveros & Gronenberg 2010), one might expect broadly polylectic species to have a broader olfactory sensory sensitivity than species in lower pollen host breadth categories. Such broader sensitivity could be the result of adaptations at peripheral or higher levels of the sensory system, or a combination of these factors.

Wcislo (1995) measured the correlation between pollen host breadth and the density of olfactory sensilla placodea using four Halictid species from two genera of *Dieunomia* (as specialists) and *Nomia* (as generalists). He found no correlation between the density of sensilla and pollen host breadth but reported significant differences between bee families in the relationship between the two traits. Therefore, given the small number of species studied, it is necessary to validate these results using other taxonomic groups of bees. Additionally, although the need to account for phylogenetic history in assessing the relationship between sensory allometry and behaviour was acknowledged, robust comparative methods could not be applied in the absence of a phylogeny.

Here, we examine the association between pollen host breadth, relative antennal length and the density of antennal olfactory sensilla for a number of species in the Australian subgenus *Lasioglossum* (*Chilalictus*). We hypothesised that broadly polylectic species have relatively longer antennae and a higher density of olfactory sensilla in comparison

with species in lower pollen host breadth categories. The species in this genus are highly suitable for this study for three reasons. Firstly, the species' pollen host breadth categories, range from oligolecty to broad polylecty (Akankunda, in prep.). Secondly, a molecular phylogeny for 51 species of the subgenus has been recently generated (Akankunda, in prep.). We used a comparative phylogenetics approach to measure the association between traits while accounting for phylogenetic non-independence due to shared ancestry between species (O'Meara 2012; Symonds & Blomberg 2014).

Methods

Data collection

To conduct comparative phylogenetic studies, we used morphological data for 51 species of *L.* (*Chilalictus*; Walker 1995). We used body length as a proxy for body size (Gathmann & Tscharntke 2002; Symonds & Elgar 2013) and took the average body length per species measured using 10 specimens each, although for some species, fewer measurements were available (Walker 1995; Data S4.1). Because the literature provided measurements of various body parts relative to head width, we converted these relative data back to actual dimensions for scape and flagellum lengths. Measurement data were tested for normality (Shapiro & Wilk 1965), and where needed, log transformed to conform to the statistical assumptions of normality. The pollen host breadth categories assigned to the species and a molecular phylogeny of *L.* (*Chilalictus*) were obtained from data in chapter 3 of this thesis.

In order to analyse the significance of the differences in the density of antennal sensilla between polylectic and broadly polylectic species, 15 species were selected for imaging using the scanning electron microscope (Fig. 4.1). The species were selected from the molecular phylogeny based on their pollen host breadth categories and taxonomic grouping (Fig. 4.1). We selected species across the molecular phylogeny to account for phylogenetic non-independence in the analysis. Voucher specimens for the selected species have been submitted to the South Australian Museum, and identifications have been verified by us (KW). When available, 3 female specimens were selected for each species to make a total of 39 specimens used for imaging (Table S4.1).

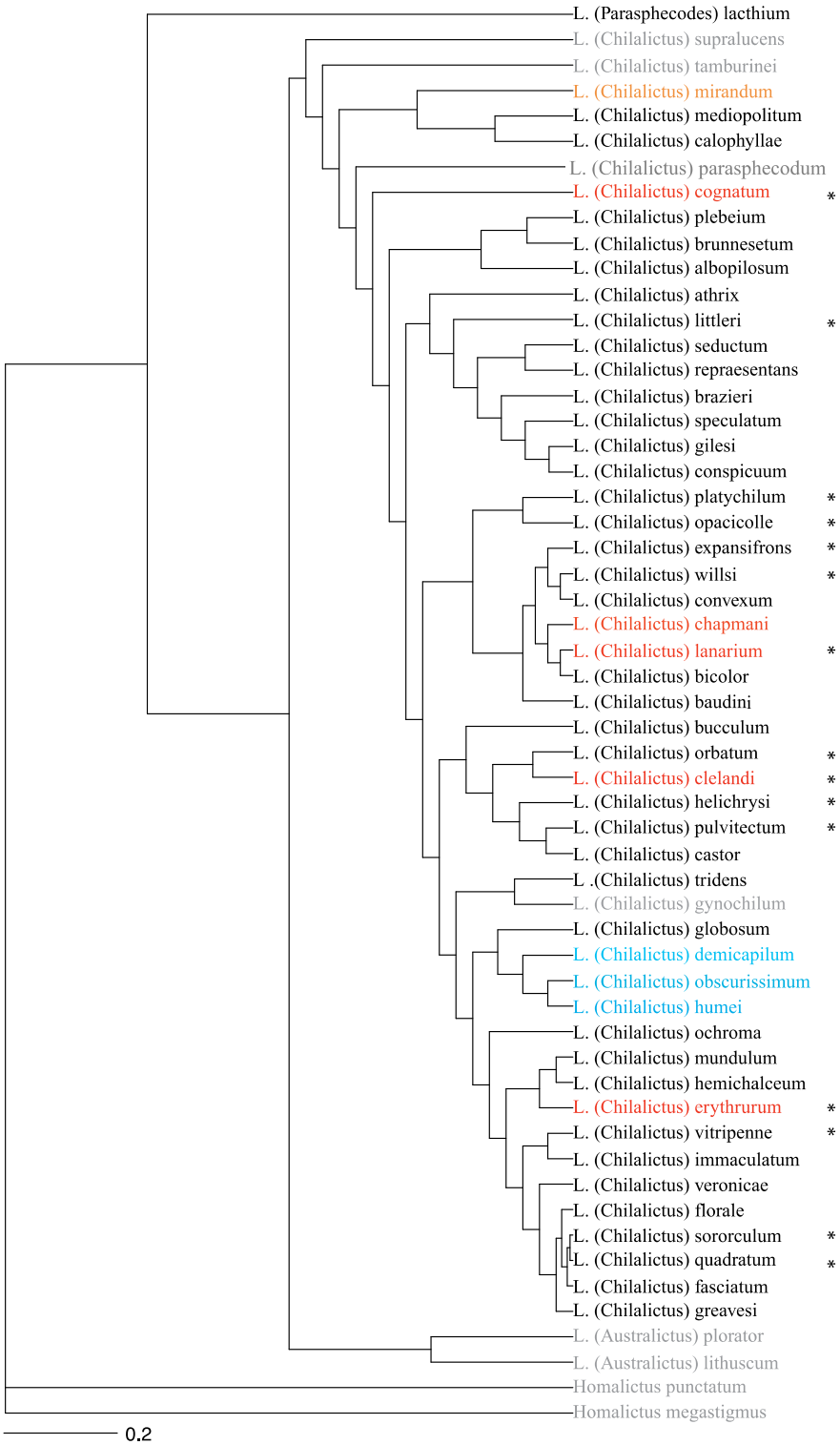
L. (Parasphecodes) lacthium
L. (Chilalictus) supralucens
L. (Chilalictus) tamburinei
L. (Chilalictus) mirandum
L. (Chilalictus) mediopolitum
L. (Chilalictus) calophyllae
L. (Chilalictus) parasphecodum
L. (Chilalictus) cognatum *
L. (Chilalictus) plebeium
L. (Chilalictus) brunnesetum
L. (Chilalictus) albopilosum
L. (Chilalictus) athrix
L. (Chilalictus) littleri *
L. (Chilalictus) seductum
L. (Chilalictus) repraesentans
L. (Chilalictus) brazieri
L. (Chilalictus) speculatum
L. (Chilalictus) gilesi
L. (Chilalictus) conspicuum
L. (Chilalictus) platychilum *
L. (Chilalictus) opacicolle *
L. (Chilalictus) expansifrons *
L. (Chilalictus) willsi *
L. (Chilalictus) convexum
L. (Chilalictus) chapmani
L. (Chilalictus) lanarium *
L. (Chilalictus) bicolor
L. (Chilalictus) baudini
L. (Chilalictus) bucculum
L. (Chilalictus) orbatum *
L. (Chilalictus) clelandi *
L. (Chilalictus) helichrysi *
L. (Chilalictus) pulvitectum *
L. (Chilalictus) castor
L .(Chilalictus) tridens
L. (Chilalictus) gynochilum
L. (Chilalictus) globosum
L. (Chilalictus) demicapilum
L. (Chilalictus) obscurissimum
L. (Chilalictus) humei
L. (Chilalictus) ochroma
L. (Chilalictus) mundulum
L. (Chilalictus) hemichalceum
L. (Chilalictus) erythrurum *
L. (Chilalictus) vitripenne *
L. (Chilalictus) immaculatum
L. (Chilalictus) veronicae
L. (Chilalictus) florale
L. (Chilalictus) sororculum *
L. (Chilalictus) quadratum *
L. (Chilalictus) fasciatum
L. (Chilalictus) greavesi
L. (Australictus) plorator
L. (Australictus) lithuscum
Homalictus punctatum
Homalictus megastigmus
0.2

Figure 4.1. Ultrametric phylogenetic tree of 51 species of *L*. (*Chilalictus*) used in this study. Pollen host breadth categories of species are indicated by colours as follows: red: broadly polylectic; black: polylectic; orange: mesolectic; blue: oligolectic; grey: unknown. Asterisks indicate species used for scanning electron microscopy of the antennae.

Scanning electron microscopy (SEM)

For each specimen, the right antenna was carefully removed from the head at the antennal base using forceps and mounted onto an SEM specimen stub covered with a sticky tab. The antennae were placed on the stubs with the dorsal view facing upwards and a 10nm thick platinum coating was applied. Imaging was done on the dorsal view using the Philips XL30 field emission scanning electron microscope with the following conditions; Acc V = 10kv, spot size = 3.0, magnification = 1, 500, working distance = 10 mm (lowered to 20 mm when the sample had to be tilted to capture the dorsal view). The sample processing and imaging were conducted at Adelaide Microscopy and images will be uploaded into Dryad Digital Repository.

Comparative phylogenetics

All the comparative phylogenetics analyses were conducted in R v 3.5.2 (R Core Team 2018), using the packages Ape v 5 (Paradis 2012), Geiger v 2.0.6.1 (Harmon et al. 2019) and Sensiphy (Paterno et al. 2018). In order to account for phylogenetic uncertainty in the analyses, 250 phylogenetic trees were subsampled from post burnin tree files of four MrBayes analyses (Chapter 3). From these, 100 randomly selected phylogenetic trees were ultrametricised using the Penalised likelihood method (Sanderson 2002) in Ape, for use in subsequent analyses. This was done to conform to the assumptions of comparative phylogenetics analysis (Symonds & Elgar 2013).

Phylogenetic signal was calculated for pollen host breadth and antennal length using Pagel's lambda (Pagel 1999; O'Meara 2012) with the fitdiscrete function in Geiger and the tree_physig function in Sensiphy, respectively. To determine the significance of the phylogenetic signal observed in pollen host breadth, we used a pchisq function in R to conduct a likelihood ratio test by comparing likelihood estimates of two models with one assuming no phylogenetic signal and the other taking on the maximum likelihood estimate of Pagel's lambda. To determine the relationship between antennal length and body length, we conducted a phylogenetic regression analysis of antennal length against body length using phylogenetic generalised linear squares; PGLS (Grafen 1989), with the tree_phylm function in Sensiphy. To determine if broadly polylectic species had larger antennae than species in lower pollen host breadth categories, we regressed antennal length against pollen host breadth while controlling for body length using PGLS (Symonds & Elgar 2013). For this analysis, pollen host breadth categories were converted into binary characters with broad polylecty = 1 and other categories = 0 (i.e polylectic, mesolectic and oligolectic). All parameter estimates were averaged across 100 ultrametricised phylogenetic trees to account for phylogenetic uncertainty. Figures were edited using Adobe illustrator v 22.

Analysis of the density of sensilla

The images were analysed using ImageJ v 1.51 (Schneider et al. 2012) and statistical analysis was conducted in the SPSS statistical package. The two olfactory sensilla counted were trichodea type A and placodea (Wcislo 1995; Frasnelli & Vallortigara 2017). Both placodea and trichodea were counted on the dorsal view for segments 5 – 11. These segments were chosen because the dorsal side of segments 1 to 3 had only

few trichodea and placodea, while segments 4 and 12 were often curved out of the field of view and could not be imaged properly for several specimens. Counting was done by hand using point markers in ImageJ to avoid repetition. On each segment, sensilla were counted within a uniform surface area of 0.006 mm^2. In order to compare the density of the sensilla between broadly polylectic and polylectic species, we summed up the counts of sensilla across seven flagellar segments for each species. This is equivalent to counting the number of sensilla on a 0.042 mm^2 surface area of the left flagellum per species. To determine whether the density of sensilla was correlated with pollen host breadth, we applied a general linear model using pollen host breadth as a factor and species as a covariate. The requirements of covariance and variance were met and we assumed the observations in each group to be sufficiently independent since the species were selected across the molecular phylogeny.

Results

There was a significant phylogenetic signal in pollen host breadth and the antennal length of the species of *L.* (*Chilalictus*) analysed (Table 4.1).

Table 4.1. Phylogenetic signal of categorical and continuous traits using Pagel's lambda.

Trait	Mean lambda	Significance
Pollen host breadth	0.72	$p < 0.01$
Antennal length	0.88	$p < 0.01$

As expected, there was a strong positive correlation between antennal length and body length ($r^2 = 0.93$, $p < 0.001$, lambda = 0.3; Fig. 4.2). However, pollen host breadth was a poor predictor of antennal length ($\beta = -0.009$ ($\pm$ 0.033), lambda = 0.318, $r^2 = 0.93$, $p = 0.8$), meaning there was no difference in antennal length between broadly polylectic species and species in lower pollen host breadth categories.

There was no significant effect of pollen host breadth on the mean average density of sensilla (Hotelling's Trace = 0.09, F = 1.56, p = 0.22) and the density of individual types of sensilla (trichodea; F = 0.74, p = 0.4, placodea; F = 3.13, p = 0.09; Fig. 4.4). The density of placodea increased distally along the flagellum for most of the species (Fig. S4.1), while for trichodea, a similar pattern was observed starting from the 6th flagellar segment (Fig. S4.2). There was no linear relationship between the density of sensilla and head width (placodea; r = 0.004, p = 1; trichodea; r = -0.09, p = 0.07).

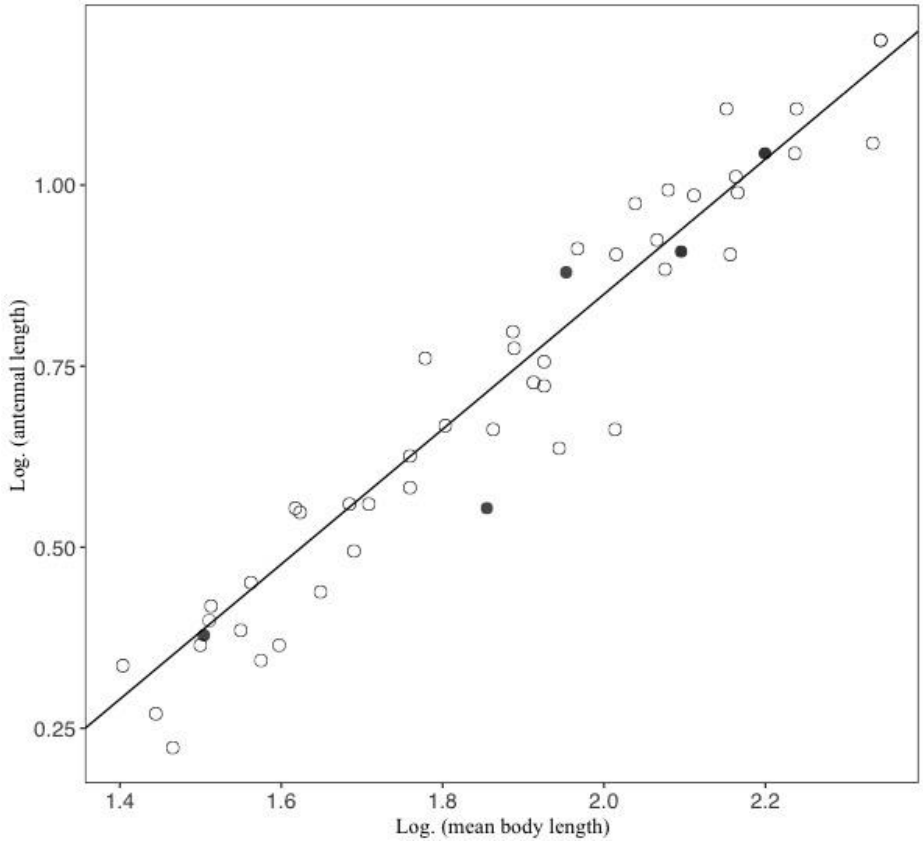

Figure 4.2. Association between antennal and body length of *L.* (*Chilalictus*) species. Pollen host breadth of species is indicated as follows: filled circles: broadly polylectic; empty circles: others (i.e. polylectic, mesolectic, oligolectic)

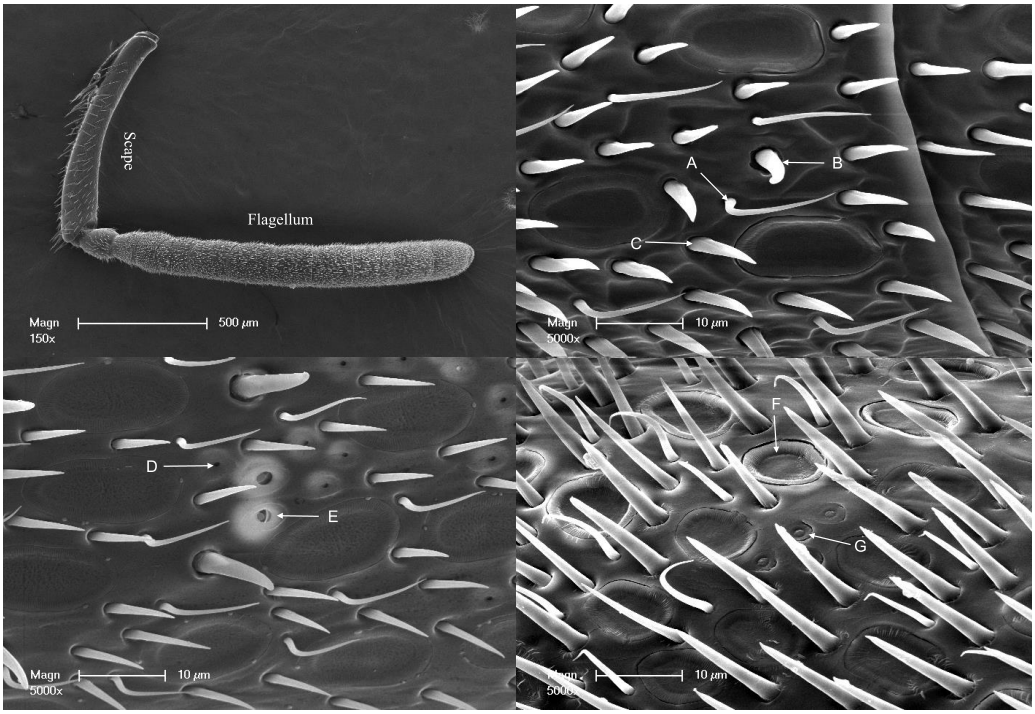

Figure 4.3. Types of sensilla found on the dorsal side of the antenna. Top left: left flagellum, Top right: A = trichodea type B, B = basiconica, C = trichodea type A, Bottom left: D = ampullacea, E = coelonica, Bottom right: F = placodea, G = coelocapitulum.

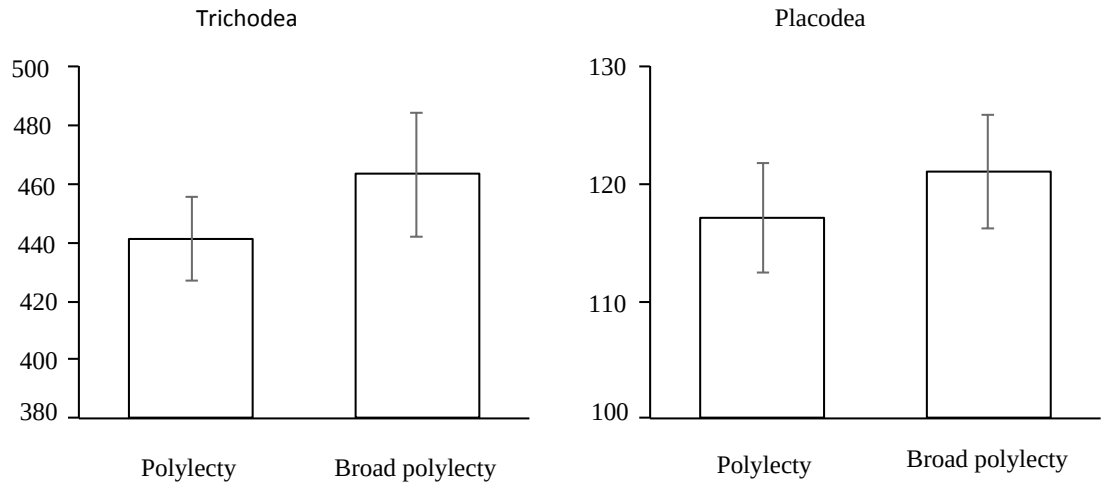

Figure 4.4. The mean total density of trichodea and placodea on the antennae of polylectic and broadly polylectic species summed over flagellar segments 5 – 11. The error bars indicate standard error of the mean.

Discussion

Contrary to our hypotheses, there was no significant difference in relative antennal length between broadly polylectic species and species in lower pollen host breadth categories. Additionally, the density of olfactory sensilla was not significantly different between broadly polylectic and polylectic species. We discuss the implications of our results on the ongoing research into factors driving and maintaining pollen host breadth in bees.

The significance of the phylogenetic signal in antennal length of the species studied was expected because morphological characters are highly heritable (reviewed in Blomberg et al. 2003). In addition, the phylogenetic signal of behavioural traits is expected to be lower than that of morphological traits because they are highly adaptive and usually affected by sampling bias and measurement error (Blomberg et al. 2003; Losos 2008). In agreement with Blomberg et al. (2003), pollen host breadth had a lower phylogenetic signal than morphological traits in this subgenus. However, the phylogenetic signal of pollen host breadth was significant, which supports suggestions that pollen host breadth in bees is not a labile trait and is highly genetically controlled (Sedivy et al. 2008). It seems likely that the phylogenetic signal of pollen host breadth was weakened by the fact that broadly polylectic species were widely distributed over the phylogenetic tree, an indication that broad polylecty has evolved by convergent evolution in this subgenus (Losos 2011).

We found no correlation between pollen host breath and relative antennal length or the density of olfactory sensilla placodea and trichodea. This is in agreement with previous research (Wcislo 1995). Given that broad polylecty evolved from polylecty in this

subgenus (Akankunda, in prep), these results suggest that the transition to broadly polylectic behaviour in bees does not coincide with morphological adaptations in the size or number of peripheral olfactory organs. However, we did not examine differences in the density of putative olfactory sensilla coelonica and basiconica between pollen host breadth categories because they occur in much lower abundances compared to sensilla placodea and trichodea (Galvani et al. 2012). Our findings also suggest that the positive correlation observed between olfactory sensory allometry and foraging efficiency found in intra-species comparisons (Spaethe et al. 2007; Riveros & Gronenberg 2010; Kapustjanskij et al. 2007), might not be supported across species. However, this study only examined gross peripheral morphological differences and future studies should compare structural differences at higher olfactory sensory levels to validate our findings.

In addition, morphological adaptations for broadly polylectic behaviour might occur in other characters not examined in this study. For example oligolectic bees exhibit differences in the length, spacing and level of branching of scopal hairs and the vestiture that aid in collecting and transporting pollen of different sizes and surface structure from their specific pollen hosts (Thorp 1979). Similar adaptations in the branching patterns of mesoventral hair shafts in oligolectic species of *L.* (*Chilalictus*), aid in collecting pollen from flowers of the genus *Wahlenbergia* (Walker 1995). Furthermore, females of *L.* (*Chilalictus*) species exhibit diverse modifications in their mouthparts, especially the labrum, which has several ornate structures such as lateral ridges, teeth and setae that vary in number and shape, and might correlate with the morphology of the flowers they visit (Walker 1995). Similar modifications on the labrum of *Habropoda laboriosa* have been previously suggested to aid in pollen

collection (Thorp 2000). Given that broadly polylectic species collect pollen from pollen hosts of several oligolectic bees (Minckley & Roulston 2006), one would expect to find modifications in their scopal hairs, vestiture or mouthparts for exploiting the diversity of flowers they visit. For example, in grasshoppers (Orthoptera), species that are polyphagous have generalised mandibles while species that specialise on grasses have evolved specialist cutting and grinding mandibles (Bernays 1998).

In closing, our results support the hypothesis that behavioural traits might be more evolutionarily labile than morphological traits (Blomberg et al. 2003), such that species adopt a behavioural pattern after which their morphological structures adapt to better suit the behaviour (Wcislo 1989). While it is theoretically possible that morphological adaptations are lagging behind broadening of the diet in broadly polylectic bees of *L.* (*Chilalictus*), it seems more likely that structural adaptations for broad polylecty in this subgenus occur at a higher olfactory sensory level than has been examined in this study. Several morphological traits that are useful for collecting and handling pollen remain unexplored and should be examined in future studies of adaptations for broad polylecty in bees.

Beside morphological adaptations, broadly polylectic species might have behavioural adaptations for exploiting multiple pollen hosts such as the ability to switch between floral sonication and scrabbling while collecting pollen, as has been reported in *B. impatiens* (Russell, Buchmann, et al. 2017). They could also have adaptations in their digestive physiology that enables them to deal with pollen of diverse chemical compositions (Sedivy et al. 2011). These too could be explored in future studies.

References

Bernays, E.A., 1998. Evolution of feeding behaviour in insect herbivores. *Bioscience*, 48(1), pp. 35–44.

Blomberg, S.P., Garland, T. & Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioural traits are more labile. *Evolution*, 57(4), pp. 717–745.

Cane, J.H. & Sipes, S., 2006. Characterizing floral specialization by bees: analytical methods and a revised lexicon for oligolecty. In N. M. Waser, Ollerton Jeff, ed. *Plant-pollinator interactions: from specialization to generalization*. Chicago University of Chicago Press, pp. 99–122.

Danforth, B.N. et al. , 2006. The history of early bee diversification based on five genes plus morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 103(41), pp. 15118–15123.

Danforth, B.N. et al. , 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual Review of Entomology*, 58(1), pp. 57–78.

Danforth, B.N., Conway, L. & Ji, S., 2003. Phylogeny of eusocial Lasioglossum reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). *Systematic Biology*, 52(1), pp. 23–36.

Davis, R.B. et al. , 2013. Degree of specialization is related to body size in herbivorous insects: a phylogenetic confirmation. *Evolution*, 67(2), pp. 583–589.

Frasnelli, E. et al. , 2010. Morpho-functional asymmetry of the olfactory receptors of the honeybee (Apis mellifera). *Behavioural Brain Research*, 209(2), pp. 221–225.

Frasnelli, E. & Vallortigara, G., 2017. Distribution of antennal olfactory and non-olfactory sensilla in different species of bees. *Symmetry*, 9(8),p. 135.

Futuyma, D.J. & Moreno, G., 1988. The evolution of ecological specialisation. *Annual Review of Ecology, Evolution, and Systematics*, 19 pp. 207-233

Galvani, G.L. et al. , 2012. Distribution and morphometric studies of flagellar sensilla in Emphorini bees (Hymenoptera, Apoidea). *Micron*, 43(5), pp. 673–687.

García-Barros, E., 2000. Body size, egg size, and their interspecific relationships with ecological and life history traits in butterflies (Lepidoptera: Papilionoidea, Hesperioidea). *Biological Journal of the Linnean Society*, 70(2), pp. 251–284.

Gathmann, A. & Tscharntke, T., 2002. Foraging ranges of solitary bees. *Journal of Animal Ecology*, 71(5), pp. 757–764.

Grafen, A., 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 326(1233), pp. 119–157.

Haider, M. et al. , 2014. Phylogeny and floral hosts of a predominantly pollen generalist group of mason bees (Megachilidae: Osmiini). *Biological journal of the Linnean Society. Linnean Society of London*, 111(1), pp. 78–91.

Harmon, L. et al. , 2019. *geiger: Analysis of evolutionary diversification*, USA.

Inkinen, P., 1994. Distribution and abundance in British noctuid moths revisited. *Annales Zoologici Fennici*, 31(2), pp. 235–243.

Kapustjanskij, A. et al. , 2007. Bigger is better: implications of body size for flight ability under different light conditions and the evolution of alloethism in bumblebees. *Functional ecology*, 21(6), pp. 1130–1136.

Komonen, A. et al. , 2004. The role of niche breadth, resource availability and range position on the life history of butterflies. *Oikos*, 105(1), pp. 41–54.

Loder, N. et al. , 1998. Body size and feeding specificity: macrolepidoptera in Britain. *Biological journal of the Linnean Society. Linnean Society of London*, 63(1),

pp. 121–139.

Losos, J.B., 2011. Convergence, adaptation, and constraint. *Evolution*, 65(7), pp. 1827–1840.

Losos, J.B., 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11(10), pp. 995–1003.

Minckley, R.L. & Roulston, T.H., 2006. Incidental mutualism and pollen specialization among bees. Plant-pollinator interactions. In N. M. Waser, Ollerton Jeff, ed. *From specialization to generalization*. Chicago: The University of Chicago Press, pp. 69–98.

Muller, A., 1996. Host-plant specialization in western palearctic anthidine bees (Hymenoptera: Apoidea: Megachilidae). *Ecological monographs*, 66(2), pp. 235–257.

Müller, A. & Kuhlmann, M., 2008. Pollen hosts of western palaearctic bees of the genus *Colletes* (Hymenoptera: Colletidae): the Asteraceae paradox. *Biological Journal of the Linnean Society*, 95(4), pp. 719–733.

Mundry, R., 2014. Statistical issues and assumptions of phylogenetic generalized least squares. In L. Z. Garamszegi, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 131–153.

Novotny, V. & Basset, Y., 1999. Body size and host plant specialization: a relationship from a community of herbivorous insects on *Ficus* from Papua New Guinea. *Journal of Tropical Ecology*, 15(03), pp. 315–328.

O'Meara, B.C., 2012. Evolutionary inferences from phylogenies: A review of methods. *Annual Review of Ecology, Evolution, and Systematics*, 43(1), pp. 267–285.

Pagel, M., 1999. Inferring the historical patterns of biological evolution. *Nature*, 401(6756), pp. 877–884.

Paradis, E., 2012. *Analysis of phylogenetics and evolution with R*, New York, NY: Springer New York.

Paterno, G.B., Penone, C. & Werner, G.D.A., 2018. sensiPhy: an R-package for sensitivity analysis in phylogenetic comparative methods. *Methods in ecology and evolution / British Ecological Society*, 9(6), pp. 1461–1467.

R Core Team, 2018. *R: A language and environment for statistical computing*, Vienna: R Foundation for Statistical Computing.

Riveros, A.J. & Gronenberg, W., 2010. Sensory allometry, foraging task specialization and resource exploitation in honeybees. *Behav Ecol Sociobiol*, 64(6), pp. 955–966.

Roulston, T.H. & Cane, J.H., 2000. The effect of diet breadth and nesting ecology on body size variation in bees (Apiformes). *Journal of the Kansas Entomological Society*, 73(3), pp. 129–142.

Russell, A.L. , Morrison, S.J., et al. , 2017. Patterns of pollen and nectar foraging specialization by bumblebees over multiple timescales using RFID. *Scientific reports*, 7, p.42448.

Russell, A.L. , Buchmann, S.L. & Papaj, D.R., 2017. How a generalist bee achieves high efficiency of pollen collection on diverse floral resources. *Behavioural Ecology*, 28(4), pp. 991–1003.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, 19(1), pp. 101–109.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W., 2012. NIH Image to ImageJ: 25

years of image analysis. *Nature Methods*, 9(7), pp. 671–675.

Sedivy, C. et al. , 2008. Patterns of host-plant choice in bees of the genus *Chelostoma*: the constraint hypothesis of host-range evolution in bees. *Evolution*, 62(10), pp. 2487–2507.

Sedivy, C., Müller, A. & Dorn, S., 2011. Closely related pollen generalist bees differ in their ability to develop on the same pollen diet: evidence for physiological adaptations to digest pollen. *Functional ecology*, 25(3), pp. 718–725.

Shapiro, S.S. & Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), pp. 591–611.

Spaethe, J. et al. , 2007. Size determines antennal sensitivity and behavioural threshold to odors in bumblebee workers. *Die Naturwissenschaften*, 94(9), pp. 733–739.

Symonds, M.R.E. & Blomberg, S.P., 2014. A primer on phylogenetic generalised least squares. In L. Z. Garamszegi, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 105–130.

Symonds, M.R.E. & Elgar, M.A., 2013. The evolution of body size, antennal size and host use in parasitoid wasps (Hymenoptera: Chalcidoidea): a phylogenetic comparative analysis. *Plos One*, 8(10), p.e78297.

Thorp, R.W., 1979. Structural, behavioural, and physiological adaptations of bees (Apoidea) for collecting pollen. *Annals of the Missouri Botanical Garden*, 66(4),p. 788.

Thorp, R.W., 2000. The collection of pollen by bees. *Plant Systematics and Evolution*, 222(1–4), pp. 211–223.

Vossler, F.G., 2018. Are stingless bees a broadly polylectic group? an empirical study of the adjustments required for an improved assessment of pollen diet in bees.

In P. Vit, S. R. M. Pedro, & D. W. Roubik, eds. *Pot-Pollen in Stingless Bee Melittology*. Cham: Springer International Publishing, pp. 17–28.

Walker, K.L., 1995. Revision of the Australian native bee subgenus *Lasioglossum* (*Chilalictus*). *Memoirs of Museum Victoria*, 55(1 & 2), pp. 1–423.

Waser, N.M. et al. , 1996. Generalization in pollination systems and why it matters. *Ecology*, 77(4),p. 1043.

Wasserman, S.S. & Mitter, C., 1978. The relationship of body size to breadth of diet in some Lepidoptera. *Ecological entomology*, 3(2), pp. 155–160.

Wcislo, W.T., 1989. Behavioural environments and evolutionary change. *Annual review of ecology and systematics*, 20(1), pp. 137–169.

Wcislo, W.T., 1995. Sensilla numbers and antennal morphology of parasitic and non-parasitic bees (Hymenoptera : Apoidea). *International Journal of Insect Morphology and Embryology*, 24(1), pp. 63–81.

Supporting information

Table S4.1. Species used for scanning electron microscopy of the left antenna

Species	Pollen host breadth	Number of specimens used
L. erythrurum	Broad polylecty	3
L. cognatum	Broad polylecty	3
L. clelandi	Broad polylecty	3
L. lanarium	Broad polylecty	3
L. vitripenne	Polylecty	3
L. sororculum	Polylecty	3
L. willsi	Polylecty	2
L. opacicolle	Polylecty	2
L. quadratum	Polylecty	3
L. orbatum	Polylecty	1
L. pulvitectum	Polylecty	3
L. helichrisy	Polylecty	2
L. platychilum	Polylecty	3
L. littleri	Polylecty	3
L. expansifrons	Polylecty	2

CHAPTER 5

GENERAL DISCUSSION

Summary

In this study, I examined a number of factors related to the evolution of pollen host breadth in bees, with a focus on the evolution of broad polylecty in the native Australian subgenus *Lasioglossum* (*Chilalictus*). Pollen host breadth is an important attribute in bees because it is associated with several life history traits, such as phenology, longevity, and sociality. It also determines a species extinction risk, its economic importance as a pollinator of crops and the significance of a species in the ecosystem through interactions with other insects and plants. In this context, broadly polylectic species are of significant interest because they are versatile pollinators of crops and wild plants, and are therefore keystone species in their habitats. They also have wide spatial and temporal distributions that protect them from going through population bottlenecks.

Despite the significance of broadly polylectic behaviour, and the general understanding that the trait is derived, the evolution of broad polylecty in bees has not been studied. In chapter 1, I reviewed the literature on the evolution of pollen host breadth in bees and polyphagy in herbivorous insects to formulate the hypothesis that constraints to the broadening of the diet ensure that broad polylecty is rare in bees. I further argued that a study of the evolution of broad polylecty in *L.* (*Chilalictus*) may help to unravel these constraints. In chapter 2 a multi-locus sequencing method was developed to generate sequences of *L.* (*Chilalictus*) that would allow the expansion of the existing molecular phylogeny of this subgenus. In chapter 3, a molecular phylogeny of *L.* (*Chilalictus*)

was generated and used to reconstruct the evolutionary history of pollen host breadth in this subgenus. In chapter 4, the association between sensory morphology and pollen host breadth in *L.* (*Chilalictus*) was studied to investigate correlated evolution between morphology and diet width. Below, the key findings of this study are summarised and recommendations for future research on the evolution of pollen host breadth in bees are made.

Why is broadly polylectic behaviour rare in bees?

An analysis of the literature suggests that broad polylecty is rare in bees and seems to be more dominant in two out of seven families, namely; Apidae and Halictidae, (Vossler 2018; Dalmazzo & Vossler 2015; Cane & Sipes 2006), although there are also broad polyleges in the family Megachilidae (Haider et al. 2014). However, it seems logic to expect evolutionary selective advantages of broad polylecty, and hence it is difficult to understand why this trait might be rare in bees. To form a hypothesis about the factors influencing the evolution of broadly polylectic behaviour in bees, I reviewed the literature on the evolution of polyphagy in herbivorous insects and pollen host breadth in bees. There are similarities in evolutionary patterns of foraging behaviour in bees and herbivorous insects (Sedivy et al. 2011), as the evolution of diet width occurs in both directions of polylecty or oligolecty depending on selection pressures acting within a species microhabitat (Sedivy et al. 2008; Janz & Nylin 2008). However, the choice of host appears to be genetically constrained (Sedivy et al. 2008; Janz & Nylin 2008). These constraints can be morphological, neurological or digestive physiological (Sedivy et al. 2008; Bernays 1998; Futuyma & Moreno 1988). Given that broad polylecty appears to be taxonomically confined, it was hypothesised that the rarity of broad polylecty in bees could be explained if there are genetic constraints to the

evolution of broadly polylectic from polylectic behaviour. I therefore hypothesised that these genetic constraints can be deduced by comparing morphological, neurological or digestive physiological differences between broadly polylectic and polylectic species.

Multilocus barcoding of pinned insect specimens using MiSeq

In order to test the hypothesis that the evolution of broad polylecty in bees is influenced by genetic constraints, I studied the evolution of broadly polylectic behaviour in a native Australian subgenus of *L.* (*Chilalictus*). For this study, it was necessary to expand the current molecular phylogeny of 24 species (Gibbs et al. 2012), in order to incorporate the spectrum of pollen host breadth in this subgenus. To establish a well-supported molecular phylogeny, three gene regions that were previously used in the molecular phylogeny of the genus *Lasioglossum*, including representatives of this subgenus, were sequenced from well-identified specimens using a novel nested amplicon sequencing approach on the MiSeq platform. This method was developed to overcome the sequence length limitation of the MiSeq in previous multi-locus barcoding projects (Sonet et al. 2018; Cruaud et al. 2017).

The novel amplicon sequencing method doubled the sequence length for both ethanol and pinned specimens previously achieved on the MiSeq platform (Sonet et al. 2018; Cruaud et al. 2017) and reduced the risk of generating chimeras by merging sequences of the target loci with contaminant sequences (Hebert et al. 2018). However, the sequencing method developed could not generate sequences from pinned specimens that were older than 10 years, which form the bulk of type specimens held in museum collections (Hebert et al. 2013). Alternative methods such as hybrid capture could be used to generate multi-locus barcodes from type specimens (Burrell et al. 2015). Using

nested amplicon sequencing, the number of species in the molecular phylogeny of *L.* (*Chilalictus*) was raised from 24 to 51. While more species remain to be sequenced, this number was sufficient to represent the spectrum of pollen host breadth categories across the *L.* (*Chilalictus*) genus of 134 species, currently described (Walker 1995).

The evolution of broad polylecty in *Lasioglossum* (*Chilalictus*)

The sequences generated in chapter two were used to create a molecular phylogeny, upon which the diet width was mapped to study the evolutionary trajectory of broad polylecty in *L.* (*Chilalictus*) and its association to geographic range size and speciation. The most likely ancestral state of pollen host breadth for this subgenus is polylectic behaviour and broad polylecty has evolved independently at least four times, while oligolectic behaviour has evolved once in three species associated with the flower genus *Wahlenbergia*. The evolution from polylecty to oligolecty is common in the family halictidae (Danforth, Conway, et al. 2003; Wcislo & Cane 1996), and might be driven by the presence of stable and highly abundant flower resource such as the genus *Wahlenbergia* in Australia (Atlas of Living Australia website at http://www.ala.org.au. Accessed June 2018). In addition, pollen host specialisation within a single clade of species supports previous suggestions that oligolectic species might be genetically constrained to a narrow diet in bees (Sedivy et al. 2008).

The factors that influence the evolution of broadly polylectic behaviour in *L.* (*Chilalictus*) are unclear but there is evidence to suggest that the loss of stable pollen resources in the environment might select for species that are capable of foraging on unrelated flowers overtime (Burrell et al. 2015). Under such selection pressure, species will expand their diet width to incorporate pollen hosts for which they are pre-adapted

to exploit. Given that broad polylecty is rare in *L.* (*Chilalictus*), we propose that polylectic species in this subgenus might lack the genetic capacity to incorporate certain plants within their diet.

Broadly polylectic species also have a wider geographical range size in comparison to species of lower pollen host breadths, which concurs with previous research (Slatyer et al. 2013; Slove & Janz 2011). Moreover, a species' geographical range size is positively correlated with its population size (Gaston & Lawton 1988). This might explain why speciation has not occurred in lineages exhibiting this trait within *L.* (*Chilalictus*). With a wide host range, wide geographic range size and high local abundance, broadly polylectic species are likely to maintain a constant gene flow between their subpopulations and as such reduce the chances of reproductive isolation and speciation. In contrast both polylectic and oligolectic species can exhibit strong but not necessarily exclusive pollen host specialisation (Haider et al. 2014; Danforth & Ji 2001; Tasei & Picart 1973), and are therefore more likely to live in smaller and isolated populations. This should facilitate reproductive isolation and subsequent speciation among polylectic and oligolectic species but less so in broadly polylectic species. However, more research is necessary to validate this hypothesis. Further studies are also necessary to determine the nature of genetic constraints that might influence the evolution of broad polylecty in this subgenus.

The association between pollen host breadth and olfactory sensory morphology in *L.* (*Chilalictus*) (Apoidea: Halictidae)

Some of the constraints to broad polylecty could be associated with morphology of sensory organs. Species with a broad diet can be expected to have a broader sensory sensitivity as they track cues from a wide range of flowers compared to species with a narrow diet. Therefore, I tested whether broadly polylectic species have longer antennae and a higher density of sensilla placodea and trichodea, the main olfactory organs on the antennae, than species in lower pollen host breadth categories. There was no significant difference in antennal length between broadly polylectic species and species in lower pollen host breadth categories. In addition, there was no significant difference in the density of sensilla placodea and trichodea between broadly polylectic and polylectic species. These results concur with previous research (Wcislo 1995). Given that broadly polylectic behaviour has evolved from polylectic ancestors in *L.* (*Chilalictus*), these results suggest that the evolution of broad polylecty in bees has not coincided with or selected for allometric growth in peripheral olfactory sensory organs. However, future studies should examine structural differences at higher olfactory sensory levels than examined in this study in order to validate our findings.

General Conclusion

This study has contributed to our understanding on the evolution of pollen host breadth in bees, specifically in the native Australian subgenus of *Lasioglossum* (*Chilalictus*). We focussed on the evolution of broadly polylectic behaviour in this subgenus because despite its evolutionary selective advantages, it appears to be rare in bees and might be taxonomically confined. As in herbivores, pollen host breadth can evolve towards oligolecty or polylecty depending on selection pressures acting within a species microhabitat but the choice of pollen host appears to be genetically constrained.

Within the species of *L.* (*Chilalictus*) examined in this study, polylectic behaviour is the ancestral trait and broad polylecty has evolved independently at least four times while, in the current phylogeny, oligolecty has evolved once. These results support the suggestions that oligolectic species are genetically constrained to their narrow diet and that polylectic species might be genetically constrained from becoming broadly polylectic. In addition, the wide distribution of broadly polylectic species and the finding that broad polylecty is rare and has evolved once in disperse lineages suggests that broad polylectic species might often have limited possibilities of speciation. Furthermore, the nature of the constraints that need to be overcome in the evolution of polylecty to broad polylecty is as yet unclear. We did not find evidence for adaptations in the peripheral olfactory sensory organs of broadly polylectic species in *L.* (*Chilalictus*). Therefore, the transition to broad polylecty in bees might not coincide with allometric growth in peripheral olfactory sensory morphology but further studies are required to examine structural differences at higher sensory levels than examined in this study.

Difficulties and recommendations for future research directions.

In this study, I set out to understand the factors influencing the evolution of broadly polylectic behaviour in bees using *Lasioglossum* (*Chilalictus*) as the study subject. For this research, it was necessary to first understand the evolutionary history of pollen host breadth in *L.* (*Chilalictus*), because it is the basis for developing and testing hypotheses on the factors that might influence the evolution of broad polylecty in this subgenus. In order to study the evolutionary history of pollen host breadth, a molecular phylogeny of this subgenus was generated using sequences from well-identified specimens. However, the molecular phylogeny is incomplete and represents 51 of the 134 species currently described. The amplicon sequencing method developed in this study could not be used to sequence highly degraded samples of the species not represented in the molecular phylogeny. Therefore, in order to produce a complete molecular phylogeny for *L.* (*Chilalictus*), future studies could employ other sequencing methods like the hybrid capture that have been optimised for sequencing highly degraded samples (Burrell et al. 2015).

In addition, although the topology of the molecular phylogeny produced is broadly consistent with the morphological phylogeny produced using 107 species (Walker, unpublished) and a molecular phylogeny previously produced with 24 species (Gibbs et al. 2012), there were several species relationships that were not resolved. This might have influenced the analysis and conclusions made about the evolutionary history of broad polylecty in *L.* (*Chilalictus*). In order to validate the results of this study, it is necessary to repeat the analyses conducted with a molecular phylogeny that is complete and better resolved. In addition to increasing the taxonomic coverage of *L.*

(*Chilalictus*), more genes will need to be added in the phylogenetic analysis in order to produce a well-resolved molecular phylogeny.

This study suggested that genetic constraints might be hampering the evolution of polylectic to broadly polylectic behaviour in species of *L.* (*Chilalictus).* Therefore, broadly polylectic species would be expected to have genetic adaptations that enable them to overcome such constraints in order to exploit the diversity of pollen hosts that they visit. To evaluate this, it is necessary to understand the nature of the constraints. We investigated whether olfactory sensory ability could be one of the constraints and whether this could be reflected in the number of sensory organs on the antennae. However, no evidence for adaptations in the peripheral sensory morphology was found in the broadly polylectic species of *L.* (*Chilalictus).* Therefore structural adaptations for broad polylecty may occur at higher sensory levels than examined in this study and future research should examine these.

Furthermore, apart from morphological and neurological adaptations, behavioural, and digestive physiological adaptations have been previously suggested to facilitate broad polylecty in bees (Russell et al. 2017; Sedivy et al. 2011), and these should be investigated in future research, in similar ways as has been done for oligolectic species (Praz et al. 2008a; Praz et al. 2008b). Particular attention should focus on the attributes of those plant species that broadly polylectic species include in their diets and that polylectic species do not utilise. This approach could give an indication whether the nature of the constraints need to be searched in the attributes of the pollen and whether it is in the plants' evolutionary interest to impede pollen collection and digestion.

Plants have evolved colourful displays, nectar production and excess production of pollen to attract and accommodate pollinators. Of the floral resources, nectar is relatively cheap and renewable, as it is produced from carbon dioxide, water and sunlight, while pollen is more costly, as it contains amino acids and fat. Therefore, plants often encourage high visitation rates through nectar production. However, it is conceivable that there is selection to reduce the amount of pollen that is rendered unavailable for pollination as it is transported to bee nests and used as bee food. This amount can be very high: Schlindwein et al. (2005) estimated that 95.5 % of pollen produced by *Campanula rapunculus* ended up as larval food, while only 3.7% became available for pollination. Mechanisms for plants to reduce the pollen utilisation could be making pollen difficult to handle, e.g. because of size and or echination (Konzmann et al. 2019; Vaissière & Vinson 1994), or to digest, e.g. through toxicity (Müller & Kuhlmann 2008; Praz et al. 2008b). Experimental approaches have strongly suggested that specialists have found ways to overcome these constraints while broad polyleges may have partially overcome them by adding them to mixed diets (Eckhardt et al. 2014; Praz et al. 2008b).

Significance of this study

This study focussed on the evolutionary constraints of broadly polylectic behaviour in bees but the findings have implications for the conservation of the more vulnerable oligolectic species. The results of this study concur with a large number of studies that show that oligolectic species might have genetic constraints that limit their diet to a small number of pollen hosts. This in turn limits their geographical range size and population size, which causes reduced genetic variation in their populations (Zayed et al. 2006; Packer et al. 2005; Danforth, Ji, et al. 2003). This makes them vulnerable to

decreasing floral resources due to anthropogenic land use and natural disasters. In this study three species of *L.* (*Chilalictus)* are likely to be genetically constrained to foraging on the plant genus of *Wahlenbergia.* The genus *Wahlenbergia* is widespread in Australia and can have a long flowering time, thus locally providing a predictable source of pollen. However, the species does not grow in areas that are dominated by crops and weeds, and flowering is highly susceptible to rainfall (Clarke 2000). Combined with a narrow geographic distribution, these oligolectic species of *L.* (*Chilalictus)* might not be able to recover from population bottlenecks caused by a local loss, or temporary lack, of pollen hosts.

On the other hand, polylectic and broadly polylectic species are versatile pollinators of crops and native plants and have a wider geographic distribution. This makes them keystone species, maintaining stable pollination networks within their habitats (Martín González et al. 2010). However, given that both polylectic and broadly polylectic species might be genetically constrained from visiting certain plant species, specialist pollinators are very important for the conservation of specific plants (Willmer 2014). This highlights the importance of maintaining a diverse pollinator assemblage in building pollination net-works that are resilient to changes in the environment (Klein et al. 2007). Altogether, this study contributes to the body of knowledge on the life history of *L.* (*Chilalictus*) species, which are among the most important native pollinators of crops and wild plants in Australia.

References

Bernays, E.A., 1998. Evolution of feeding behaviour in insect herbivores. *Bioscience*, 48(1), pp. 35–44.

Burrell, A.S., Disotell, T.R. & Bergey, C.M., 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, 79, pp. 35–44.

Cane, J.H. & Sipes, S., 2006. Characterizing floral specialization by bees: analytical methods and a revised lexicon for oligolecty. In N. M. Waser, Ollerton Jeff, ed. *Plant-pollinator interactions: from specialization to generalization*. Chicago University of Chicago Press, pp. 99–122.

Clarke, P.J., 2000. Plant population processes in temperate woodlands in eastern Australia: premises for management. In R.J. Hobbs & Yates C.J. eds Temperate woodlands in Australia: Biology, conservation, management and restoration. Surrey Beatty and Sons, Sydney, pp. 248 – 270.

Cruaud, P. et al. , 2017. High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific reports*, 7,p. 41948.

Dalmazzo, M. & Vossler, F.G., 2015. Pollen host selection by a broadly polylectic halictid bee in relation to resource availability. *Arthropod-plant interactions*, 9(3), pp. 253–262.

Danforth, B.N., Conway, L. & Ji, S., 2003. Phylogeny of eusocial *Lasioglossum* reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). *Systematic Biology*, 52(1), pp. 23–36.

Danforth, B.N. & Ji, S., 2001. Australian *Lasioglossum* + *Homalictus* form a monophyletic group: resolving the "Australian enigma". *Systematic Biology*, 50(2), pp. 268–283.

Danforth, B.N., Ji, S. & Ballard, L. , 2003. Gene flow and population structure in an

oligolectic desert bee , *Macrotera* (*Macroteropsis*) *portalis* (Hymenoptera : Andrenidae). *Journal of the Kansas Entomological Society*, 76(2), pp. 221–235.

Eckhardt, M. et al. , 2014. Pollen mixing in pollen generalist solitary bees: a possible strategy to complement or mitigate unfavourable pollen properties? *The Journal of Animal Ecology*, 83(3), pp. 588–597.

Futuyma, D.J. & Moreno, G., 1988. The evolution of ecological specialisation. *Annual Review of Ecology, Evolution, and Systematics*.

Gaston, K.J. & Lawton, J.H., 1988. Patterns in the distribution and abundance of insect populations. *Nature*, 331(6158), pp. 709–712.

Gibbs, J. et al. , 2012. Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Molecular Phylogenetics and Evolution*, 65(3), pp. 926–939.

Haider, M. et al. , 2014. Phylogeny and floral hosts of a predominantly pollen generalist group of mason bees (Megachilidae: Osmiini). *Biological journal of the Linnean Society. Linnean Society of London*, 111(1), pp. 78–91.

Hebert, P.D.N. et al. , 2013. A DNA "barcode blitz": rapid digitization and sequencing of a natural history collection. *Plos One*, 8(7), p.e68535.

Janz, N. & Nylin, S., 2008. The oscillation hypothesis of host-plant range and speciation. In K. Tilmon, ed. *Specialization, speciation, and radiation. The evolutionary biology of herbivorous insects*. University California Press, pp. 203–215.

Klein, A.M. et al. , 2007. Importance of pollinators in changing landscapes for world crops. *Proceedings. Biological Sciences / the Royal Society*, 274(1608), pp. 303–313.

Konzmann, S., Koethe, S. & Lunau, K., 2019. Pollen grain morphology is not exclusively responsible for pollen collectability in bumble bees. *Scientific reports*, 9(1),p. 4705.

Martín González, A.M., Dalsgaard, B. & Olesen, J.M., 2010. Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity*, 7(1), pp. 36–43.

Müller, A. & Kuhlmann, M., 2008. Pollen hosts of western palaearctic bees of the genus *Colletes* (Hymenoptera: Colletidae): the Asteraceae paradox. *Biological Journal of the Linnean Society*, 95(4), pp. 719–733.

Packer, L. et al. , 2005. Conservation genetics of potentially endangered mutualisms: reduced levels of genetic variation in specialist versus generalist bees. *Conservation Biology*, 19(1), pp. 195–202.

Praz, C.J., Müller, A. & Dorn, S., 2008a. Host recognition in a pollen-specialist bee: evidence for a genetic basis. *Apidologie*, 39(5), pp. 547–557.

Praz, C.J., Müller, A. & Dorn, S., 2008b. Specialized bees fail to develop on non-host pollen: do plants chemically protect their pollen? *Ecology*, 89(3), pp. 795–804.

Russell, A.L. , Buchmann, S.L. & Papaj, D.R., 2017. How a generalist bee achieves high efficiency of pollen collection on diverse floral resources. *Behavioural Ecology*, 28(4), pp. 991–1003.

Schlindwein, C. et al. , 2005. Pollination of *Campanula rapunculus* L. (Campanulaceae): How much pollen flows into pollination and into reproduction of oligolectic pollinators? *Plant Systematics and Evolution*, 250(3–4), pp. 147–156.

Sedivy, C. et al. , 2008. Patterns of host-plant choice in bees of the genus *Chelostoma*: the constraint hypothesis of host-range evolution in bees. *Evolution*, 62(10), pp.

2487–2507.

Sedivy, C., Müller, A. & Dorn, S., 2011. Closely related pollen generalist bees differ in their ability to develop on the same pollen diet: evidence for physiological adaptations to digest pollen. *Functional ecology*, 25(3), pp. 718–725.

Slatyer, R.A., Hirst, M. & Sexton, J.P., 2013. Niche breadth predicts geographical range size: a general ecological pattern. *Ecology Letters*, 16(8), pp. 1104–1114.

Slove, J. & Janz, N., 2011. The relationship between diet breadth and geographic range size in the butterfly subfamily Nymphalinae - a study of global scale. *Plos One*, 6(1), p.e16057.

Sonet, G. et al. , 2018. Using next-generation sequencing to improve DNA barcoding: lessons from a small-scale study of wild bee species (Hymenoptera, Halictidae). *Apidologie*, 49(5), pp. 671–685.

Tasei, J.N. & Picart, M., 1973. Le comportement de nidification chez *Osmia* (*Osmia*) *cornuta* Latr. et *Osmia* (*Osmia*) *rufa* L. (Hymenoptera Megachilidae). *Apidologie*, 4, p.195 –225 .

Vaissière, B.E. & Vinson, S.B., 1994. Pollen morphology and its effect on pollen collection by honey bees, *Apis Mellifera* L. (Hymenoptera: Apidae), with special reference to upland cotton, *Gossypium Hirsutum* L. (Malvaceae). *Grana*, 33(3), pp. 128–138.

Vossler, F.G., 2018. Are stingless bees a broadly polylectic group? an empirical study of the adjustments required for an improved assessment of pollen diet in bees. In P. Vit, S. R. M. Pedro, & D. W. Roubik, eds. *Pot-Pollen in Stingless Bee Melittology*. Cham: Springer International Publishing, pp. 17–28.

Walker, K.L. , 1995. Revision of the Australian native bee subgenus *Lasioglossum* (*Chilalictus*). *Memoirs of Museum Victoria*, 55(1 & 2), pp. 1–423.

Wcislo, W.T., 1995. Sensilla numbers and antennal morphology of parasitic and non-parasitic bees (Hymenoptera : Apoidea). *International Journal of Insect Morphology and Embryology*, 24(1), pp. 63–81.

Wcislo, W.T. & Cane, J.H., 1996. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. *Annual Review of Entomology*, 41, pp. 257–286.

Willmer P. Climate change: bees and orchids lose touch. *Current Biology.* 2014 Dec 1;24(23):R1133–5.

Zayed, A. et al. , 2006. Increased genetic differentiation in a specialist versus a generalist bee: implications for conservation. *Conservation genetics (Print)*, 6(6), pp. 1017–1026.