THE UNIVERSITY
of LIVERPOOL

Bee Immunity:
A Multifaceted Investigation Considering Social, Genetic, Evolutionary and Microbial Ecological Factors

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by Lauren Mee.

March 2023

Dedicated to a dear friend,

whose memory lives on as a reminder

to always look for the good in life.

"Thankful for all the brightness

in life that can overshadow

any darkness"

Danielle Mc Laughlin

1989 - 2017

Acknowledgements

I would like to thank Seth Barribeau, the OG, for all his guidance and encouragement throughout this ordeal of a PhD. It has not been a smooth ride but I have appreciated your seemingly endless optimism as you dragged us both through. Thanks to Jonathan Bollback for stepping in and covering all the administrative support I needed after Seth's departure, and for trusting me with `sudo` access on the Babbage server. Thanks also to Greg Hurst and Andrea Betancourt for their input which helped me me get back on track when Covid and happenstance had derailed me.

Thank you to the various people who worked with the bees and understood the special highlights and stresses of doing so over the years. To Hawra and Rama, it was a tonic to be able to bond about our time in the red room.

Thank you to all of my fellow students over the years. You all made office 219 a very special place to work. Thank you to Fiona, Jordan, Bryony and Eve for always having time for tea in the early days when we were still finding our feet. Thanks also to Jack for being on this journey from way back in the undergraduate days. Special thanks to Simon who helped so much with my final chapter when I'd unwittingly entered a new field for my last few weeks: I would have really struggled without you. And a shout out to Hazel for being willing to be my personal photographer.

I would also like to express my gratitude to Heather Allison for her remarkable supportiveness and readiness to offer assistance during challenging times. You were twice a light in otherwise dark times, both in my undergraduate and during the making of this thesis, and I will never forget your kindness.

Thank you to my dear, dear friends for still putting up with me and checking in after this PhD ate up all my time and personality. Thank you to Natalie and Louise for all the Vibe dates and adventures. A very special mention for my flatwife Colette - Flat 6 was the port in the storm throughout two PhDs and a pandemic. Thank you for supplying soda bread and comic relief throughout.

I am forever grateful to my family for their love and support. Thank you to my uncle Jimmy for inspiring me to stay curious and humble, I wish you could still be here to celebrate the end of this journey. Thank you to my not so little brother Michael for your encouragement and for always cheering me on, but for also keeping me humble the way only siblings can. Especially grateful to my parents, Michael and Debra: without your financial and emotional support I would never have been able to finish.

Last, but not least, I would like to thank my long suffering partner Martin. At times you have literally carried me through this gauntlet and I could never articulate how much that has meant to me. Thank you for putting up with me when I'm being an absolute crank and making sure I'm fed and adequately walked. You're the best.

Thank you all x

Thesis Abstract

Bees have a limited number of immune genes relative to model insects, but they have managed to thrive and diversify globally. RNA-Seq studies have highlighted bee candidate immune genes that fall outside of canonical pathways and bee microbiomes have demonstrated strong protective effects against pathogens. This thesis explores these avenues to help widen our understanding of bee immunity. I first examine the patterns of adaptive evolution in canonical and candidate immune genes in 11 bee species, and assessed the potential interplay between sociality and immunity. I found that candidate genes exhibit similar levels of selection and genomic change to canonical immune genes, consistent with their potential role in the immune response and host-pathogen coevolution. I used RNA-Seq data to highlight differentially expressed genes in response to three different immune challenges across four Hymenopteran species. I found shared and species-specific immune response genes, including candidate genes already highlighted by previous work. I conducted a broad-scale survey of the microbial content of 18 bee genera that vary in social structure using transcriptomic data and a cloud-computing resource. I showed that the bacterial communities of bees are influenced by host location, phylogeny, and sociality, with higher diversity in genera with more complex societies. Notably, bacteria with known anti-pathogenic properties were present across social bee genera, suggesting that symbioses enhancing host immunity are important. Collectively, these results help build a holistic picture of bee immune potential and help explain possible mechanisms of how they may compensate for their depauperate immune gene repertoire. I finish by stressing the need to continue to research species from outside of the social corbiculates to fully understand the complexities of bee immunity.

Contents

Chapter 1

General Introduction

1.1 Preface

The aim of this chapter is to introduce key concepts necessary for understanding the motivations and ideas behind the work included this thesis. These fall largely under three categories: sociality, insect immunity and the bee microbiome, with some overlap. Each of these topics could be discussed at great length, but I have endeavoured to provide a concise overview to aid comprehension of the thesis. I begin with sociality in the insects - the accepted definitions, costs, benefits and how sociogenomics helps us understand evolutionary changes driving the transition to complex behaviours. I then touch on insect immunity in general, focusing more on what is known about the innate immune response in bees and a host of behaviours that constitute social immunity in bee societies. Next, I provide an overview of what is known about the bee microbiome and discuss the potential role it plays in supplementing bee immunity. I finish with the research aims, and how each data chapter endeavours to address these questions.

1.2 The evolution and ecology of social behaviour in the insects

1.2.1 Defining social lifestyles

Whether in the swarming masses of locusts or the intricate hierarchies of elephant herds, social lifestyles have evolved across a wide variety of species as a means of survival in challenging environments. Within insects, there are multiple examples of social living strategies that are mainly defined according to the presence or absence of three factors: cooperative brood care, overlapping generations, and the reproductive division of labour (Table 1.1, Michener 1969). When all three of these factors are present, the society is considered complex (or advanced) eusocial, the most elaborate form of sociality. Eusociality has independently evolved several times in insects, namely in the ants, bees, wasps and termites (da Silva, Jack 2021; Rehan and Toth 2015; Wilson and Hölldobler 2005).

Table 1.1: Social lifestyle categories and the behavioural traits that define them. Cells are coloured dependent on the presence of the column behaviour: dark shaded cells are present and fixed, light shaded cells are sometimes present and empty cells denote where the behaviour is absent. Modified from Rehan and Toth 2015.

Social Category	**Cooperative Brood Care**	**Reproductive Division of Labour**	**Overlapping Generations**
Complex Eusocial			
Primitive Eusocial			
Incipiently Social			
Subsocial			
Solitary			

Eusociality can manifest in both primitive and advanced forms (Table 1.1), with the former being either facultative (e.g. *Megalopta genalis* can form both solitary and social nests, Wcislo et al. 2004) or obligate (e.g. the bumblebees, which exhibit many features of complex eusociality, Kocher and Paxton 2014). Eusocial animals form colonies that contain at least two different member castes: one caste produces offspring and passes on their genes to the next generation, and the other foregoes or reduces individual reproduction in order to care for the offspring of the former.

In cases where the reproductive and worker castes display significant morphological distinctions, certain complex eusocial societies can be viewed as a unified superorganism. This analogy, first described by Wheeler 1911, likens the reproductive caste to the metazoan germ-line and the worker caste to the soma. In superorganisms, differential caste physical traits arise through caste-specific developmental pathways, and once established, these castes become irreversible (Boomsma and Gawne 2018). Considering this definition, only honey and some stingless bee species can be considered superorganisms in the bees (Boomsma and Gawne 2018).

1.2.2 The molecular basis of sociality: understanding social behaviour through the lens of sociogenomics

The study of sociogenomics aims to uncover the molecular mechanisms underlying social behaviour (Robinson et al. 2005). Some social lifestyles can be plastic, with taxa exhibiting both solitary and social living dependent on resources or geographical location (Davison and Field 2016; Gibbs et al. 2012; Kocher and Paxton 2014; M. P. Schwarz et al. 2007; Soro et al. 2010; Wcislo et al. 2004). The transition from separate organisms to a superorganism, however, is one of the major evolutionary transitions (Szathmáry and Smith 1995). This transition is facilitated by the domination of the gene pool by reproductive individuals in social insect societies. Consequently, selection shifts from the individual to the colonial level, leading social species to reach an evolutionary point of no return (Boomsma and Gawne 2018).

There is much debate about how this transition to superorganismality occurs, but the prevailing theories (and increasing evidence) can be categorised as occurring due to either linear or nonlinear molecular mechanisms (Sumner et al. 2022). Linear mechanisms include those that drive evolution via small, incremental processes such as gradual change in transcription leading to changes in gene regulation or point mutations driving protein evolution. Evidence suggests that complexity of gene regulation using a conserved "genetic toolkit" is positively correlated with social elaboration rather than there being sociality-associated genes that evolved *de novo*

in different social taxa (Berens et al. 2015; Kapheim et al. 2015; B. Qiu et al. 2018; Rehan and Toth 2015; Shell et al. 2021; Toth and Robinson 2007; Toth and Rehan 2017). Similarly, comparing transcription between reproductive/worker castes in species of varying social lifestyles has shown that the more complex the social background of the taxa, the greater the difference in gene regulation (Arsenault et al. 2018; Patalano et al. 2015; Rehan et al. 2018; Standage et al. 2016).

On the other hand, nonlinear mechanisms would consist of large-scale genomic changes. These events could include whole genome duplication, gene family expansion and contraction, genome rearrangements (insertions or deletions) or disturbance of regulatory regions by transposable elements (Erwin 2000). Such events could be responsible for the "point of no return" in major transitions. One example of this is the "social chromosome" which determines the social organisation of *Solenopsis invicta* (J. Wang et al. 2013), but evidence of similarly dramatic genomic events in other social insects is scarce. It is likely that shifts in sociality and the behaviours that accompany it result from both linear and nonlinear mechanisms. With the growing availability of genomes from related taxa exhibiting varying social lifestyles, comparative studies have the potential to shed more light on the evolutionary history that has transformed eusocial insects from their more solitary ancestors.

To fully understand any aspect of social insect biology, it is crucial to consider its evolutionary context. In order to do so, it is important that we carefully choose the species to compare. Some taxa, while wildly successful, lack appropriate comparators. The ants, for example, exhibit the most advanced elaborations of sociality in insects, but as they are all highly eusocial, the appropriate comparison group is absent. Focusing elsewhere in Hymenoptera, however, can be very informative. Not only can every social lifestyle still be found extant but some single origins have led to monophyletic groups that exhibit a whole spectrum of social behaviours. For example, Apidae (bees) and Vespidae (wasps) exhibit all social states, from solitary to advanced eusociality (Shell et al. 2021; Toth and Rehan 2017). The main emphasis of this thesis is on comparative analyses of multiple bee species in order to gain a

deeper understanding of immunity in bees.

1.2.3 The social network: exploring the advantages and disadvantages of living in groups

Social living can confer a number of advantages to insect societies. Social insects are capable of incredibly complex behaviours that can broadly be classified thus: 1) how they interact with and shape their environment, 2) their learning, memory and cognitive capabilities and 3) their physiology. Many social insects construct their own nesting environments and are capable of structures of architectural marvel (Sane et al. 2020), most famously in the case of the termite mounds that stand like skyscrapers in the African Savannah. Meanwhile, wasps and bees use a diverse range of building materials to construct their nests, ranging from wood (Figure 1.1), to wax, soil and plant fibres (Sane et al. 2020) and some ants use larval silk to construct vast arboreal nests from leaves (Crozier et al. 2010). Before humans, termites and ants were engaging in agriculture, with both practising fungiculture (Mueller et al. 2005) and some ant species keeping aphids to farm their honeydew (Nelson and Mooney 2022).

There is increasing evidence to suggest that social insects have impressive cognitive and learning abilities (Brown and Austin 2021; Chittka and Rossi 2022). Honeybees have their "waggle" dance that can communicate direction and distance of floral resources to hivemates (Biesmeijer and Seeley 2005; von Frisch, K 1946) and understand that zero is less than one (Cordes 2019; Howard et al. 2018). Some studies have shown ants using problem-solving to rescue entrapped hivemates (Nowbahari et al. 2009), bumblebees can be trained in "tool use" (Alem et al. 2016) and wasps have impressive facial recognition (Avarguès-Weber et al. 2017; Tibbetts 2002). Coordinated behaviour can also allow for complex hive defence against predators, such as in the case of Asian honeybees that surround an invading hornet and elevate their own body temperature to effectively cook the invader (Ono et al. 1995) Additionally, coordination of behaviours can minimise pathogen risk in a phenomenon known

Figure 1.1: *Vespula* wasps make nests by combining masticated wood fibres with their saliva to make a papery building material. This foundress was spotted chewing on a bench in Ness Botanic Gardens, United Kingdom, in spring 2018 (photo credit: Lauren Mee).

as "social immunity" (discussed below, Cremer et al. 2007). Social living can also dramatically increase a species' potential lifespan (Lucas and Keller 2020), with some ant queens living up to 30 years (Keller 1998). Lastly, social living - and the consistent and constant interactions it imparts - allows for the vertical transmission of microbial partners across generations, allowing the coevolution of host-microbe symbiotic relationships (Lombardo 2008; Sanders et al. 2014; Zhang and Zheng 2022).

An obvious drawback of eusociality for worker castes is the sacrifice of individual reproduction in order to care for and allow the propagation of the reproductive caste. Furthermore, in eusocial insect societies, all individuals originate from a single pair or a limited number of reproductive individuals. As a result, relatedness within the

colony is high, while genetic diversity is low. Social insect communities are therefore perfect breeding grounds for pathogens (Schmid-Hempel 1995), where the relatedness of individuals and the frequency of physical contact between conspecifics allows for rapid disease transmission (Hess 1996; Naug and Camazine 2002; Shykoff and Schmid-Hempel 1991). It is therefore not surprising that social insects are associated with a plethora of pathogens (Paul Schmid-Hempel 1998). As a response, eusocial taxa have evolved various strategies to counteract the threat of disease, which are discussed in detail below (Cremer et al. 2007; Cremer et al. 2018; Wilson-Rich et al. 2009).

1.3 The genetic underpinning of the insect innate immune response

1.3.1 Individual insect immunity: an overview

Insects rely on their cuticle, which serves as both an exoskeleton and a lining for the foregut, hindgut, tracheae, and reproductive ducts, as their primary line of defence against pathogens (Moret and Moreau 2012; Moussian 2010). Should this primary barrier be breached or circumvented, insects employ a sophisticated pathogen recognition system. For this to occur, signatures of invading pathogens (pathogen-associated molecular patterns or PAMPs), such as peptidoglycans, lipolysaccharides or β-1,3 glucans are bound by host pattern recognition receptors (PRRs). PRRs are diverse, including, but not limited to: C-type lectins (CTLs), thioester-containing proteins (TEPs), peptidoglycan recognition proteins (PGRPs), Gram-negative bacteria-binding proteins (which are synonymous for beta-glucan receptor proteins; GNBPs or GBRPs respectively), galectins and leucine-rich repeat containing proteins (LRRs). Each recognition protein has its own combination of pathogens it is able to bind to, and each induces particular immune responses. Some directly elicit immune effector responses, some activate immune signalling pathways, and some can induce both (Choe et al. 2002; Levashina et al. 2001). These pathways incorporate different

signalling cascades, with the most well described being the Toll, IMD, JAK/STAT and RNAi pathways (Brutscher and Flenniken 2015; Hillyer 2016; Obbard et al. 2006; Paul Schmid-Hempel 2005; W. Zhang et al. 2021).

Each pathway will eventually cause an immune effector mechanism, such as phagocytosis, autophagy, apoptosis, RNA interference, melanisation or the production of antimicrobial peptides (AMPs) that serve to neutralise or seriously hinder the invading pathogen (Hillyer 2016; W. Zhang et al. 2021). Some of these pathways are specialised at recognising and eliminating particular pathogens. For example, the different compositions of peptidoglycan characteristic of Gram-negative and Gram-positive bacteria will be picked up by different receptors in *Drosophila melanogaster* and activate the IMD and Toll signalling pathways, respectively (Bischoff et al. 2004; Leulier et al. 2003; Michel et al. 2001; Royet et al. 2011). What we know about the genes driving these responses we owe to extensive work in model insect species such as *D. melanogaster*, though experimental transcriptomics continues to expand our knowledge of immune responses in ever more diverse insect species.

1.3.2 Strength in numbers: exploring social immunity

As social insects exist in densely-populated, genetically homogeneous communities, the risk that pathogen exposure and spread poses is high. In order to counter this, social insects have evolved social immunity - a suite of behaviours that prevent or else slow infection (Cremer et al. 2007). Social immunity in its original definition - the set of disease-controlling behaviours specific to eusocial insects - is likely an aspect of the unique and complex behaviours social insects have evolved after selection moved to the colony rather than the individual level (Cremer et al. 2018). Social immunity includes behaviours such as allogrooming (Wilson-Rich et al. 2009), corpse removal (Sun et al. 2018), exclusion/sacrifice of/by the sick (Drum and Rothenbuhler 1985; Shorter and Rueppell 2012), immune priming through contact (Traniello et al. 2002), and the organisation and separation of colony functions and spaces (Pie et al. 2004; Schmid-Hempel and Schmid-Hempel 1993; Stroeymeyt et al. 2014).

Some of these behaviours can be linked to specific genes. For example, up-regulation of *malvolio* in the honey bee switches honeybees into foragers after infection, moving disease away from the colony heart (Antonio et al. 2008; Ben-Shahar et al. 2004). Honey bees also exhibit a hygienic behaviour that involves uncapping and removing dead larvae from cells, which has been associated with resistance to American foulbrood (Spivak and Reuter 2001). Interestingly, these behaviours can be selectively bred for, and the genetic factors that contribute to them have been at least partly identified (Harpur et al. 2019; Lapidge et al. 2002; Pérez-Sato et al. 2009; Rothenbuhler 1964).

There is reason to theorise that, as the major transition to sociality shifts selection from the level of the individual to that of the colony, that the evolution of these behaviours may lead to relaxed selection on personal immunity of insect society members. Though there is evidence of AMPs being more potent in social and group-living insects (Hoggard et al. 2011; Stow et al. 2007; Turnbull et al. 2011; Turnbull et al. 2012), other studies have found evidence of relaxed investment in individual immunity. A comparison of the encapsulation responses in solitary insects and their social relatives reveals that solitary species produce stronger responses across multiple insect orders (López-Uribe et al. 2016). Further, studies suggest that there have been contractions of immune gene families during the evolution of sociality in termites (He et al. 2021). Both of these studies - and population-level selection analyses in the honeybee (Harpur and Zayed 2013) - support the hypothesis that the development of social immunity leads to reduced investment in individual immunity in social insect species.

1.3.3 Bees exhibit a depauperate immune gene count in contrast to other model insects

In 2006 the first social insect genome was published, that of the honeybee (Honeybee Genome Sequencing Consortium and others 2006). Considering the estimated pathogen load of social insects (Schmid-Hempel and Schmid-Hempel 1993; Schmid-

Hempel 1995), particular interest was paid to the immunological portions of the genome (Evans et al. 2006). Compared to solitary Dipterans, the number of immune genes in the honeybee were drastically reduced (Evans et al. 2006) and, similarly, ants were later observed to have this immune gene paucity too (Gadagkar 2011; Gadau et al. 2012). Though this initially seems to further support the theory of sociality relaxing selection on individual immunity to the extent of gene loss, later work revealed this depauperate immune gene repertoire to be a characteristic of bee genomes that predated sociality, rather being a consequence of it (Barribeau et al. 2015). If a depleted immune gene set is ancestral in the bees, then this gene loss cannot be due to the major transition to complex eusociality and relaxed selection as discussed above (Szathmáry and Smith 1995), but instead may simply be a characteristic of bee genome architecture.

This then raises the main question of this thesis: what else may be happening in the bees to allow them to compensate for this apparent lack in immune genetic potential? As discussed above, our current knowledge of insect immunology and immune protection genes largely comes from extensively researched model species such as *D. melanogaster*. Immune function is then assigned via orthology to species separated from bees by approximately 300 million years of divergence (Hennig 1981), making it highly probable that taxon-specific genes are overlooked (Otani et al. 2016; Sackton 2019).

When the first genomes of *Nasonia* became available it was noted that, like the honeybee, the parasitoid wasp exhibited a paucity of immune genes relative to other model insects (Wurm and Keller 2010). However, a suite of wasp-specific immune response genes were later identified in *Nasonia vitripennis* when immune challenged (Sackton et al. 2013). Similarly, work in the honeybee has highlighted a wealth of candidate, non-canonical immune genes (Alaux et al. 2011; Doublet et al. 2017; Richard et al. 2012). Perhaps, then, the genetic potential of the bee immune response is not as depauperate as first investigations suggest?

1.4 The bee microbiome and its importance for pollinator health

1.4.1 The characterised corbiculate core microbiota

The composition of an animal's microbiome can have profound effects on its health and well-being. In humans, relatively recent work continues to highlight the role microbiomes play in both health and disease states (B. Wang et al. 2017). In insects, microbes can wield massive influence on the biology of their host, affecting how the animal feeds (Andersen et al. 2012; Cheng et al. 2019), reproduces (Bourtzis et al. 1996; Singh and Linksvayer 2020; Werren et al. 2008) or fights infection (Benoit et al. 2017; Bian et al. 2010; Duplouy et al. 2015).

The obligate eusocial bees - honeybees, bumblebees and stingless bees - have a well-characterised, shared "core" microbiome (Hammer et al. 2021; Kwong and Moran 2016b; Kwong et al. 2017a; Martinson et al. 2011; Raymann and Moran 2018) that is thought to play important roles in many aspects of bee health (Raymann and Moran 2018). This corbiculate core consists of a limited number of genera (Table 1.2): *Snodgrassella*, *Gilliamella*, *Bifidobacterium* and two clusters of lactic acid bacteria: *Bombilactobacillus* (previously *Lactobacillus: Firm-4*) and *Lactobacillus: Firm-5* (also referred to as *Lactobacillus* near *melliventris*). While this core set of bacteria are found consistently in honeybees, bumblebees and stingless bees (Kwong et al. 2017a), the abundances, compositions and exact species can differ (Table 1.2). There are also host-specific species. For example, *Frischella* and *Bartonella* seem mostly confined to *Apis*, and *Bombiscardovia* and *Candidatus Schmidhempelia* to *Bombus*.

Table 1.2: Overview of the corbiculate core and other associated microbes as identified in Kwong et al. 2017a.

Microbe Taxa	First Isolated	Ref(s)
	Corbiculate Core	
Snodgrassella		
S. alvi	Honeybee and bumblebee guts	Kwong and Moran 2013
S. gandavensis, *S. communis*	Bumblebees	Cornet et al. 2022
Gilliamella		
G. apicola	Honeybee and bumblebee guts	Kwong and Moran 2013
G. bombicola, *G. intestini*, *G. bombi*, *G. mensalis*	Bumblebee gut	Praet et al. 2017
G. apis	Honeybees	Ludvigsen et al. 2018
Bombilactobacillus		
B. mellis, *B. melllifer*	Honeybee stomach	Olofsson et al. 2014
B. bombi	Bumblebee gut	Killer et al. 2014b
Lactobacillus: Firm-5		
L. helsingborgensis, *L. melliventris*, *L. kimbladii*, *L. kullabergensis*	Honeybee stomach	Olofsson et al. 2014
L. apis	Honeybee stomach	Killer et al. 2014a
L. panisapium	Honeybee bread	C. Wang et al. 2018
L. bombicola	Bumblebee gut	Praet et al. 2015b
Bifidobacterium		
B. asteroides	Honeybees	Scardovi and Trovatelli 1969
B. commune	Bumblebee digestive tract	Killer et al. 2009
B. bohemicum, B. actinocoloniiforme	Bumblebee gut	Killer et al. 2011
B. commune	Bumblebee gut	Praet et al. 2015a
	Other Corbiculate Associates	
Apibacter		
A. adventoris	Honeybee gut	Kwong and Moran 2016a
A. mensalis	Bumblebee gut	Praet et al. 2016
Apilactobacillus		
A. kunkeei	Grapes, wine, flowers and honey	Edwards et al. 1998; Endo et al. 2012
A. apinorum	Honeybee stomach	Olofsson et al. 2014
A. bombintestini	Bumblebee gut	Heo et al. 2020; Mattarelli et al. 2021
A. apisilvae	Stingless bees	Oliphant et al. 2022
Commensalibacter		
Strain AMU001	Honeybee gut	Siozios et al. 2019

Continued on next page

Table 1.2: continued

Microbe Taxa	First Isolated	Ref(s)
Bombella/ Parasaccharibacter		
Bo. apis / P. apium	Honeybee larvae / honeybee midgut	Corby-Harris et al. 2014; Smith et al. 2019; Yun et al. 2017
Bo. intestini	Bumblebee crop	Li et al. 2015
Bartonella		
Ba. apis	Honeybee gut	Kešnerová et al. 2016
Frischella		
F. perarri	Honeybee gut	Engel et al. 2013
F. japonica	Eastern honeybee gut	Wolter et al. 2021

The corbiculate core bacterial community is strain-rich and species-poor (Ellegaard et al. 2015; Engel et al. 2014). Notably, some strains from *Apis* bees exhibit as much genetic variation from each other as they do from strains found in *Bombus* bees (Engel et al. 2014). The shared presence of these core microbes in bumblebees, honeybees, and stingless bees suggests an ancient origin of this community, presumably established prior to the divergence of these three obligate eusocial corbiculates around 55 million years ago (Peters et al. 2017). It can be hypothesised that this last common ancestor (LCA) was living socially, and that contact between parents, offpsring and nestmates facilitated transmission of these microbes across generations, allowing for the coevolution of a stable and consistent microbial community alongside the early corbiculates.

1.4.2 Protection from non-host sources: the microbiome as an extension of host immunity

Koch and Schmid-Hempel 2012 found that the microbial community of bumblebees played more of a role in host resistance to infection by the trypanosomatid *Crithidia bombi* than did the genotype of either host or parasite. Since then, there has been numerous studies looking into the role the standard bee microbial taxa described in the corbiculates play in bee immune defense, both as a community or individually (Table 1.3). The importance of the microbiome when considering bee health appears

to be substantial (Engel et al. 2016; Raymann and Moran 2018; R. S. Schwarz et al. 2015).

Table 1.3: Examples of bee microbial taxa that protect against infection.

Microbe Taxa	**Effect**	**Reference(s)**
Apibacter	Resistance to *Crithidia*	Mockler et al. 2018
Apilactobacillus	Resistance to *Paenibacillus larvae*	Butler et al. 2013; Forsgren et al. 2010; Kačániová et al. 2020; Kiran et al. 2022
	Resistance to *Nosema*	Arredondo et al. 2018
	Resistance to *Melissococcus plutonius*	Endo et al. 2012; Endo and Salminen 2013; Vásquez et al. 2012; Zendo et al. 2020
	AMP production	Dyrhage et al. 2022; Zendo et al. 2020
Bombella / Parasaccharibacter	Antifungal properties	Miller et al. 2021
Gilliamella	Resistance to *Crithidia*	Mockler et al. 2018
Lactobacillus: Firm-5	Resistance to *Crithidia*	Mockler et al. 2018
	Resistance to *M. plutonius* and *P. larvae*	Killer et al. 2014a
Lactic Acid Bacteria	Inhibitory against *M. plutonius*	Vásquez et al. 2012
Snodgrassella	Resistance to *Serratia marcescens*	Horak et al. 2020
Entire community	Resistance to *Crithidia bombi*	Koch and Schmid-Hempel 2012
	Resistance to *Lotmaria passim*	Raymann et al. 2017

The presence of a healthy, established microbiome can protect individuals from pathogen invasion and persistence in a phenomenon termed "colonisation resistance" (Lawley and Walker 2013). This is where the resident microbial community of an organism contributes to the host being inhospitable to invading pathogens. The mechanisms behind this phenomenon include the priming of the host immune system (Horak et al. 2020; Kwong et al. 2017b; Lang et al. 2022; Näpflin and Schmid-Hempel 2016; Sadd and Schmid-Hempel 2006). During this process the symbiotic microbiota activate host immunity at a nominal level which in turn renders the host more resistant to pathogen threat (Prigot-Maurice et al. 2022). Horak et al. 2020, for example, found that the presence of the symbiont *Snodgrassella alvi* in the honeybee gut enhanced host AMP production, and increased survivability post infection with

Serratia marcescens.

Otherwise, symbionts may protect their host from pathogenic threat via direct antagonistic interaction, such as the production of antimicrobial substances (Dyrhage et al. 2022; Endo and Salminen 2013; Koch and Schmid-Hempel 2012; Steele et al. 2017; Vásquez et al. 2012; Zendo et al. 2020). For instance, the symbiont *Apilactobacillus kunkeei* has been found to confer resistance to multiple pathogens (Table 1.3) including *Melissococcus plutonius*, the causative agent of European foulbrood. A bacteriocin, named kunkecin A, was isolated from *A. kunkeei* and found to exhibit high antibacterial activity against *M. plutonius in vitro* (Zendo et al. 2020). This exemplifies a microbiological conflict wherein the symbiont may assume a direct protective role in defending its host against pathogenic invasion.

Resident bacteria can otherwise produce biofilms that act as physical barriers to pathogen establishment in occupied tissue (Horak et al. 2020; Kwong and Moran 2013; Martinson et al. 2012; Powell et al. 2016). Established microbiomes may also keep pathogens at bay as symbionts out-compete invasive microbes for resources necessary for pathogen establishment and spread (Prigot-Maurice et al. 2022).

Further to these anti-pathogenic qualities, a stable and established microbiome can help support host immunity by less direct means. An effective host immune response requires adequate nutritional support (Alaux et al. 2010; Cotter et al. 2011; Dolezal and Toth 2018; Negri et al. 2019) and the microbial community in the bee gut aids fermentation and digestion (Ellegaard et al. 2015; Engel et al. 2012; Kešnerová et al. 2017; Kwong et al. 2014; Kwong and Moran 2016b); indeed, germ-free bees tend to weigh less than those with intact microbiomes (Zheng et al. 2017). All these factors underscore the significance of a healthy microbiome for the overall health of eusocial corbiculates. The immune potential of these bees cannot be evaluated in isolation, but rather must be considered in the context of their microbial partners' composition.

1.4.3 What about the solitary bees?

Most of what is currently available in the literature regarding bee microbiomes are from studies using honeybees or bumblebees. While there are increasing numbers of studies of bees outside Apinae, there remains a paucity of work assessing the microbial community of solitary bees (reviewed in Voulgari-Kokota et al. 2019). Solitary bees lack the consistent social interactions of obligately eusocial corbiculates, which has implications for how their microbiomes are acquired and established and how, or if, they can be passed down through generations.

As outlined in Table 1.1, solitary bees do not care for their brood and contact between mother and offspring often ends at egg laying. However, passing on microbes orally via trophallaxis or through the physical nest environment is not the only way microbes can be vertically transmitted. *Wolbachia*, for instance, is an intracellular, maternally-inherited bacterium that manipulates insect reproduction so successfully that it is found in as much as 52% of insect species (Bourtzis et al. 1996; Weinert et al. 2015; Werren et al. 2008). Interestingly, *Wolbachia* has been found in a number of solitary bee species but is rare in the obligately eusocial corbiculates (Gerth et al. 2011; Gerth et al. 2015; Ramalho et al. 2021; Saeed and White 2015) and has been proposed as a driver of *Andrena* diversification (McLaughlin et al. 2023). It's relative absence from social bees may be due to a number of factors (Ramalho et al. 2021), including the protective potential of the established corbiculate microbiome.

Outside of reproductive manipulators such as *Wolbachia*, what are the drivers of solitary bee microbiomes? One would assume that they would be more dependent on the environment and the lack of social interaction would make them less stable across geographical and generational scales, and more hetereogeneous in general. Indeed, the bee's geographic location appears to be a major driver behind microbial community composition in studies that exclusively looked at solitary bees (Kapheim et al. 2021; Keller et al. 2013; McFrederick et al. 2017).

When considering the bees that fall into social lifestyles somewhere between solitary and eusocial the picture is less clear. In some cases environment remains the

most significant contributing factor behind microbiome composition (McFrederick and Rehan 2016; McFrederick and Rehan 2019) whereas others find the bee species to be a more powerful identifier (Dew et al. 2020). Yet others find both to be important drivers (McFrederick et al. 2012; Shell and Rehan 2022). These findings do not need to be viewed as contradictory but perhaps rather a reflection of the complexity of the different social lifestyles of the bees studied. Perhaps as bees become more social, social contact facilitated vertical transmission shifts the main microbiome acquisition route from the environment to conspecifics from the community. What this means for the immune response potential of these non-obligately eusocial bees remains unexplored.

1.5 Significance

Bees play a crucial role as pollinators (Figure 1.2), and the reproduction of many vegetable, fruit and other crops depend upon them and their activity (Klein et al. 2007). For example, the value of pollinator-dependent crops to the USA economy was estimated to be as high as $16 billion (Calderone 2012). Further, pollinator-dependent crops are estimated to be up to five times the value of pollinator independent crops (Gallai et al. 2009). Crops also have better yield and quality when available pollinator species are diverse (Bänsch et al. 2021) with wild bees often being the more efficient pollinators than honeybees (Garibaldi et al. 2011). Accordingly, there has been an increase in the use of managed pollinators and an increase in the diversity of species used for pollination globally (Osterman et al. 2021).

Despite their importance to ecosystems and human food security, wild bees are in decline (Biesmeijer et al. 2006; Cameron and Sadd 2020; Goulson et al. 2008; Goulson et al. 2015; Kosior et al. 2007; Potts et al. 2010). The factors leading to these declines are myriad: bees are accosted by climate change, pesticide use, pollutants and land cover change (reviewed in LeBuhn and Luna 2021). Beyond this, bees are also detrimentally affected by disease. As managed species such as honeybees and bumblebees have been transported around the world, as too have

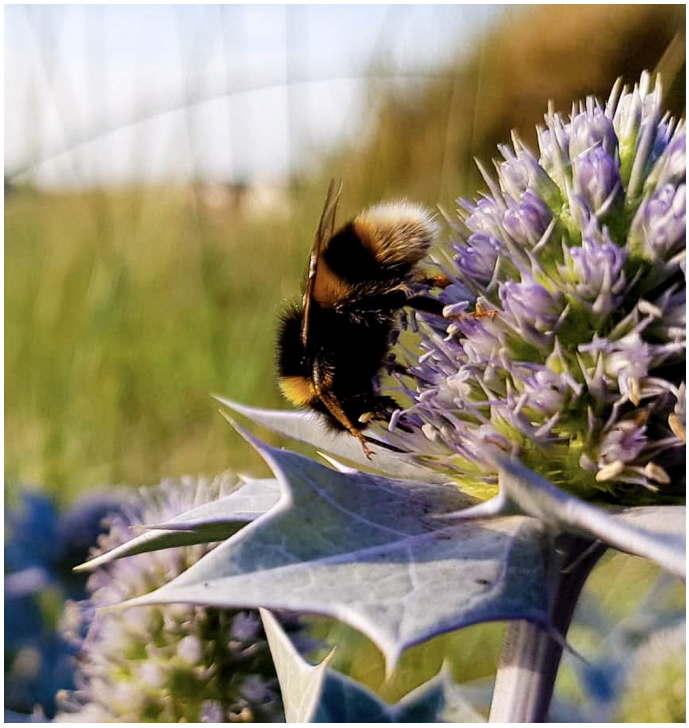

Figure 1.2: *Bombus terrestris* feeding on a sea holly plant near Crosby. As the bee moves from flower to flower, it unintentionally transfers pollen grains, facilitating the essential process of pollination (photo credit: Lauren Mee).

their parasites, which can run rampant amongst local naive pollinator communities, that lack any evolved resistance (Goka et al. 2001; Schmid-Hempel et al. 2014). Commercial bumblebees, for example, can harbour pathogens, potentially acting as disease reservoirs that threaten the health of wild populations (Colla et al. 2006; Martin et al. 2021; T. E. Murray et al. 2013; Schmid-Hempel et al. 2014). Similarly, viruses primarily associated with managed bee species have been found in local wild species (Alger et al. 2019; Manley et al. 2015; E. A. Murray et al. 2019; Schoonvaere et al. 2018) indicating spillover events.

Understanding the workings of the immune system of bees - including their behaviour and the interplay with their microbiome - can, therefore, help inform ways to resist disease. Supporting our bees in turn protects the economics of bee-associated production but, more importantly, the biodiversity and health of our environment which is especially important in a world of rapidly changing climate.

1.6 Thesis overview

1.6.1 Research aims

The purpose of this thesis is to investigate the immune system of the bee whilst considering evolutionary, social and microbial ecological factors. The depauperate immune gene repertoire is a feature of bees from all social lifestyles (Barribeau et al. 2015), and yet they have successfully spread globally and diversified into more than 20,000 species (Ascher and Pickering 2020). If bees aren't hampered by their restricted immune gene number, is this because they have evolved ways to compensate? Or is it perhaps an artefact of genome annotation, wherein assigning immune genes via orthology to Dipteran models means non-canonical, Hymenoptera-specific immune genes are being overlooked? What can we postulate about the microbial composition of bees, and how does this differ in the context of social lifestyle? The focus of this thesis will be to address and investigate these ideas and to contribute to our understanding of bee immunity.

1.6.2 Chapter 2: contrasting patterns of immune gene evolution across multiple origins of sociality

Doublet et al. 2017 used metatranscriptomics to identify a number of candidate immune genes in *A. mellifera* triggered in response to immune challenges that fell outside of the canonical immune gene set (informed almost entirely by homology to Dipteran model organisms). This chapter set out to assess the evolutionary context of the immune genes, both canonical and proposed candidate immune genes, by assessing patterns of positive selection across 11 bee species. The phylogeny used in this approach included two origins of sociality (transition from solitary to social variations) and two elaborations to advanced eusociality (in the honeybees and stingless bees). Branch-site tests of positive selection were used to then assess what evolutionary forces, if any, were present on the solitary, post-origin and post-elaboration of sociality branches and consider the differences between them. In

particular, how alike non-canonical immune genes sourced from Doublet et al. 2017 and canonical immune genes were was evaluated, with the intent of potentially lending support to their putative role in the bee immune response.

1.6.3 Chapter 3: using transcriptomics to elucidate the Hymenopteran immune response

This chapter was concerned with capturing the innate immune response of three bees (and one wasp) species by using, for the first time, a standardised protocol across species to assess immune activation. I utilised differential expression analyses to characterise the change in gene expression of each species when exposed to the same antigen exposures. I compared these responses across the four species to identify what, if any, genes were uniformly regulated in response to the immune challenges. Again, genes were then considered in context of whether they were canonical immune or else non-canonical, assessing their putative immunological function.

1.6.4 Chapter 4: comparing the microbial complements of bees from different social lifestyles

For this chapter I amassed 254 publicly accessible RNA-Seq samples from 18 bee genera and used a bioinformatic pipeline to identify non-host, microbial reads. Using this approach, I constructed a picture of the microbial community of diverse bees, supported by the pipeline's ability to recapitulate the corbiculate core microbial community in honeybees, bumblebees and stingless bees. I then used this data to assess whether location, phylogeny or social lifestyle were more important drivers of microbial composition. Furthermore, I identified microbial taxa associated with different bee tribes and examined their potential to support bee immunity.

Chapter 2

Selection analyses support putative immune function of non-classical immune response genes in the bees

---

2.1 Abstract

Hosts and their parasites and pathogens are locked in antagonistic coevolution. The genetic consequence of this can be seen in the rates of adaptive evolution in immunologically important loci in many taxa. As the risk of disease transmission increases we might also expect to see greater rates of adaptive evolution on genes of immune function. The evolution of sociality and its elaborations in insects represent enormous shift in disease transmission risk. Here, I examine whether sociality in the bees corresponds to changes in the rate of adaptive evolution in both classical canonical immune genes, and genes with putative immune functions identified from meta-analyses of honeybee transcriptomic responses to infection. I find that measures of gene-wide adaptive evolution is greater in both classical receptor genes and non-canonical candidate genes, and that branch-site adaptive evolution does increase with sociality regardless of gene category. I identify two genes with putative roles in immunity that warrant particular attention (Vitellogenin and a *Tret1-2* homolog).

There are more gene family changes after the origin of sociality across all gene classes - especially in candidate immune genes - with contractions occurring after the elaboration of sociality to complex eusociality. There are few genes or functions under adaptive selection that appear to be shared outside of specific lineages, suggesting that evolution of the immune system may be specific to individual species and their pathogen interactions.

2.2 Introduction

Bees vary widely in their social structure, from wholly solitary to advanced eusociality with distinct castes. The advantages conferred by sociality have contributed to ants, bees, wasps and termites becoming ecologically dominant in their respective habitats (Brady et al. 2006; Cardinal and Danforth 2011), but also carries greater risk of infectious disease. Insect societies can contain between tens to millions of genetically similar individuals packed into dense, interconnected communities that are ideal for pathogen transmission.

This increased risk of infection was presumed to exert strong selection for the maintenance and expansion of immune genes in highly social insects. However, the whole genome sequencing of honeybees found a surprising reduction in the number of immune genes relative to solitary insects (Evans et al. 2006; Honeybee Genome Sequencing Consortium and others 2006). One explanation for the depauperate individual immunity initially uncovered by genome sequencing is that eusocial insects benefit from a suite of behavioural defences that constitute "social immunity". These defences include the exclusion of sick individuals, allogrooming, hygienic and guarding behaviours, all of which help prevent the spread of disease within a social group (Cremer et al. 2007; Cremer et al. 2018; Dolezal and Toth 2014; Oxley et al. 2010; Wilson and Hölldobler 2005). These behaviours are absent in solitary insects and may have relaxed the need for extensive investment in individual protection, resulting in evolutionary losses of immune genes. However, subsequent sequencing of other Hymenoptera, including solitary species, suggest that the diminished immune

repertoire described in honeybees (Evans et al. 2006) preceded the origin of sociality, rather than being a consequence of it (Barribeau et al. 2015; Kapheim et al. 2021).

An alternative explanation for this reduction in immune genes is that we might simply be looking at the wrong genes. Much of what we understand about insect immunology and the genes that produce immune protection is drawn from well characterised model species, such as *Drosophila melanogaster*. While studies in model species have been instrumental in our understanding of insect immunity, Dipterans and Hymenopterans are separated by approx. 300 million years of evolution (Hennig 1981), and it is possible that lineage-specific genes and pathways are being overlooked (Otani et al. 2016; Sackton 2019). A number of recent studies have used transcriptomic or microarray experiments to describe the immune responses of honeybees after infection or immune activation (Alaux et al. 2011; Doublet et al. 2017; Richard et al. 2012), and find a great number of immune-responsive genes that are not part of the classical immune suite from model species. These non-canonical, candidate immune genes and pathways highlight potentially novel immune responses that may explain the scarcity of canonical immune genes gleaned from *D. melanogaster*.

While further experimental work is needed to confirm the immunological role of these non-canonical genes, we can take advantage of a feature of immune genes to infer which of these candidate genes are most likely to protect against infection. Hosts and pathogens are locked in antagonistic coevolution as any adaptation for host resistance exerts strong selection on pathogens, and *vice versa*. As a consequence, immune genes show elevated rates of adaptive evolution relative to non-immune genes (Obbard et al. 2006; Obbard et al. 2009; Sackton et al. 2007; Sackton 2020). How social structure and the commensurate greater risk of infection affects immune gene evolution remains unclear. While disease transmission risk is greater in denser populations - at its extreme in colonies of eusocial insects - the opportunity for social reduction of risk is also greatest.

Differing patterns of adaptive evolution has been found on conserved canonical

immune genes between advanced eusocial honeybees, primitively eusocial bumblebees and the solitary leaf cutting bee (Barribeau et al. 2015). Here I expand on this preliminary survey by analysing the canonical and non-canonical immune responsive genes to identify whether there are comparable patterns of evolution between the classic and candidate immune genes across bees, and whether the patterns differ among social and solitary clades.

Table 2.1: The 11 bee genome assemblies included in these analyses. Most transcriptomes and proteomes can be found on NCBI. *Melipona quadrifasciata* and *Lasioglossum albipes* are avaiable at Beebase (http://hymenopteragenome.org/beebase/). Species that are considered subsocial[1] or socially polymorphic[2] are categorised as facultatively social.

Family	**Species**	**Social structure**	**Assembly**	**Reference**
Apidae	*Habropoda laboriosa*	Solitary	Hlab1.0	Kapheim et al. 2015
	Ceratina calcarata	Facultatively social[1]	ASM165200v1	Rehan et al. 2016
	Bombus terrestris	Primitively eusocial	Bter1.0	Sadd et al. 2015
	Bombus impatiens	Primitively eusocial	BIMP_2.0	Sadd et al. 2015
	Melipona quadrifasciata	Complex eusocial	Mqua1.0	Kapheim et al. 2015
	Apis mellifera	Complex eusocial	AmelHAv3.1	Wallberg et al. 2019
	Apis florea	Complex eusocial	Aflo_1.0	Kapheim et al. 2015
	Eufriesea mexicana	Facultatively social	Emex1.0	Kapheim et al. 2015
Halictidae	*Lasioglossum albipes*	Facultatively social[2]	Lalb2.0	Kocher et al. 2013
	Dufourea novaeangliae	Solitary	Dnov1.0	Kapheim et al. 2015
Megachilidae	*Megachile rotundata*	Solitary	Mrot1.0	Kapheim et al. 2015

The bees as a group are ideal for examining the potential consequences of sociality on other genomic features, such as genes responsible for immunity, as 1) sociality evolved multiple times - there are at least three independent origins of sociality and two independent transitions to superorganismality (Boomsma and Gawne 2018; Kapheim et al. 2015; Rehan and Toth 2015; Rehan et al. 2016); 2) all social lifestyles are extant, sometimes within the same monophyletic group; and 3) genomic resources are rapidly expanding. Here I use the whole genome sequence of 11 species of bees (Table 2.1) with representatives from across Anthophila with multiple social lifestyles (Trapp et al. 2017, Figure 2.1). I then ask whether the patterns of molecular evolution of canonical and non-canonical immune genes change in tandem with events such as the introduction of social living or elaborations from eusocial societies to superorganisms with morphologically and developmentally distinct reproductive and

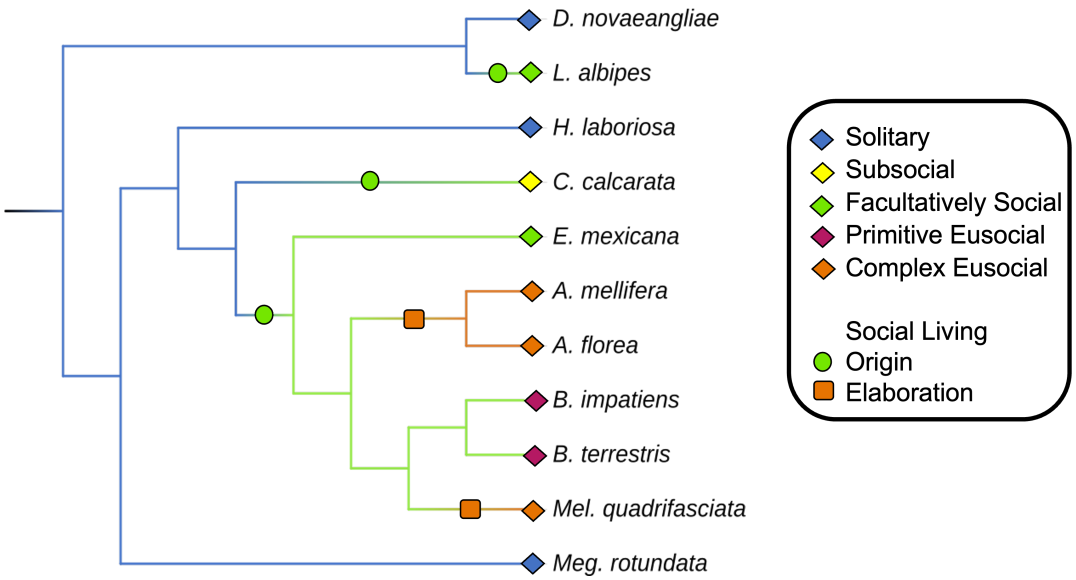

Figure 2.1: Phylogeny of the examined bee taxa. Branches are designated either solitary (blue), post an origin of sociality (green) or an elaboration to superorganismality (orange) event. These are the three categories of branch test used in the selection analyses to assess long-term shifts in selection. Branch lengths are not indicative of evolutionary time. Figure produced using iTol's web tool (Letunic and Bork 2021).

worker castes (Figure 2.1).

I examine three origins of sociality and the successive branches leading to five bees exhibiting various non-solitary lifestyles. These include the socially polymorphic *Lasioglossum albipes*, the subsocial *Ceratina calcarata*, two primitively eusocial bumblebees *Bombus terrestris*, *B. impatiens* and the facultatively social *Eufriesea mexicana*. I also consider two transitions from social to complex eusocial / superorganismality (Figure 2.1), which precede the honeybees *Apis mellifera* and *A. florea*, and the stingless bee *Melipona quadrifasciata*. Finally, the analysis also considers three solitary branches leading to *Dufourea novaeangliae*, *Habropoda labriosa* and *Megachile rotundata*.

I predict that if sociality increases evolutionary pressure on immune genes then immune genes of social clades will evolve more rapidly than in solitary clades (both in adaptive sequence evolution and gene family expansion and contraction). Further, if our candidate non-canonical genes indeed serve as functional immune genes they too will be subject to the same selective pressure as canonical immune genes and demonstrate elevated rates of adaptive evolution and gene turnover relative to genes

without immune function.

2.3 Materials and methods

2.3.1 Data resources

I used publicly available data from 11 bee species (Table 2.1). Whole genome sequencing derived transcriptomes and proteomes were downloaded from NCBI (NCBI Resource Coordinators 2018) or from BeeBase (Elsik et al. 2016). I divided genes into one of three categories - canonical immune genes, non-canonical candidate immune genes, and other genes with no known or putative immune function as a background comparison set. The canonical gene list (n = 279) was compiled from previous literature (Evans et al. 2006), the NCBI Biosystems pathway resource (Geer et al. 2010), and OrthoDB (Zdobnov et al. 2017). Each immune gene was given a function of either receptor, signalling or effector. The putative non-canonical immune genes I used were those found to be significantly differentially expressed across 19 transcriptomic studies of honeybees post immune activation (Doublet et al. 2017). Transcripts from these experimental studies were recovered from the current *A. mellifera* genome assembly (AmelHAv3.1) and translated into their protein products (n = 413). The background sample set was compiled from the rest of the *A. mellifera* genome with no known role in immunity (n = 9,229). Phylogeny (Figure 2.1) was taken and adapted from Rubin et al. 2019, originally based on work by Ramirez et al. 2010 and Branstetter et al. 2017.

2.3.2 Generating codon alignments

I translated all genes from the *A. mellifera* genome into protein products using `BLAST+` (Camacho et al. 2009). I consider these *A. mellifera* proteins as "anchor" proteins, against which other species are compared. Orthologs in the other 10 species were identified using reciprocal best-hit BLAST, keeping only those where a one-to-one ortholog was found across all species. For all genes with one-to-one orthologs across

all 11 species, I produced multiple sequence protein alignments using `MAFFT` (Katoh et al. 2019). Each protein ortholog was reverse translated (using `tBLASTn`) into nucleotide sequences. These MSA/fasta pairs were then used by `PAL2NAL` (Suyama et al. 2006) to produce codon alignments with gaps removed for downstream PAML analyses. Alignments from non-coding transcripts were removed. By the end of this process there were 186, 138 and 5640 alignments ready for `codeml` analysis in the canonical immune, non-canonical immune and background gene classes respectively. GC content was recorded per species per gene from each of these alignments as the proportion of G and C nucleotides of the sequence length. Assessing if there was an overall difference in GC content across different gene functions was determined using the Kruskal-Wallis rank sum test. I tested for differences in GC content between gene functions using Dunn's test of multiple comparisons with Bonferonni correction (Dunn 1964).

2.3.3 Positive selection analyses

First, to assess the overall evolutionary rate for each gene, I used the PAML (Yang et al. 1997; Yang 2007) program `codeml` (`model=0, NSsites=0, ncatG=1`) to estimate dN/dS ratios for each alignment. I assessed differences between evolutionary rate and gene function using the same approach as above when considering GC content. To examine whether the distinct origins of sociality and social elaboration alter rate of evolution on immunologically relevant genes, I then used branch-site likelihood methods for detecting long-term shifts in positive selection in `codeml` (Yang and Nielsen 2002; J. Zhang et al. 2005). In these models, only designated foreground branches are tested for positive selection, while all other branches in the tree are considered background. In different tests, the foreground branches were either solitary lineages, or branches following an origin or an elaboration of sociality. I consider three branches solitary, three distinct origins of sociality (one on the branch leading to the Xylocopinae, one in the corbiculate bees and one in Halictidae), and two separate elaborations to complex eusociality (one in the branch leading to the two

Apis species and one to the stingless bees, Figure 2.1). I considered each lineage separately and in combination with other lineages of the same sociality (*i.e.* all post-origin of sociality branches, all solitary branches etc.). I used the branch-site likelihood method for detecting signals of positive selection (J. Zhang et al. 2005). The resultant LRT statistic was compared against a χ^2 distribution ($d.f = 1$), at $\alpha = 5\%$. I adjusted p-values using the Holm-Bonferonni procedure to control for multiple testing (Holm 1979). Those with a p-value below 0.05 after correction were considered as being under a positively selected gene (PSG).

The number of PSGs with known canonical immune and non-canonical candidate immune function were compared against the number of PSGs in the background gene class using χ^2 tests. As there were a number of cases where PSG counts were too low to make χ^2 approximation appropriate, p-values were computed by Monte Carlo simulation with 10,000 iterations using `chisq.test` in R (4.1.2 Hope 1968; R Core Team 2020), with Benjamini-Hochburg corrections for multiple testing (Benjamini and Hochberg 1995). To assess trends in the proportion of genes under selection across socialities I used `prop.trend.test` (R 4.1.2).

As the branch-site test is vulnerable to assigning positive selection to changing codon sites that may have occurred due to multinucleotide mutation (MNM, Venkat et al. 2018), I also assessed the likelihood of codon change occurring due to more than one simultaneous nucleotide change. In order to do this, I used the `hyphy` package `FitMultiModel` (`FMM`, Lucaci et al. 2021). I fed my alignments and the gene trees predicted by `codeml` to `FMM`, which assesses the log likelihood of alignments being affected by instances of MNM. This produces an evidence ratio (ER) per codon, with values above 5 suggesting strong support for change in the codon occurring due to two- or three- hit phenomena (2H or 3H), depending on the model being assessed. I looked at both the 2H and 3H models for all genes that were considered under positive selection by the branch site tests, recording whenever sites under selection were likely to be codons of multiple differences (CMD). Sites were considered as likely under selection if they had a PAML-estimated Bayes Empirical Bayes posterior

probability over 0.9.

2.3.4 Gene family analysis

Beyond adaptive evolution at coding regions, I explored whether gene families of each category grew or contracted depending on social structure. To do so, SB (see Contributions 2.5.4) expanded the orthology grouping to include paralogs using `fastOrtho` (Davis et al. 2020) on the isoform reduced proteomes used earlier. Any gene families that had fewer than two constituent species were dropped from the analysis. Of the resultant 8777 gene families only those that contained at least one ortholog from *A. mellifera* were analysed (n = 7193). `CAFE5` (Mendes et al. 2021) was then used to statistically assess the evolutionary change in gene family number. `CAFE5` uses lambda (λ) as a measure of the probability of both gene gain and loss (assumed equally probable) per gene per time unit across the phylogeny (De Bie et al. 2006). λ was calculated for each of the three types of branch tests (solitary, post an origin of sociality, and post an elaboration of sociality, Figure 2.1).

2.3.5 GO analysis

I inferred gene function by assigning gene ontology terms to the isoform-reduced honeybee proteome (AmelHAv3.1) using `eggNOG`'s sequence mapper (version 4.5) set to default parameters (Huerta-Cepas et al. 2016). I then determined whether any of the gene ontology terms associated with PSGs of each gene class were significantly enriched against the complete proteome using `topGO` (Alexa and Rahnenfuhrer 2016) with Fisher's exact tests (version 2.46.0). GO terms were also marked as to whether they were immune associated or not.

2.4 Results

2.4.1 Evolutionary rate and GC content

In sum, I analysed 5964 genes for gene-wide evolutionary rate and compared these between the three canonical immune functions (receptor, signalling, effector), the putative non-canonical immune genes and the background set of genes (Figure 2.2A). There were significant differences between dN/dS ratios, the measure of evolutionary rate, between the gene classes ($P < 0.001$, $df = 4$, $\chi^2 = 982.07$, Kruskal-Wallis rank sum test, see **Appendix** 6.2: Table 6.1; also **S2.1**). All immune associated gene classes significantly differed from background (see **Appendix** 6.2: Table 6.1), except in the cases of background and effector ($P.adj = 1$, $Z = 0.4823$, Dunn's test).

In the post- origin and elaboration of sociality branch tests, GC content increases in genes under positive selection, but this pattern is not seen in solitary lineages (Figure 2.2B). The average gene GC content per species varied (see **S2.2**), with the greatest GC content in *C. calcarata* and *L. albipes* and the smallest in the two *Apis* species. Meanwhile, in a pattern similar to dN/dS ratios, mean GC content per species in each branch test category (solitary, post-origin and post-elaboration of sociality) was significantly higher in receptor immune and non-canonical immune genes across all branch tests (Figure 2.2C, see **Appendix** 6.2: Table 6.2; also **S2.3**).

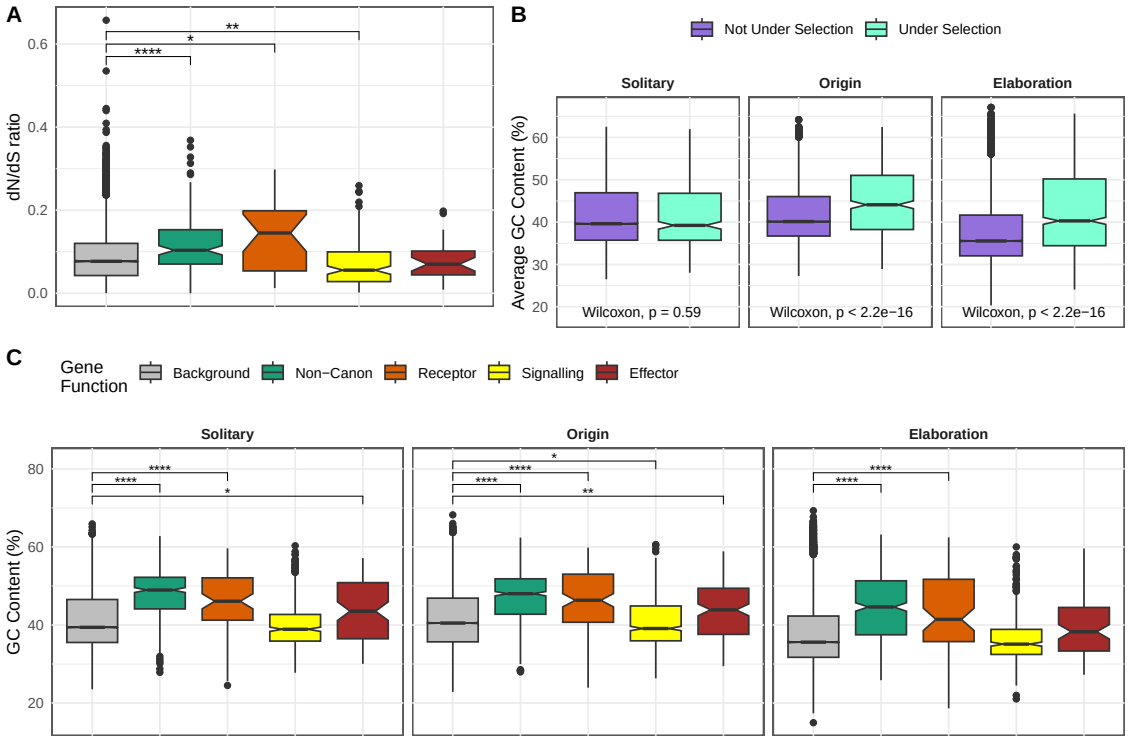

Figure 2.2: (A) Significant differences ($P < 0.001$, $df = 4$, $\chi^2 = 982.07$, Kruskal-Wallis rank sum test) were found between evolutionary rate and gene function. Differences in dN/dS ratios between immune/candidate immune and background genes were assessed with Dunn's test (see **S1**). Group level significant differences are indicated with $P < 0.05 =$ *, $P < 0.01 =$ **, $P < 0.001 =$ ***, $P < 0.0001 =$ ****. Receptor genes had a slightly higher evolutionary rate, but the most significant difference was between non-canonical immune and background genes. (B) Mean GC content was compared for genes under selection and those not under selection using the Wilcoxon rank sum test. No significant difference was found for solitary lineages, but a significant difference ($P = < 0.001$) was observed post-origin and elaboration of sociality. (C) Non-canonical immune and receptor genes consistently showed significant differences in GC content compared to background genes across gene functions per branch test category (see **Appendix** 6.2: Table 6.2; also **S2.2**).

2.4.2 Branch-site models of positive selection

I analysed each gene with 11 branch foregrounds (three combined by-branch test type, three solitary branches, three origins of sociality, two elaborations) resulting in 65,879 tests of positive selection (see **S2.4-2.6**) across the genome.

I tested for an enrichment of PSGs of canonical or candidate immune genes using the background genes to calculate an expectation of PSGs per lineage (see **Appendix** 6.2: Table 6.3 and Table 6.4). There were few canonical PSGs, so I assessed them as a group rather than separating them into the three sub-functions. There were a number of branch tests where no canonical immune genes were under selection: on the branches leading to *Ceratina* and the corbiculates, and the tests that included all branches post an origin or elaboration of sociality combined (see **Appendix** 6.2: Table 6.3). When all solitary branches were considered simultaneously, I found there were significantly more canonical PSGs than background ($\chi^2 = 10.561$, $P = 0.029$, χ^2 test with Monte Carlo simulation), but this result did not remain significant after correcting for multiple testing. There were no other cases where there were significantly more canonical immune genes - or non-canonical candidate genes - under selection than background in any of the branch tests (see **Appendix** 6.2: Table 6.3 and Table 6.4).

I do, however, find that the overall number of PSGs increases with social complexity of the branch tested. This pattern holds regardless of the category of gene examined (canonical immune genes: $\chi^2 = 7.8413$, $df = 1$, $P = 0.005$; non-canonical putative immune genes: $\chi^2 = 17.351$, $df = 1$, $P < 0.001$; background genes: $\chi^2 = 260$, $df = 2$, $P < 0.001$ [see **Appendix** 6.2: Table 6.5]). PSGs tend to be taxon specific with the most shared genes under selection occurring between Apidae and Halictidae (Figure 2.3).

In total, there were 42 out of the 186 canonical immune genes that exhibited signals of positive selection (see **S2.4**), some of which were shared among multiple lineages (Table 2.2). Dorsal, (*Dl*) was under selection in *Habropoda* and *Novaeangliae* branches and when all solitary lineages were considered simultaneously, but did not

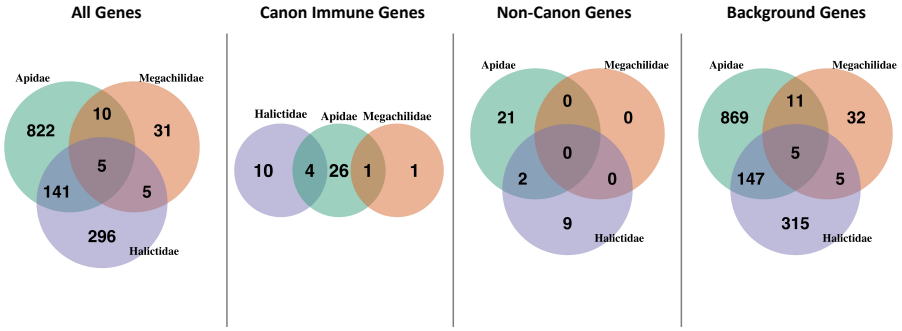

Figure 2.3: Shared genes under positive selection across the phylogeny overall, in canonical immune genes, and non-canonical candidate immune genes. There is some overlap between Apidae and Halictidae families.

show evidence of positive selection in any social lineages.

Table 2.2: Canonical immune genes under positive selection across multiple tested branches.

Gene	**Branch(es) Tested**	**Immune Function**
CTL4 (*LOC412825*)	*Habropoda*, *Lasioglossum*	Effector
Dicer-1 (*Dcr-1*)	*Habropoda*, *Melipona*	Receptor
dorsal (*dl*)	All Solitary, *Habropoda*, *Dufourea*	Signalling
LOC411115	*Lasioglossum*, *Apis*	Signalling
RPTOR (*LOC551668*)	*Habropoda*, *Megachile*	Signalling
ATG3 (*LOC552315*)	*Habropoda*, *Melipona*	Signalling
Tab *LOC726947*	*Lasioglossum*, Melipona	Signallng
LOC727634	All Solitary, *Megachile*	Signalling

In the non-canonical candidate immune class there were 32 PSGs from an initial group of 138 (see **S2.5**). There were only two genes under positive selection in more than 2 lineages. Vitellogenin (*Vg*) shows positive selection in the post-social origin lineages *Ceratina* and *Lasioglossum* and the solitary lineage *Dufourea*. A facilitated trehalose transporter *Tret1-2* homolog (*LOC413576*) showed positive selection in both branches leading to *Lasioglossum* and *Melipona*. Though *Vg* appeared to have sites under selection that were not considered likely to be due to MNM, all of *Tret1-2* homolog's positively selected codons were flagged as CMDs (see **S2.7-2.8**).

In total, there were 28040 codon sites that were assessed by `codeml` to likely

be under selection from the PSGs across the 11 branch tests. Of these, 1953 were identified by `FMM` as likely to be CMDs caused by a 2H phenomena, and 1525 by a 3H. When broken down by the category of branch test, those that lie post an elaboration of sociality event have considerably more CMDs than after the origins of sociality or in solitary clades (2H: 8.94% vs 4.63% and 4.96%, 3H: 6.70% vs 3.35% and 5.21% in the origin and solitary tests, respectively, see **S2.7-2.8**).

2.4.3 Gene family analysis

I find the highest probability (λ) of significant gene family change in non-canonical putative immune genes in the post-social origin lineages (Table 2.3), driven particularly by expansion (Figure 2.4). In the canonical gene sets, the greatest change consisted of contractions in the post-elaboration lineages (see **S2.9-2.10**).

Table 2.3: The most change in the canonical immune gene families largely occurred in the branches post an elaboration of sociality, driven mainly by gene family contraction. Overall, the highest rate of gene family expansion and contraction is in the post-origin of sociality non-canonical immune genes. These changes appear to be largely expansions (Figure 2.4). I denote the highest rate within branch test category by italics and the highest rate of change within gene category with boldface.

Model	λ **Solitary**	λ **Origin**	λ **Elaboration**
Background	0.078	**0.190**	0.163
Non-Canonical	0.112	***0.279***	*0.264*
Canonical	0.118	0.191	**0.223**
Receptor	0.129	0.195	**0.261**
Signalling	0.111	0.194	**0.231**
Effector	*0.164*	**0.185**	0.131

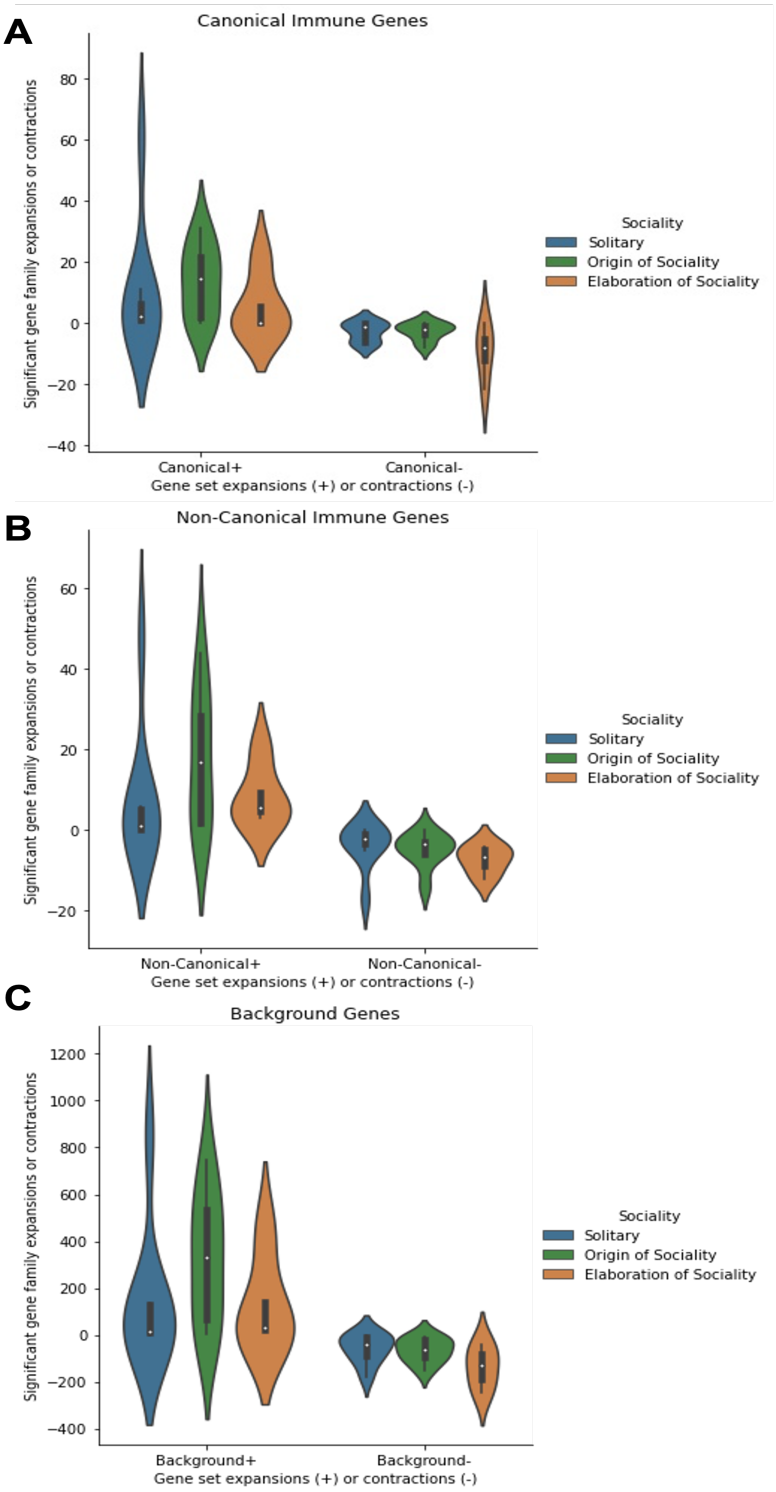

Figure 2.4: Post elaboration of sociality branches have higher rates of significant (A) canonical immune gene family contractions than other categories of branch tests, while post origins of sociality branches have more gene family expansions than solitary or post elaboration clades in (B) non-canonical gene sets. Y axis denotes the number of paralogs gained or lost per family.

2.4.4 GO analysis

When comparing across the three categories of branch test, the majority of enriched GO terms were unique to the combination of lineage and class (Figure 2.5). Within the canon immune PSGs four GO terms (GO:0031667, GO:0009991, GO:0006915, GO:0045860) were shared between all branch test type. Two of these terms suggest responsive functions (to nutrient levels and extracellular stimuli), one with positive regulation of kinase activity and one with apoptosis. There was no overlap between the non-canonical PSG's enriched GO terms between post-social origin, post-social elaboration and solitary branch tests, and nor were there common terms between canon and non-canon gene sets within each branch test category (see **S2.11-2.13**).

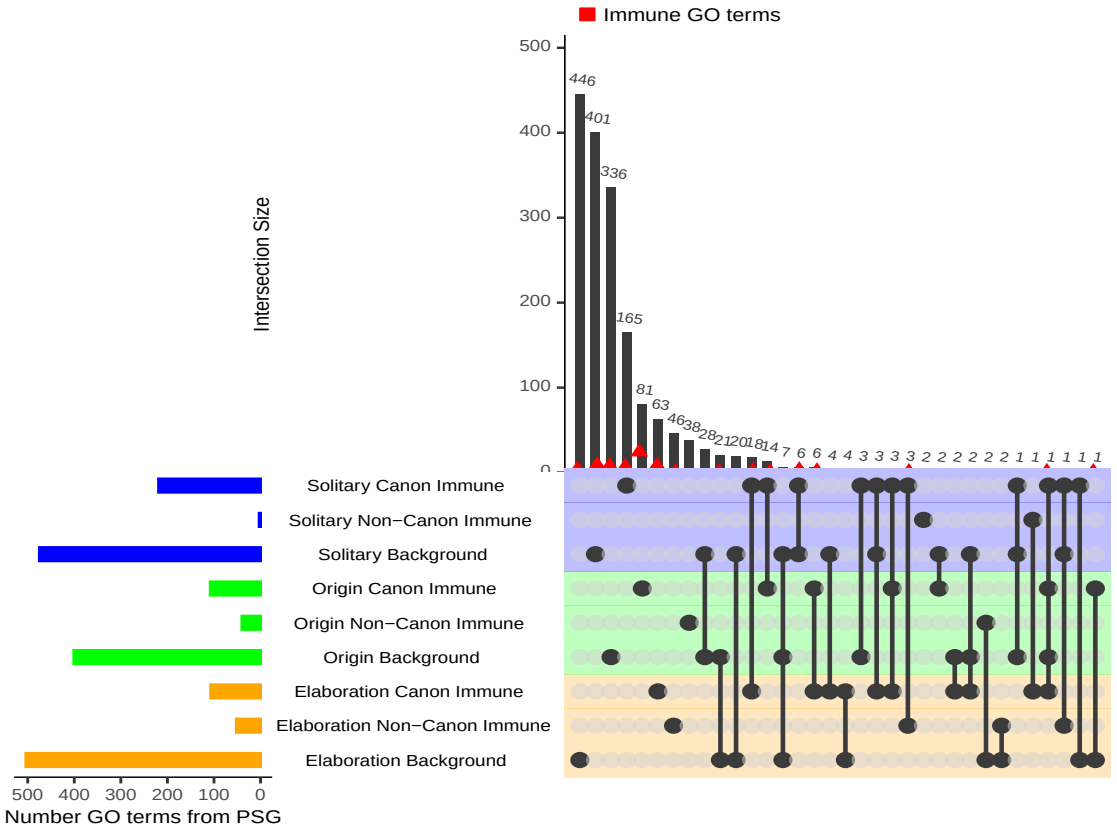

Figure 2.5: Though there are some shared enriched GO terms between the canonical immune PSGs of the three classes of branch test (highlighted in blue, green and yellow for solitary, post-social origin, and post-social elaboration branches, respectively) there is otherwise little overlap between enriched GO terms of candidate immune PSGs. There are more immune-associated GO terms found enriched in PSGs from the canonical immune gene set in post-social origin tests than any other test or gene class. Numbers of GO terms that are considered immune-associated are indicated by a red marker.

2.5 Discussion

2.5.1 Divergent patterns of positive selection on immune/candidate immune genes in solitary and social lineages

It is near dogma that genes important for immunity evolve more rapidly than other genes. This is intuitive given the requisitely antagonistic coevolution between hosts and pathogens generally, and in genes that affect infection and immunity specifically (Sackton 2019). This is seen in the elevated dN/dS ratios of immune receptor and signalling genes and the non-canonical immune genes in my analysis, all of which being significantly higher than that of background genes (Figure 2.2A). This lends weight to the hypothesis that these candidates may indeed be involved in the innate immune response of the bees, or, at the least, in the honeybees.

I also find that the non-canonical immune genes tend to have greater GC content (Figure 2.2C), again exhibiting a pattern similar to that consistently found in the canonical recognition genes. It is already documented that honeybees have very low GC content when compared to other insects (Honeybee Genome Sequencing Consortium and others 2006; Jørgensen et al. 2007), and here, too, I find the two *Apis* species with the lowest GC content across single-copy orthologs used in my selection analyses (see **S2.2**). Jørgensen et al. 2007 first described how there are both GC-poor and GC-rich regions across the honeybee genome, the latter of which appeared to be evolving slightly faster. Similarly, here, I see that there is higher average GC content in the PSGs of the post- origin and elaboration of sociality branches (Figure 2.2B) compared to genes not considered under selection.

DNA repair mechanisms during recombination events tend to preferentially use G/C nucleotides to fix breaks, a process known as GC-biased gene conversion (Duret and Galtier 2009). Over evolutionary time, this leads to a correlation between recombination and GC-content, which has already been extensively explored in the honeybee when mapping recombination (Beye et al. 2006; C. F. Kent et al. 2012; Wallberg et al. 2015). It is likely, therefore, that the elevated GC content detected in

canonical and candidate immune genes in this study are indicators of recombination.

Assessing GC content revealed a significant disparity between the GC content of PSGs and non-selected genes in non-solitary and superorganismal species, but not, interestingly, in solitary taxa (Figure 2.2B). This is likely due to the difference in recombination rates between solitary and social bee species (J. C. Jones et al. 2019). Eusocial insects - including honeybees, bumblebees and stingless bees - have all been found to exhibit very high levels of recombination compared to other insects and vertebrates (Beye et al. 2006; Kawakami et al. 2019; Waiker et al. 2021; Wilfert et al. 2007). This variation in recombination rates is likely influenced by differences in effective population size (Ne). Eusocial insects appear to have low Ne relative to non-eusocial species (Romiguier et al. 2014), which is expected given their high reproductive skew (Graur 1985).

Low Ne is known to increase linkage disequilibrium, and therefore the likelihood of deleterious alleles becoming linked to advantageous or vital genes in a phenomenon known as Hill-Robertson interference (Hill and Robertson 1966; Keightley and Otto 2006). Thus, the ability of selection to act on bringing advantageous mutations to fixation is reduced.

Recombination, however, can reduce interference between linked mutations and therefore restore the efficacy of selection (Hartfield and Keightley 2012; Kent and Zayed 2013; Webster and Hurst 2012). Indeed, recombination has been indicated as a particularly important driver of genome evolution in the honeybee (Wallberg et al. 2015). This is likely due to a feedback loop wherein elevated recombination rates in eusocial taxa are then selected for (Kent and Zayed 2013), which is likely behind the difference in patterns seen in this study (Figure 2.2B). Further, it has been proposed that, in the social bees, recombination is likely a driver of diversification of genes involved in worker differentiation (C. F. Kent et al. 2012; H. Liu et al. 2015) or immunity (Fischer and Schmid-Hempel 2005; Kerstes et al. 2012), the latter of which appear to be supported by my results.

Here, I found the highest average GC-content in the receptor and non-canonical

immune genes (Figure 2.2C). Finding such indications of diversification - via positive selection and recombination - in immune genes, particularly those involved in pathogen recognition, supports the hypothesis that genes involved in host-pathogen interactions experience more evolutionary change relative to genes of other functions due to a coevolutionary arms race (Fischer and Schmid-Hempel 2005; Kerstes et al. 2012; Sackton 2019). Similarly, non-canonical genes exhibit an overall increased evolutionary rate consistent with other immune genes. However, additional research is required to determine whether this is due to immunological pressure or their involvement in the transition to social behaviours, such as caste differentiation, which is another probable driver of evolutionary change (C. F. Kent et al. 2012; H. Liu et al. 2015).

I proposed that if the putative immune genes identified from transcriptomic studies had immune functions then these genes ought to be more commonly under positive selection. This was the case in some social lineages (*Apis*, *Melipona*, *Lasioglossum*, see **Appendix** 6.2: Table 6.4), but was not seen in any solitary lineages. In all solitary branch tests, however, the proportion of canonical immune genes under selection was greater than those in the background, though none of these differences were statistically significant after multiple testing correction (see **Appendix** 6.2: Table 6.3). This potentially supports the suggestion that sociality affords less reliance on canonical immune genes (Evans et al. 2006). Thus, as solitary bees invest more in individual immunity, I see greater evidence of adaptive molecular evolution at known immunological loci. On the other hand, the non-canonical genes may be subject to less constraint and are thus more likely to be utilised in the adaptation to social lifestyles and, perhaps, to increasing pathogen pressure.

My GO analysis revealed there was little overlap between enriched terms of PSGs from different combinations of branch test and gene class (Figure 2.5). Immune associated GO terms were most enriched in the canonical immune genes under selection post the three origins of sociality, supporting the theory that the advent of social interactions would bring with it increased pathogen risk and the need for

improved immune responses. This is in contrast to the relative lack of positive selection I identify in canon immune genes post origins of sociality. I postulate that though there is relatively fewer genes showing adaptive evolution after the origin of sociality, the genes that are under selection are those that have diverse immunological functions.

2.5.2 Immune-associated genes found under selection across multiple tested lineages

The canonical immune gene most frequently under selection was dorsal (*dl*). *Dl* has long been known to play a role in the innate immune response of insects, having been linked to anti-microbial, -fungal and -viral responses through activation of the Toll pathway (Bangham et al. 2006; Belvin and Anderson 1996; Ferreira et al. 2014; Sheehan et al. 2018; Silverman and Maniatis 2001). In Hymenoptera specifically, *dl* orthologs have been found to regulate expression of defensin (Anete Pedro Lourenço et al. 2018). *Dl* also plays a role in development, and so it cannot be said with absolute certainty that the pressure behind the positive selection in the *Habropoda* and *Novaeangliae* lineages were down to adaptive immune function.

In the non-canonical candidate immune genes, vitellogenin (*Vg*) is under selection in several lineages (social: *Lasioglossum* and *Ceratina*, solitary: *Habropoda*). Though it was not detected by my analysis, *Vg* has previously been found to be exhibiting high levels of adaptive evolution, both recent and ongoing, in various *Apis* species (C. F. Kent et al. 2011). *Vg* is an ancient yolk precursor protein found throughout the Metazoans and functions as a large lipid transfer protein important in a number of functions, including immune responses (Hayward et al. 2010; Salmela and Sundström 2018). *Vg* protects against oxidative stress (Havukainen et al. 2013; Seehuus et al. 2006) and is involved in trans-generational immune priming (Harwood et al. 2019; Sadd et al. 2005; Salmela et al. 2015), where pathogen resistance is transferred from adults to offspring (Milutinović et al. 2016). *Vg* is also down-regulated upon *Varroa* mite or viral infection (Alaux et al. 2011; Doublet et al. 2017; Anete P

Lourenço et al. 2009). Behaviourally, down-regulation of *Vg* results in workers switching from colony-based work to foraging, effectively removing the infection from the heart of the colony (Denison and Raymond-Delpech 2008; Nelson et al. 2007). Shifting worker tasks to minimise risk within the colony serves as a demographic and behavioural defensive adaptation which can be considered a component of social immunity. More directly, *Vg* can also recognise and bind pathogens (Shicui Zhang et al. 2011), suppress microbial growth, and is up-regulated in response to bacterial challenge in *Apis cerana* (Park et al. 2018). It is important however to note that *Vg* functions in immunological and non-immune processes. Thus, while I do see adaptive evolution of *Vg* in my tests, discerning the relative role of immune or other pressures in driving these coding changes remains challenging.

LOC413576, another multi-test non-canonical immune PSG, is a facilitated trehalose transporter *Tret1-2* homolog, referred to hereafter as *Tret1-2*. Trehalose is the main sugar found in insect heamolymph and *Tret1-2* is responsible for the trehalose release from the fat body and its subsequent incorporation into tissues in need (Kanamori et al. 2010). Trehalose is a source of nutrients for invading parasites in mosquitoes (K. Liu et al. 2013), and knocking down the transporter gene resulted in increased resistance to *Plasmodium* infection. In Hymenoptera, *Tret1-2* was found to be involved with stress resistance and the subsequent increase in longevity in ant queens (Harrison et al. 2021); and is down-regulated after bacterial injection (Viljakainen et al. 2018). In Doublet et al. 2017's meta-analysis of experimental transcriptomic studies, *Tret1-2* was particularly down regulated in honeybees infected with *Nosema* and Deformed Wing Virus (DWV) - individually and as a coinfection - Israeli Acute Paralysis Virus (IAPV) and *Varroa*. This suggests that *Tret1-2* may play a role in a general response to infection, possibly limiting pathogen access to resources.

2.5.3 Advent of sociality accompanies dynamic genome change

The number of PSGs detected was found to be positively correlated with the complexity of sociality in the tested branches, indicating a clear relationship between the branch test category and PSG detection (see **Appendix** 6.2: Table 6.5). This occurred regardless of gene category, and may be due to the increased recombination/genetic variation phenomena described above. Transitioning between social lifestyles, particularly when evolving into complex eusociality, would require dramatic genomic changes as genes and processes involved in reproduction, caste differentiation, longevity, communication and other processes beyond immunity take on new roles. Thus my results may reflect this overarching genome-wide change.

As branch-site tests can suffer from excess false positives based on multinucleotide mutations (Venkat et al. 2018), I used `hyphy`'s `FMM` (Lucaci et al. 2021) to identify potential codons of multiple differences (CMDs). I again found considerably more instances of CMDs on branches after the elaboration of sociality (see **S2.7-2.8**), indicating a much more dynamic genome environment. CMDs may still be a consequence of positive selective forces, but may also have become fixed due to neutral selection. As branch-site tests cannot distinguish between these two, I highlight these codons and the genes they occupy as less certain PSGs.

Beyond sequence level adaptive evolution, I further explored whether gene turnover - the expansions or contractions in the number of genes within a gene family - differ among gene categories and among the different categories of branch tests. I find substantially more gene family expansions in branches after the origin of sociality than either solitary species or species after the elaboration to highly eusocial societies (Figure 2.4). This is likely due to the host of new behaviours and adaptations necessary to facilitate a transition from solitary to social living. This is most pronounced in non-canonical immune genes, but I cannot say whether this is due to putative immune function or other processes associated with living in insect societies.

Strikingly, clades past elaborations to complex eusociality have more gene family contractions than solitary or social clades. The number of expansions within these

clades is greatest in the non-canonical immune genes and background gene sets (which recapitulates the findings of Kapheim et al. 2015). It is noteworthy that the pattern of increased gene family number in social taxa disappears post elaboration to advanced eusociality, where instead, I see contractions in gene families, most dramatically in the canonical immune genes (Figure 2.4, see **S2.9-2.10**). This may be that as complex eusociality introduces an evolutionary "point of no return" (Wilson and Hölldobler 2005), the continual loss of social plasticity translates to gene loss in particular families. It may well be that the presence of social immunity does indeed lead to a reduced arsenal of immune genes in the advanced eusocial insects, as other genes and pathways are utilised instead. These analyses suggests that in bees the core canonical immune genes do not vary much in sequence evolution relative to background genes in more social clades. The non-canonical genes, however, may be less constrained and more amenable to evolutionary change. Whether the changes that I do detect are driven largely by adaptation to pathogen pressure or other factors that accompany social living is open to further exploration.

With the use of RNA-Seq to understand insect immunity becoming more common, a complicated picture is emerging where immune responses are challenge-specific, recruiting genes existing outside of what is considered canonical in individual host-pathogen couplings (Doublet et al. 2017; Sackton 2019; Troha et al. 2018). Individual species may invest resources into specialised pathways dependent on the prevalence of certain pathogens or challenges in their unique ecological niches. Thus, it may be difficult to detect such patterns of adaptive selection at a broad phylogenetic and geographic scale. The likely diverse range of taxonomically restricted genes are difficult to capture without population level sequencing.

2.5.4 Conclusion

I predicted that increased sociality increases infectious disease risk and would thus drive the rate of adaptive evolution of the genes that encode the immune system writ large; including both the classical definition of the immune response, and

candidate physiological, demographic, or behavioural components. I find that the classical immune genes are evolving more rapidly in solitary bees, with non-canonical candidate genes changing more in the social branches. I also find more change in the coding genes of species that have evolved past either an origin or elaboration of social behaviour. I interpret this as the result of solitary bees needing greater investment into individual immunity, whereas social animals have more dynamic patterns of selection in areas other than immunity as they transition into different social lifestyles. I demonstrate that the candidate non-canonical immune genes and classical immune receptor genes have elevated evolutionary rate and a trend towards greater GC content. I propose that the non-canonical immune genes assessed here are good candidates for further investigation and suggest that future work hones in on specific host-pathogen challenges in closely related species or among populations to better unravel the genetic underpinning of Hymenopteran immunity.

Contributions

I want to emphasise the specific role that my supervisor, Seth Barribeau, played in this chapter, which involved conducting the `CAFE5` gene family analysis.

Chapter 3

Comparing Hymenopteran immune response in three bees and a wasp identifies five genes with shared expression

---

3.1 Abstract

Insects are the most abundant class of animals and, as such, represent an unparalleled resource for pathogens. The insect immune response is a complex series of barriers, recognition proteins, signalling cascades, and effectors that have evolved to prevent and counter infection. Much of what we currently understand about insect immunity comes from important work with *Dipteran* and *Lepidopteran* model species. The use of these model species has been largely successful but cannot completely describe non-model or taxa-specific immune adaptations. The recent development of affordable next-generation sequencing has led to an expansion of transcriptomic assays of immunity in diverse insects; however each employs distinct methods, exposures, doses and time points for capturing immune expression. To better characterise the Hymenopteran immune response collaborators and I challenged four species: a

honeybee *Apis mellifera*, a bumblebee *Bombus terrestris*, a largely solitary carpenter bee *Ceratina australensis*, and the wasp *Polistes lanio*. Treatment samples were either exposed to a standardised concentration of heat-killed Gram positive bacteria, Gram negative bacteria, or sterile injury with naive controls. RNA sequencing was used to capture transcriptomic response of immune activation after six hours. I identified differential expression of genes that were shared among these species or unique to individual taxa. I find five differentially expressed genes that are shared across comparisons. Among these are the immune genes *toll*, NF-κ-B inhibitor *cactus* and a serpin, as well as genes with less clear roles such as *ninjurin-1* and *Art3*. I find a particular reliance on nutritional immunity among the bees, and identify three candidate immune genes in the bees that warrant further investigation. Often, where I find overlaps in gene expression I also see variation in the patterns of expression, suggesting that the dynamics and patterns of gene transcription in immunity is complex, dynamic, and varies according to species.

3.2 Introduction

The extraordinary success of insects, being found on every continent including Antarctica (Basset et al. 2012; Larsen et al. 2017; Teets and Denlinger 2014) brings them into contact with diverse pathogens and parasites such as viruses (Asgari and Johnson 2010; Bonning 2019; McMenamin and Flenniken 2018; Nouri et al. 2018), bacteria (Fünfhaus et al. 2018; Vallet-Gely et al. 2008), fungi (Shang et al. 2015; Wang and Wang 2017), protozoans (Roditi and Lehane 2008; Sadd and Barribeau 2013) and macroparasites (Kraaijeveld et al. 2002; D. Lu et al. 2016; Pennacchio and Strand 2006). An insect's first line of defence is its cuticle, which encompasses the insect body as an exoskeleton and also lines the foregut, hindgut, tracheae and reproductive ducts (Moret and Moreau 2012; Moussian 2010). Should these barriers be breached, insects have numerous innate immune responses that have evolved to recognise and neutralise the invaders. Each of these responses can be categorised into three stages: pathogen recognition, immune signalling and the deployment of

immune effectors (Hillyer 2016; Paul Schmid-Hempel 2005; W. Zhang et al. 2021).

Most of what we understand about insect immunology comes from work in relatively few insect orders, predominantly Diptera and Lepidoptera. Work in Hymenoptera, arguably the most speciose order of the insects (Forbes et al. 2018), is considerably less mature, though strides have been taken in the lab-amenable *Nasonia vitrepennis* (Sackton et al. 2013), the commercially important honeybee, *Apis mellifera* (Alaux et al. 2011; Doublet et al. 2017; Richard et al. 2012), and some ant species (Gupta et al. 2015; Viljakainen et al. 2018; Yek et al. 2013). Often, these studies have found immune-activated genes that do not have homologs in Diptera, suggesting that immune responses can be order- or even species-specific.

Comparing transcriptomic responses between different experiments is possible, but challenging, as each study uses different challenge protocols leading to enormous sources of variation among data sets (Doublet et al. 2017; Sackton 2019; Troha et al. 2018). Therefore, in order to best compare across different taxa, experiments should be as uniform as possible; exposing animals to the same pathogens, in the same fashion, to the same pathogenic material, and capturing immune responses at the same time points.

Here, I compare the immune response of four Hymenopteran species (Figure 3.1) using a standardised challenge protocol and striving to minimize any variation among the four species. Each of the Hymenopterans - *Apis mellifera*, *Bombus terrestris*, *Ceratina australensis* and *Polistes lanio* - were subjected to the same immune challenges and I use RNA sequencing to characterise the changes in gene expression on immune challenge. I then took an orthogroup-level approach to identify shared and taxonomically distinct patterns of gene expression. This approach can identify both novel, taxon-specific immune adaptations and conserved Hymenopteran elements of immunity that may have been missed by previous surveys based on orthology and homology to Dipteran immune genes.

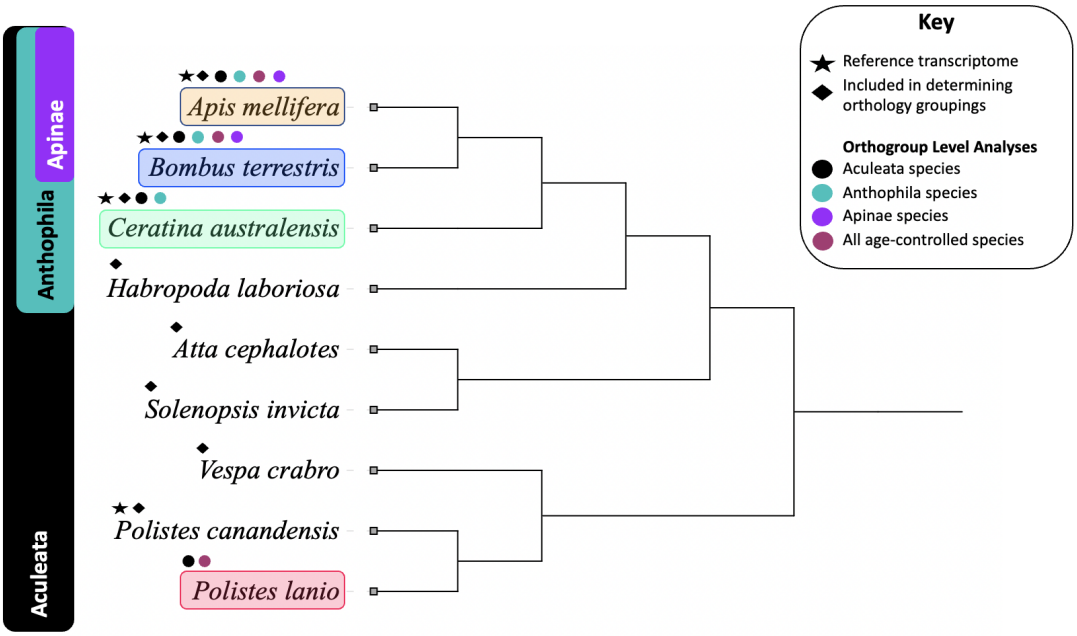

Figure 3.1: Phylogenetic relationship of the study species. Branch lengths are not indicative of evolutionary time; topography taken from Branstetter et al. 2017. Experimental species have been highlighted. Reference transciptomes which had reads mapped against are indicated by a star ($\star$), and all species included in the `OrthoFinder` (Emms and Kelly 2019) orthology sorting run are indicated with a diamond. The four combinations of species used in the orthogroup levels are indicated with coloured circles: all species (Aculeata, black); all bee species (Anthophila, teal); *Apis* and *Bombus* species (Apinae, purple) and all age-controlled species (*Apis*, *Bombus* and *Polistes*, red). Figure produced using iTol's web tool (Letunic and Bork 2021).

3.3 Materials and methods

3.3.1 Species information

Apis mellifera

Honeybees came from a single brood frame donated from Fera UK, delivered to the University of Liverpool. I checked the brood frame once per day for newly eclosed workers. Bees from each daily collection were allowed to feed on the honey from the frame before being housed in plastic cup bee houses that I had modified with chicken-wire sections to reduce internal humidity. Bees had constant access to 1:1 sugar-water (E.H. Thorne Ltd, Lincolnshire, UK). I kept the bees in a temperature controlled insectory at 27°C until five days post eclosure when they were collected

for immune challenge.

Bombus terrestris

I collected the bumblebees from a single colony delivered to the University of Liverpool (Agralan Ltd, Wiltshire, UK). The colony was kept in the provided housing box at 27°C. I monitored the colony twice daily for newly eclosed workers, which I removed and individually marked using a "Queen marking Kit" (E.H. Thorne Ltd, Product Code: Q0020) before returning them to the colony. Colonies were fed using pollen and sugar-water as above. I collected marked individuals on the fifth day post-eclosure for immune challenges.

Ceratina australensis

The carpenter bees, *C. australensis*, were field caught in Adelaide, South Australia by KO and SR (see Contributions 3.5.5). No permits were required for collection or export at this location. These bees were not able to be aged controlled.

Polistes lanio

The wasps in this study were collected in the field in Trinidad, under a Special Game Licence awarded by the Trinidad & Tobago Wildlife Section (2019) to RS (see Contributions 3.5.5). Wasps were aged upon emergence by taking daily censuses and marking newly emerged individuals with paint markers (Uni POSCA) on the thorax with different days corresponding to different colours. Individuals were then left in their nests until reaching five days old.

3.3.2 Immune challenge

Lab-reared animals and field-observed *P. lanio* were challenged 5 days after eclosure. *C. australensis* were caught in the field and not age-controlled. Each individual animal was randomly allocated to one of four treatments: naive, sterile wound, Gram positive and Gram negative. Naive animals were handled in the same way as the

other experimental individuals, but without challenge.

Individuals were anaesthetised on ice or by refrigerating them for short periods of time at -4°C. Treatment animals received an exposure via cuticle piercing between the teguments on the dorsal side of the abdomen. Injection sites were kept consistent per species. Size 1 insect pins were dipped in either sterile phosphate buffered saline (PBS), heat-killed *Staphyloccous lentus* or *Serratia marcescens* culture for the sterile wound, Gram positive and Gram negative treatments, respectively. Bacteria were cultured in LB broth at 23°C, quantified using optical density, spun, pelleted, and re-suspended in sterile PBS to reach the challenge dose of $1x10^8$ cfu/mL.

After being challenged, honeybees, bumblebees and wasps were kept in 50mL Falcon tubes with holes cut for air and ad libitum food and water (pollen and sugar water-soaked cotton wool for the bees, water-soaked cotton wool and mango for the wasps). *C. australensis* were individually housed in 100mm petri dishes. After six hours, all lab-reared and *C. australensis* individuals were snap frozen in liquid nitrogen and stored at -80°C. For *C. autralensis*, challenges were administered in the field and samples were put into RNAlater before being snap-frozen. *P. lanio* samples had abdomens removed and added directly to RNAlater. These were kept at 4°C overnight before being frozen. All field-caught samples were shipped at low temperatures in RNAlater to the University of Liverpool for RNA extraction.

3.3.3 RNA extraction, library preparation and sequencing

I kept samples at -80°C until extraction. Abdomens were homogenised in Trizol™ using a Qiagen Tissuelyser (Cat.No: 85220, 2007) and sterile Qiagen stainless steel 5mm beads (Cat.No: 69989) at 25Hz. Duration was 1.5 minutes for larger *Apis*, *Bombus* and *Polistes*, and 30 seconds for *Ceratina*.

Each homogenate was centrifuged at 12000 x g at 4°C for 5 minutes to remove insoluble material. I then added 0.2mL chloroform to each extracted supernatant, mixed well, and allowed samples to stand at room temperature for 10 minutes. These were then again centrifuged at 4°C for 15 minutes. I removed aqueous layers and

transferred these to fresh tubes along with equal amounts of 100% ethanol and mixed well. For the rest of the extraction I used Direct-Zol™ RNA MiniPrep Plus kits (Zymo Research) following manufacturer's instructions. RNA quality was checked using Agilent 2100 Bioanalyzer (Agilent Technologies).

Further quality checks, library preparation and sequencing was conducted by NovoGene UK on the Illumina HiSeq PE150. Ultimately, there were 12 samples per species - three per experimental condition - that were sequenced.

3.3.4 Pseudoalignment and orthology assignment

I trimmed paired reads using `Trim Galore` (version 0.6.6, a `CutAdapt` wrapper, Martin 2011). `Kallisto` (version 0.43.0) was used for read quantification (Bray et al. 2016). These reads were mapped to their respective reference transcriptomes, when available. *P. lanio* reads were mapped against a close relative (Table 3.1). As the *C. australensis* build lacked RefSeq gene identifiers, those of a close relative - *Ceratina calcarata* - were used when available, determined by reciprocal best hit (RBH) using `Blast+` (Camacho et al. 2009).

Table 3.1: The four reference transcriptomes used in this analysis are available via NCBI (Sayers et al. 2021). *P. lanio* reads were aligned against *P. canadensis*. RefSeq or GenBank genome build identifiers are included with each assembly.

Species	**Assembly**	**Accession**
Apis mellifera	AmelHAv3.1 (Wallberg et al. 2019)	GCF_003254395.2
Bombus terrestris	iyBomTerr1.2 (Darwin Tree of Life Project Consortium 2022)	GCF_910591885.1
Ceratina australensis	ASM430768v1 (Rehan et al. 2018)	GCA_004307685.1
Polistes canadensis	ASM131383v1 (Patalano et al. 2015)	GCF_001313835.1

To determine orthologous gene groups, the whole-genome sequence derived transcriptomes (cDNA sequence from every predicted gene) and proteomes (amino acid sequence) were reduced to one isoform per gene. I did this in two steps: first, validated RefSeq entries (NP/NR/NM) were prioritised above model RefSeq entries (XP/XR/XM). If there were no validated RefSeq entries, or multiple validated gene products, then the longest isoform per gene was assigned as its representative. I used these reduced transcriptomes and proteomes to identify orthologous groups using `OrthoFinder` (Emms and Kelly 2019). Four addi-

tional Hymenopteran genomes were added to improve orthology prediction: *Atta cephalotes* (AttaCep1.0: GCF_000143395.1, Suen et al. 2011), *Habropoda laboriosa* (Hlab1.0: GCF_001263275.1, Kapheim et al. 2015), *Solenopsis invicta* (UNIL_Sinv_3.0: GCA_016802725.1), and *Vespa crabro* (iyVesCrab1.2: GCF_910589235.1, Darwin Tree of Life Project Consortium 2022). Orthogroups determined by `Orthofinder` primarily came from the proteome sorting run. To produce a complete orthogroup list from across the genome, I then added groups of non-coding RNA transcripts from the `Orthofinder` transcriptome results. Gene trees were produced per orthogroup as part of the OrthoFinder pipeline.

To infer functions of differentially expressed genes and orthogroups I took advantage of *A. mellifera* annotations. I classified genes into three broad categories: 1) canonical immune genes based on their described role in the literature (Evans et al. 2006; Waterhouse et al. 2020) or orthology to known immune genes in the OrthoDB database (Zdobnov et al. 2017); 2) putative immune genes based on a meta-transcriptomic analysis of immune responses in honeybees (Doublet et al. 2017); and 3) any remaining genes that were considered non-immune. I assigned orthogroups one of these functional classes based on orthology to *A. mellifera* genes. Genes that were not assigned to orthogroups or had no clear ortholog in *A. mellifera* I classified as "un-annotated" (see **S3.1**).

3.3.5 Differential expression analyses

For the species level differential expression (DE) analyses, I converted `Kallisto` estimated transcript abundances to gene-level count tables per species using the package `tximport` (version 1.24.0, Soneson et al. 2015) in R (version 4.2.1, R Core Team 2020). Genes with counts < 10 across samples were filtered out of the tables. I then normalised transcript abundances using `DESeq2` (version 1.36.0, Love et al. 2014). I used principal component analyses (PCA), expression heatmaps, and sample to sample distance matrices to identify potential outliers. As I was attempting to identify conserved patterns of differential expression across considerable phylogenetic

distance, with limited sample sizes, I excluded samples that diverged markedly from treatment norms. As this may be removing genuine biological variation, I also ran all analyses without removing outliers and include these results in the supplementary materials for completeness (see **S3.2, S3.11**).

I used `DESeq2` to run differential expression analyses, contrasting the three treatment conditions to the naive un-manipulated individuals. I also assessed expression at the orthogroup level by merging species into single count tables per four combinations: 1) all four species, 2) all three bee species (excluding the wasp *P. lanio*), 3) *A. mellifera* and *B. terrestris* together and 4) all age-controlled species (*A. mellifera*, *B. terrestris* and *P. lanio*). I refer to these groupings as Aculeata, Anthophila, Apinae and Age-Controlled, respectively.

For each of these combinations, transcript abundance files were converted to orthogroup-level expression count tables using `tximport`, with `OrthoFinder` orthogroup designations in place of gene IDs. This was first done per species to allow `tximport` to quantify the transcript abundance per species and then these outputs were combined manually. Only orthogroups that had detectable expression from each of the species in each grouping were included (Aculeata, $n = 7198$; Anthophila, $n = 7648$; Apinae, $n = 8871$; Age-Controlled, $n = 8142$). Genes or orthogroups with $FDR < 0.1$ per contrast were considered differentially expressed "genes" (DEG).

3.3.6 GO enrichment analyses

I assigned gene ontology terms to each species' isoform-reduced proteome using `eggNOG` (version 5.0) with default parameters (Cantalapiedra et al. 2021; Huerta-Cepas et al. 2019) in R. I performed GO enrichment analysis using `topGO` (version 2.48.0, Alexa and Rahnenfuhrer 2016). For each species, terms from differentially expressed genes were compared against those of all genes that had not been filtered out by `DESeq2` to identify over-represented terms and associated functions. Enriched GO terms with a weighted Fisher $P < 0.05$ were considered significantly enriched.

3.4 Results

3.4.1 RNA extraction, pseudoalignment and orthogroup sorting

Sequenced samples had an average of 49 million raw reads (with a range of 39M:71M; see **Appendix** 6.3: Table 6.6). `Kallisto` (Bray et al. 2016) pseudoaligned 68.13%, 91.20%, 59.75% and 76.29% of the reads for *A. mellifera*, *B. terrestris*, *C. australensis* and *P. lanio*, respectively (see **Appendix** 6.3: Table 6.7). Final protein and transcript orthology sorting using `OrthoFinder` (Emms and Kelly 2019) assigned 96.0% and 82.4% of the input amino and nucleotide sequences to orthogroups (see **S3.1**).

3.4.2 Differential expression analyses

I find varying numbers of significantly differentially expressed genes across the species and conditions (see **Appendix** 6.3: Table 6.8; also **S3.3-3.6**) with no standard response in terms of whether genes were majority up- or down-regulated in any condition (Figure 3.2). In the bees, more immune (canonical or putative) genes were significantly differentially expressed than genes of other function in most instances. This pattern was not apparent in the wasp *P.lanio* (Figure 3.3).

The differentially expressed genes were largely species-specific, although there were some overlaps (see **Appendix** 6.3: Figure 6.1). Three orthologous sets of genes were significantly differentially expressed in the four separate analyses: 1) a group of modular serine proteases (MSPs), 2) a group of uncharacterised serine-rich proteins referred to as probable serine/threonine-protein kinases (S/TKs) and 3) a number of lethal(2)essential for life protein coding genes (L(2)EFL)s.

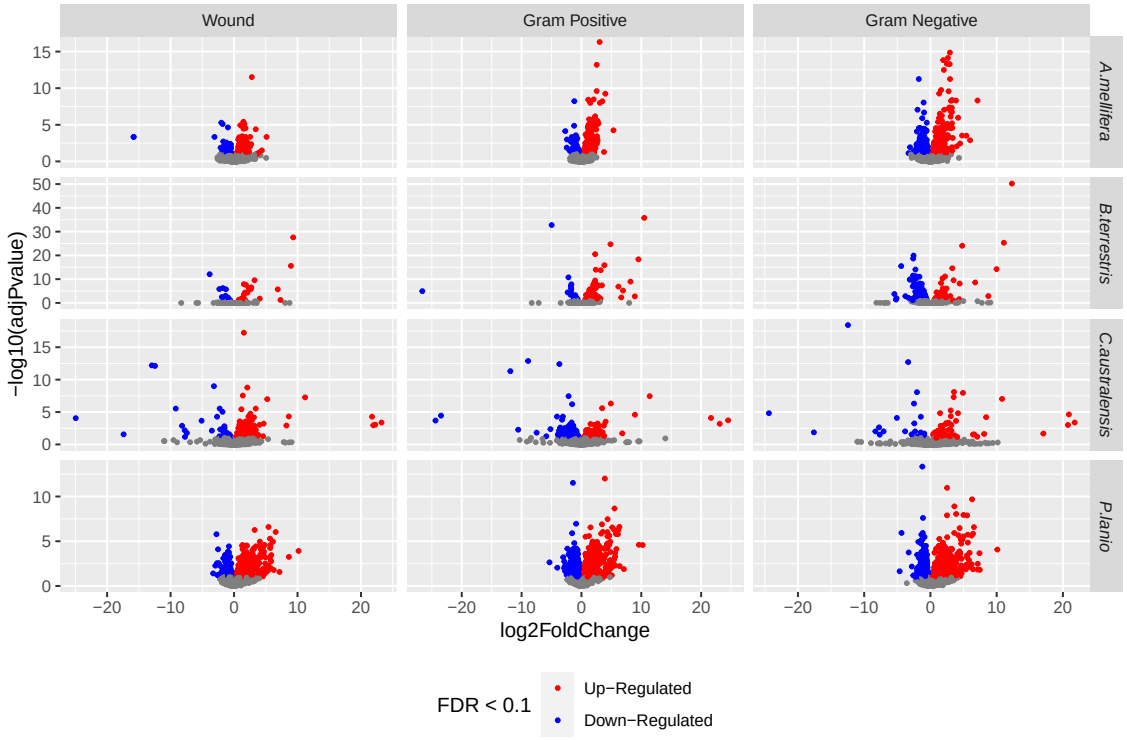

Figure 3.2: Volcano plots of gene expression when comparing the naive condition to sterile wounding, Gram positive and Gram negative exposure across the four species. Coloured points represent significantly up (red) or down (blue) regulated genes. Grey points represent genes that did not significantly vary with treatment.

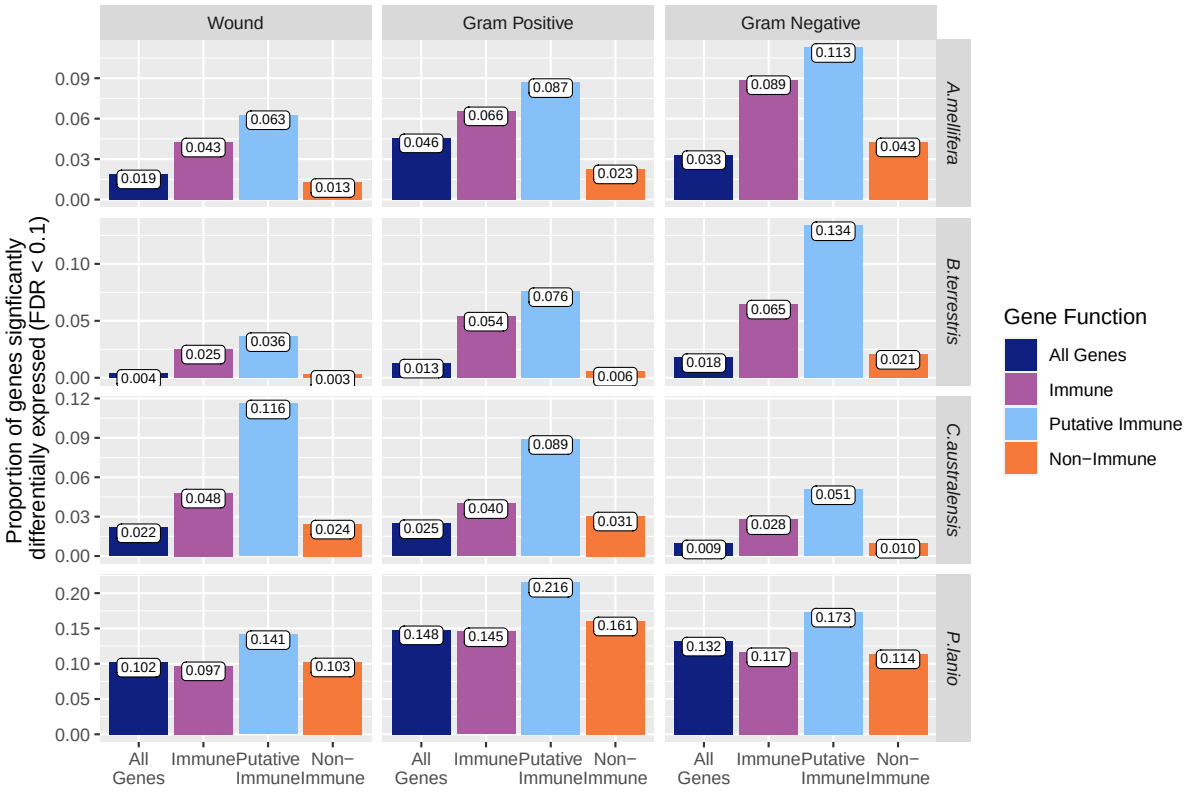

Figure 3.3: Proportion of significantly differentially expressed genes of each function (immune in purple, putative immune in pale blue, non-immune in orange and all genes combined in navy) in each contrast (naive vs sterile wound, naive vs Gram positive and naive vs Gram negative).

However, these shared genes were differentially expressed under different challenges or were expressed in different directions. There were three differentially expressed MSP orthologs in *P.lanio*: one was down-regulated in the sterile wound treatment, one was up-regulated in the Gram positive condition only and the other was up-regulated in both bacterial challenges. There was only one MSP in each of the bees that was up-regulated in all Gram negative treatments. This MSP was also up-regulated on sterile wounding in *A. mellifera* and in *B.terrestris* when exposed to Gram positive bacterial challenge. In *B. terrestris* sterile wounding condition, the S/TKs were up-regulated, while in sterile wounded *P. lanio*, Gram-positive exposed *C. australensis*, and Gram-negative challenged *A. mellifera*, they were down-regulated. In *A. mellifera* and *B. terrestris*, L(2)EFL was significantly down-regulated in Gram negative treatments, and also in Gram positive exposed *B. terrestris*. The L(2)EFL gene in *C.australensis* was up-regulated in the Gram positive exposed bees. In *P.lanio*, there were four differentially expressed L(2)EFL genes. Using *Polistes canadensis* gene RefSeq IDs: *LOC106785761* was down-regulated in both bacterial treatments; *LOC106785771* was up-regulated in sterile wound and Gram positive conditions; *LOC106790160* was up-regulated in the Gram negative condition and *LOC106785762* was up-regulated across all conditions. Of the three genes, there was only one instance of universal expression: the MSP significant up-regulation in the Gram negative condition.

There were 15 orthologous genes differentially expressed across the bee species including classic immune genes such as NF-κ-B inhibitor *cactus* and β-1,3-glucan-binding protein 1 (β*GBP-1*,also known as *b-gluc2* or *gnbp-1*). However, they vary somewhat in patterns of expression (Figure 3.4), and were not always uniformly significant. There were four strikingly consistently expressed genes: serpin-5/88Ea, fibroin heavy chain (FHC), *malvolio*, and an uncharacterised orthogroup of genes referred to here as OG0002961 that were significantly up-regulated in all conditions across all bee species (Figure 3.4).

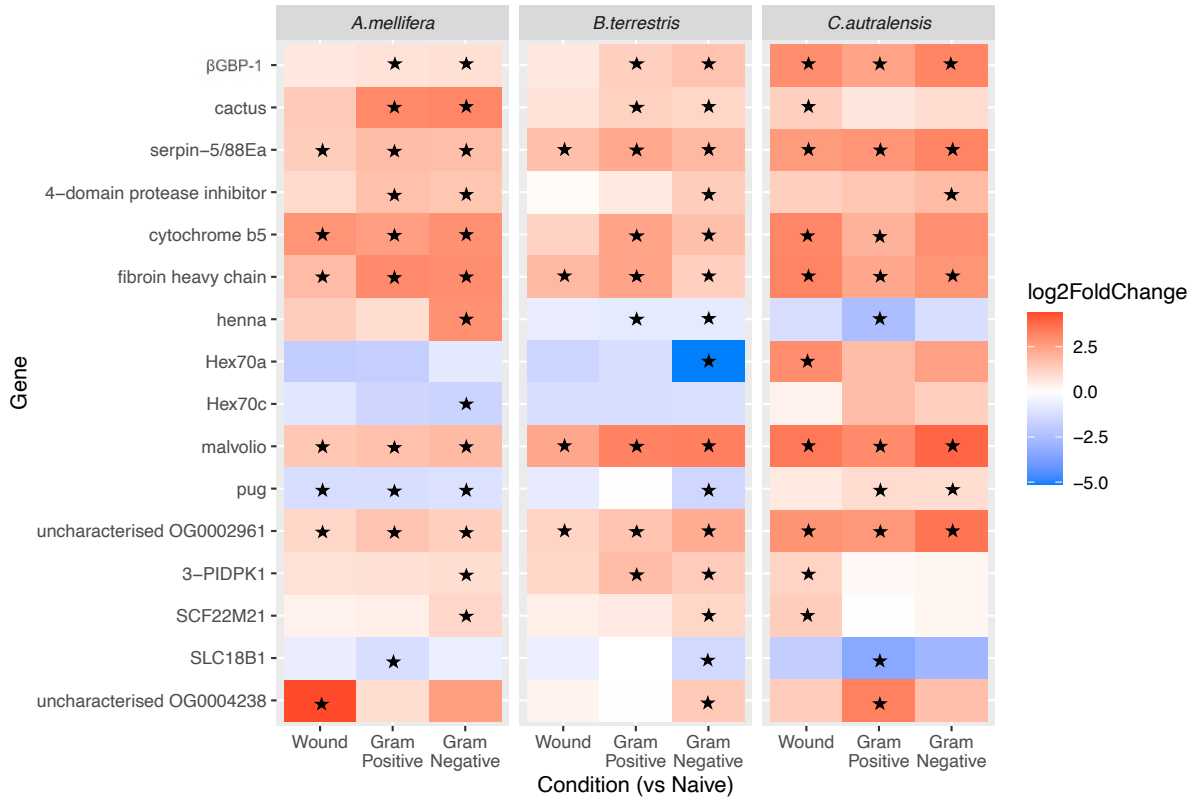

Figure 3.4: Heatmap of the log-2 fold change in gene expression of 15 genes found to be significantly differentially expressed in at least one condition across all three bee species relative to naive samples. *A. mellifera* gene names are used when available. Occurrences where genes are significantly differentially expressed are indicated with black stars ($\star$). 3-PIDK1: 3-phosphoinositide-dependent protein kinase 1; SCF22M21: solute carrier family 22 member 21; SLC18B1: MFS-type transporter SLC18B1.

To identify orthogroups with conserved expression across experimental conditions, I also examined orthogroup-level expression using four combinations of the species: Aculeata, Anthophila, Apinae and Age-Controlled (see **S3.7-3.10**; for results without removing outliers see **S3.11**). There were considerably fewer significant differentially expressed orthogroups, except in the Apinae analysis, with few orthogroups found universally across combinations and treatments (Figure 3.5).

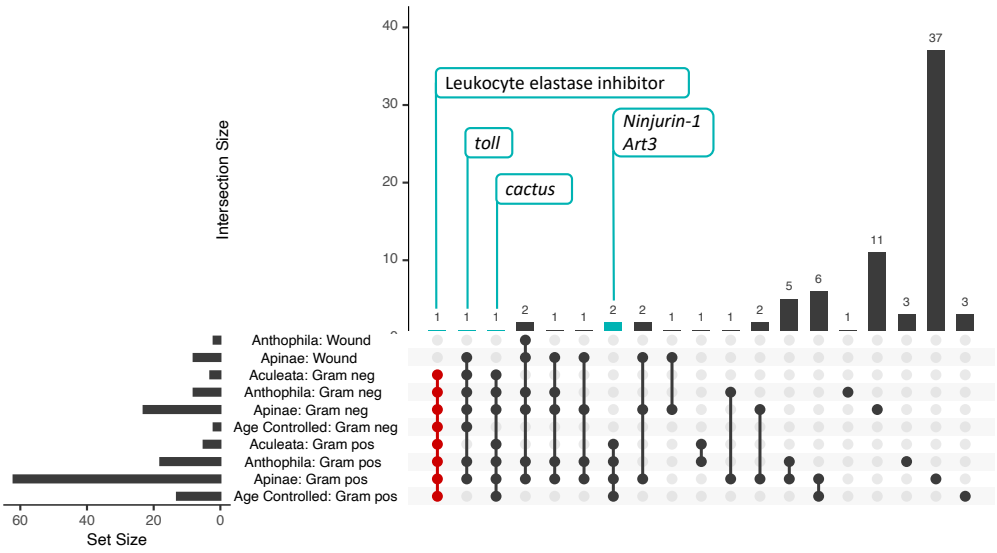

Figure 3.5: Differentially expressed orthogroups and where they intersect across the different species groupings and treatments. Orthogroups that were found differentially expressed in at least one treatment in every combination are highlighted and labelled in blue (using *A. mellifera* orthogroup members). One intersection - a leukocyte elastase inhibitor orthogroup - was only but consistently differentially expressed in bacterial treatments (highlighted in red)

There were five orthogroups that were significantly expressed in all orthogroup level analysis sets (Figure 3.6). A leukocyte elastase inhibitor (LEI) - a serpin - was up-regulated in all bacterial challenges; *ninjurin-1* was up-regulated in the Gram positive challenge in all species' combinations, and a group containing NF-κ-B inhibitors *cactus*, *cactus1* and *cactus2* was consistently significantly up-regulated in all Gram positive challenges, and, with the exception of in the Age-Controlled analysis, all Gram negative challenges. A group of *toll* genes was significantly up-regulated in all Gram negative challenges, as well as Gram positive in Anthophila and Apinae, and in the sterile wound challenge in Apinae. Finally, an arginine methyltransferase

3 (*Art3*) was significantly up-regulated in all Gram positive challenges.

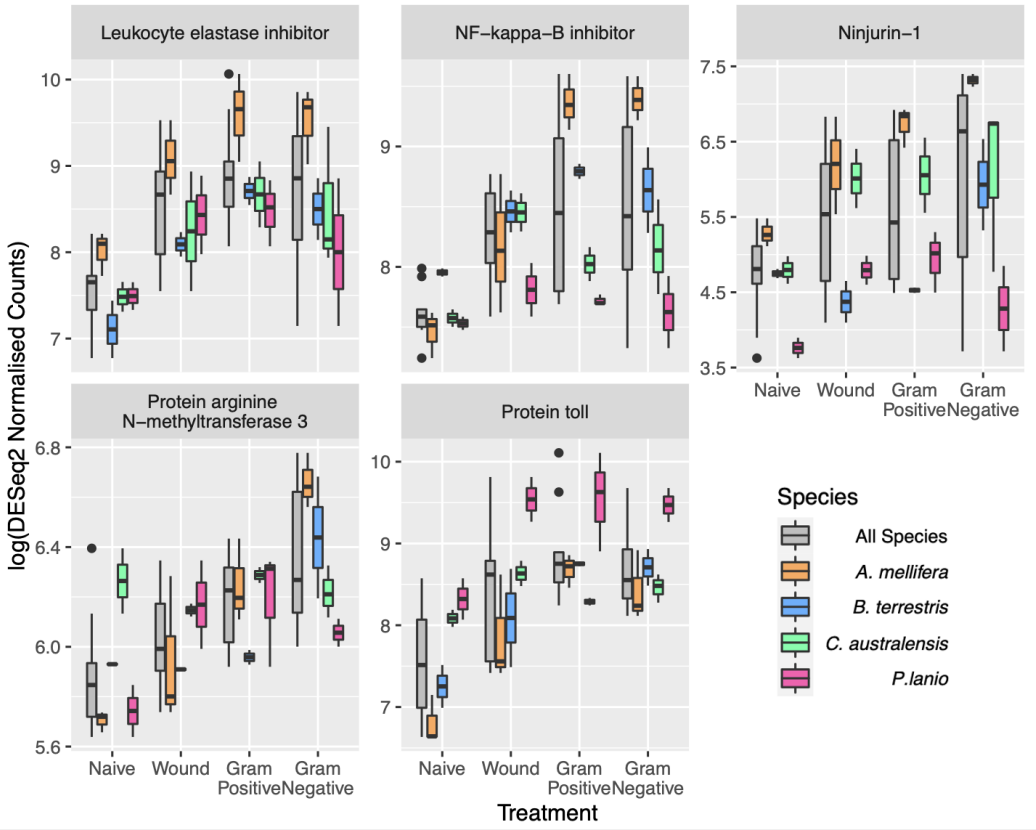

Figure 3.6: Expression boxplots of the five differentially expressed orthogroups. Grey boxplots represent all species, with individual species' expression in coloured boxes (orange: *A. mellifera*; blue: *B. terrestris*; green: *C. australensis*; red: *P. lanio*). Individual outliers are marked with a dark circle.

3.4.3 GO enrichment analyses

In each of the four species I find enriched GO terms for immune function (see **Appendix** 6.3: Table 6.9; also **S3.12-3.15**). *A. mellifera* and *B. terrestris* had the most clearly immune-associated responses, and *C. australensis* had a number of terms associated with pathogen pattern recognition up-regulated throughout all three conditions. *P.lanio*, on the other hand, had few GO terms that appeared to be clearly associated with immune responses.

The most frequently enriched immune GO terms included terms associated with detecting fungal pathogens (*response to fungus*, *defense response to oomyctes*, *detection of molecule of fungal origin*) enriched in all four species, all up-regulated

except in the Gram negative and wound conditions of *B. terrestris*; *Toll-Interleukin receptor (TIR) domain binding* was found to be frequently up-regulated, including an instance in *P.lanio*; *(1− >3)-β-D-glucan binding* and *serine-type endopeptidase inhibitor (serpin) activity* were also frequently up-regulated. Overall, however, the most frequently identified enriched GO terms were associated with ion homeostasis (*multicellular organismal iron ion homeostasis*, *manganese ion transmembrane transport*, *manganese ion transmembrane transporter activity*, *iron import into cell* and *iron ion transmembrane transporter activity*) which were associated with up-regulated genes in multiple conditions in the three bee species.

3.5 Discussion

3.5.1 All species show differential expression in three gene families

Despite extensive taxon specific gene expression in response to immune activation, I find three gene families that show differential expression across taxa. The L(2)EFL genes are heat-shock proteins that are cytoprotective and involved in stress-resistance and longevity in *D. melanogaster*, likely via the JNK signalling pathway (Jo and Imm 2017; Gan et al. 2021; M. C. Wang et al. 2005). In honeybees, L(2)EFL has been identified as a putative immune candidate as it is often up-regulated in response to pathogens (Brutscher et al. 2017; Doublet et al. 2017; Grozinger and Flenniken 2019). There are numerous L(2)EFL gene family members across the taxa included in my orthology sorting (Figure 3.7), several of which were significantly differentially expressed in each of the experimental species. However, there was not consensus in the direction of expression or the experimental condition that induced differential expression except in one instance (Figure 3.7:4) where orthologs in *P. lanio* and *C. australensis* were both up-regulated in Gram positive treatments. Interestingly, the one *A. mellifera* L(2)EFL gene (*LOC724367*) was significantly down-regulated in the Gram negative response and is a highly connected hub gene in immune co-expression pathways (Doublet et al. 2017).

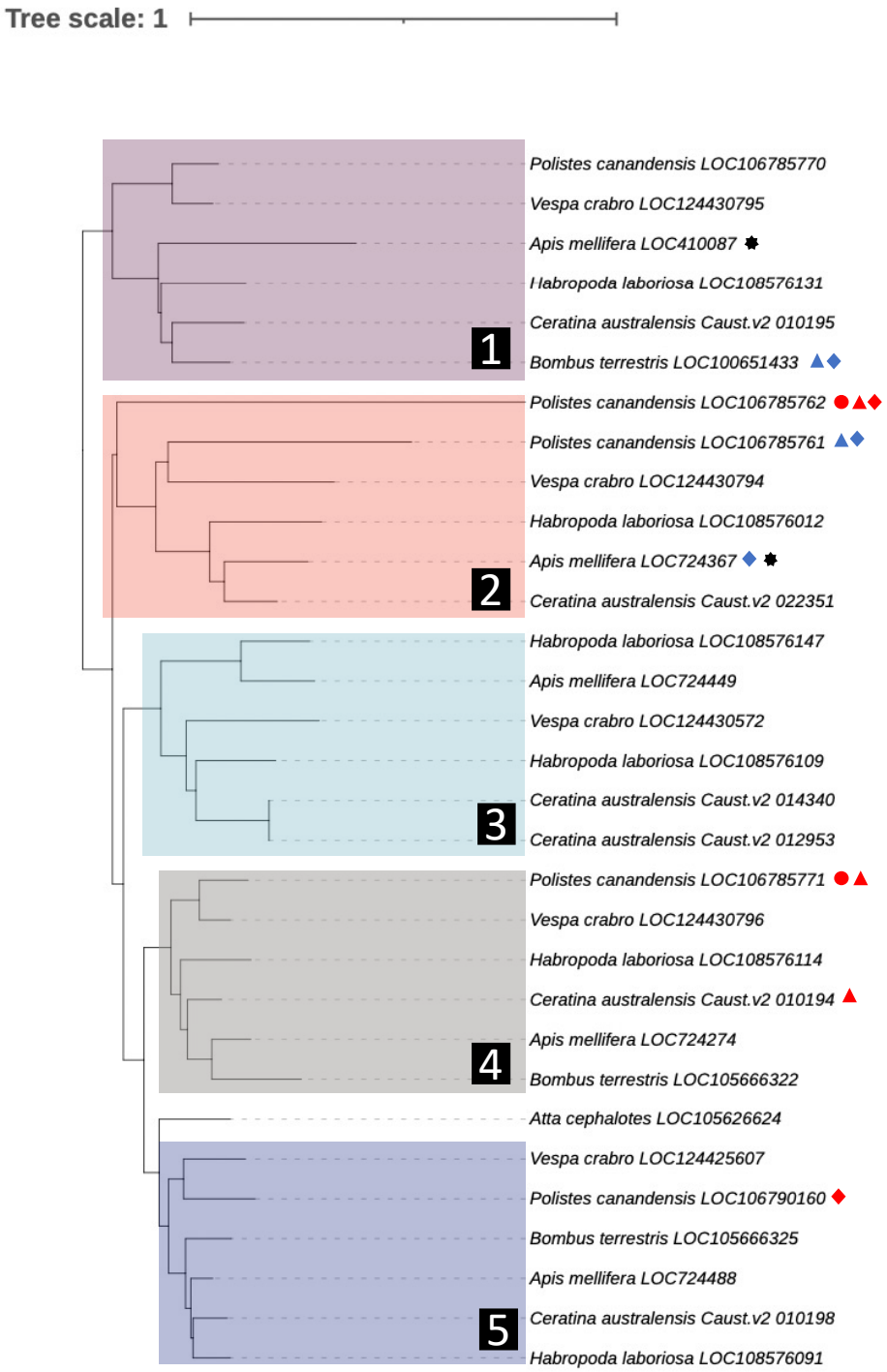

Figure 3.7: Gene tree of protein lethal(2)essential for life (L(2)EFL) orthogroup as produced by `Orthofinder` (Emms and Kelly 2019). Highlighted are potential different homologous gene groupings. I mark differentially expressed genes according to condition (wound: circle; Gram positive: triangle; Gram negative: diamond) and by direction of gene expression (red: up; blue: down). Genes found as differentially expressed in Doublet et al. 2017 or Brutscher et al. 2017 are marked with a black star ($\star$).

The gene family designated S/TK consist of three single copy one-to-one orthologs with no convincing homology to sequences outside Hymenoptera (Altschul et al. 1990). InterPro annotation suggest they each contain a basic helix-loop-helix (bHLH) domain and are part of the transcription factor ATOH8 superfamily (IPR011598, IPR032660, Jones et al. 2014) suggesting that these genes are putative transcription factors.

The modular serine protease (MSP) group had multiple differentially expressed members, but this was only consistently differentially expressed across species experiencing Gram negative bacterial exposure. The nearest *D. melanogaster* ortholog to these MSPs is a signalling molecule in the Toll immune pathway (Ali Mohammadie Kojour et al. 2020; Valanne et al. 2011). However, this is generally a pathway activated by fungi and Gram positive bacteria, and yet I see MSPs up-regulated most frequently in the Gram negative condition. While the Toll pathway is commonly thought to be limited to Gram positive bacteria and fungi there is evidence of both cross-talk across insect immune pathways and Toll pathway activation by Gram negative bacteria in other insect species (Edosa et al. 2020; Park et al. 2019; Yu et al. 2010; Yu et al. 2020), including to the specific bacterial species I used here (Z. Wang et al. 2021).

3.5.2 Toll pathway genes consistently differentially expressed in the bees

I find that a higher proportion of immune-associated genes are significantly differentially expressed than genes without known or putative immune function in the bees, although this was not the case in the wasp, *P. lanio* (Figure 3.3). There was also greater overlap in the shared differentially expressed genes among the bees than when including *P. lanio*. Of the 15 overlapping gene families among the bees (Figure 3.4), three have known canonical immune functions and eight have putative immune roles (see **S3.1**). These results lend further weight to their likely role in immune responses not just in the honeybee (where this possible immune function was inferred, Doublet et al. 2017), but across the surveyed bees which span approximately 100

million years of divergence (Kumar et al. 2022).

Among these canonical immune genes, I find the pathogen recognition receptor β-1,3-glucan-binding protein 1 (*βGBP-1*, also known as *gnbp-1*). βGBPs are humoral pattern recognition receptors (PRR) that can interact with other receptors, commonly activating the prophenoloxidase (PPO) cascade. This cascade can trigger a number of defense mechanisms not limited to melanisation and wound healing (A. Lu et al. 2014; Vetvicka and Sima 2017; W. Zhang et al. 2021). As my treatments were administered via injection, wound healing through PPO activation is intuitive. Interestingly, in the mealworm *Tenebrio molitor* this pathway has been shown to more specifically be initiated when a Gram negative binding protein (gbnp) binds to microbial peptidoglycan and, as a complex, interacts with a MSP (Cerenius et al. 2010; Park et al. 2007). This interaction leads to a signalling cascade that can induce both Toll and PPO pathways. This, taken in combination with my finding of a MSP up-regulated across all species in the Gram negative conditions, suggests I am detecting the initialisation of one or both of these pathways in response to my administered treatments.

More generally, *βGBP-1* is well known to initiate the *toll* signalling cascade, with some cross-talk with the IMD pathway, in the model fruitfly *Drosophila melanogaster* (P. Qiu et al. 1998; Towb et al. 2001; Valanne et al. 2011). Expression of *βGBP-1* was only differentially expressed in the bacterial treatments in the lab-reared species. In *C. australensis* there was also *βGBP-1* differential expression in the sterile wounding response, but, as these animals were field caught and not age-controlled, their immune history can not be known and will likely affect immune expression. Two other members of the Toll pathway were also consistently differentially expressed in the bees. These were *cactus*, a NF-κ-B inhibitor that suppresses downstream *toll* signalling and also plays a role in haematopoiesis and melanisation, and serpin-5/88Ea, a signalling molecule that regulates the Toll pathway and functions as a serine protease inhibitor (Ahmad et al. 2009).

3.5.3 Three candidate immune genes exhibit universal up-regulated response to all challenges in the bees

Outside of canonical immune pathways, there were three differentially expressed genes in the bees that have previously been highlighted as putative immune genes (Doublet et al. 2017). These three were consistently up-regulated in all experimental conditions in each of the bee taxa. The first, OG0002961 (based on the identifier allocated by `OrthoFinder`, Emms and Kelly 2019, see **S3.1**) consists of single-copy one-to-one orthologs across the Hymenopteran species I analysed, and are largely uncharacterised. Annotations from the hornet *V. crabro* and two ant species assign possibilities: circumsporozoite-like protein (*LOC105617077*), collagen alpha-1(IV) chain (*LOC105204660*), and S/TK (*LOC124432034*). The bee members of this group are mysterious and there is little currently available to elucidate function, with no annotated domains detected using Interpro (Jones et al. 2014). Most of the potential orthologs outside of Hymenoptera, found using `BLAST` (Altschul et al. 1990), are also uncharacterised.

The second of these genes are annotated as fibroin heavy chain (FHC) proteins or spindroins. I find no clear orthologs to this gene outside of Hymenoptera based on `BLAST` searches (Altschul et al. 1990). FHCs and spidroins are both constituents of silk (Andersson et al. 2016; Herold and Scheibel 2017; Zhou et al. 2000). Honeybees, bumblebees, ants and hornets are known to spin silk as larvae (Sutherland et al. 2012), consisting of proteins from four paralogous genes, *fibroin*1-4 (Sutherland et al. 2006). The FHC protein here, however, does not cluster with these well-understood silk fibroins, and as such I cannot conclude that this is a protein used in silk production, especially since I see it up-regulated in adult bees. Interestingly, the *Bombyx mori* FHC - which is a well-characterised silk fibre constituent - mediates cross-talk between silk glands and the fat body (Q. Chen et al. 2015). The insect fat body is an important player in orchestrating innate immune responses (Hillyer 2016) and perhaps, in Hymenoptera, this is a protein structurally similar to fibroin that works as a communication molecule for the fat body. Of course, this is speculative and

would require further research to investigate any associations with the Hymenopteran fat body.

While not normally viewed as a canonical immune gene, *malvolio* has long been linked to immune responses in the honeybee, where an increase in gene expression results in infected bees switching to foraging behaviour, thus moving infection away from the heart of colony (Alaux et al. 2011; Antonio et al. 2008; Ben-Shahar et al. 2004). As such, *malvolio* may be an example of a gene involved in social immunity - a suite of behaviours in eusocial insects that limit or eliminate infection (Cremer et al. 2007; Cremer et al. 2018; Dolezal and Toth 2014; Oxley et al. 2010; Wilson-Rich et al. 2009).

This may be the case for the eusocial *A. mellifera* and *B. terrestris* - though it has not, to my knowledge, been demonstrated in the latter - but *C. australensis* is only socially polymorphic, living both solitarily and as part of small groups (Harpur and Rehan 2021). The up-regulation of *malvolio* in *C. australensis* makes it more likely that this gene plays a part in direct individual immune mechanisms and that perhaps its role in moving infected social bees away from vulnerable colony components evolved secondarily.

Further support for this gene putatively being an ancient and fundamental component of the insect immune response comes from an interesting possible connection to ion sequestration. In the bees, I find enriched GO terms associated with metal ion homeostasis were the most frequently shared (see **Appendix** 6.3: Table 6.9), specifically concerning three ions: iron, manganese and copper (see **S3.12-3.15**). It is likely that these terms reflect the utilisation of nutritional immunity by these samples.

Nutritional immunity includes the phenomena wherein hosts respond to pathogens by limiting access to transition metal ions essential for the survival and propagation of the invaders (Hood and Skaar 2012), or else boost ion levels to increase cytotoxicity or improve functionality of other immune processes (Hrdina and Iatsenko 2021). Iron sequestration is perhaps the most understood process of nutritional immunity,

wherein the host restricts free iron from being accessible by pathogens (Ong et al. 2006). In Hymenoptera, there is evidence of intracellular parasites taking advantage of this mechanism (Rodríguez-García et al. 2021). In Diptera, this has been suggested to occur when the host transports iron in the haemolypmh to the fat body (Iatsenko et al. 2020), and in *D. melanogaster*, there is evidence that this process is mediated by transferrin proteins, specifically transferrin-1 (Weber et al. 2020).

Though I found no significantly increased expression of bee transferrins, the Drosophila *malvolio* ortholog is also implicated in iron and manganese transportation (Bettedi et al. 2011; Orgad et al. 1998; S. Wu et al. 2022). In the fruitfly, *malvolio* is expressed in both haemocytes and in Malpighian tubules and has been implicated in susceptiblity to Sindbis virus (Rose et al. 2011), suggesting that this gene may have a long history of playing a part in innate insect immune response that predates the evolutionary split between Diptera and Hymenoptera.

After iron homeostasis, the most recurrently up-regulated GO term associated with sterile wound, Gram positive (all bees), and Gram negative (*B. terrestris* and *C. australensis*) conditions was manganese ion transport. Manganese ions are found to be sequestered in mammalian innate immune responses (Corbin et al. 2008; Kehl-Fie et al. 2011) but the role they play in insect-pathogen interactions is less characterised. As many microbial species require manganese to be viable (Colomer-Winter et al. 2018; Porcheron et al. 2013) it is probable that manganese sequestration is a generic host response intended to starve microbial invaders much in the manner of iron sequestration. Again, *malvolio*'s up-regulation is likely a driver of these GO terms being over-expressed.

Copper ions have been demonstrated to play a role in melanisation, wherein they support the action of copper-dependent tyrosinases (Dudzic et al. 2019; Nappi and Christensen 2005). Copper levels have also been found to correlate with pigment intensity in *D. melanogaster* (Vásquez-Procopio et al. 2020). The up-regulation of genes annotated with copper transport in experimental conditions in the bees may reflect a melanisation immune response. The fact that the increase in ion movement

encompasses all three experimental conditions suggests that nutritional immunity may be a generic response to the cuticle being breached and does not require specific pathogen recognition. However, as with all metal ions, there are many other host functions that these metals may be being gathered to support, and the lack of evidence of this occurring in *P. lanio* undermines any conclusions over the uniformity of the response in Hymenopteran species.

3.5.4 Five orthogroups exhibit conserved expression across treatments

My aim was to identify shared immunologically reactive genes across Hymenopterans. I find five orthologous groups that have conserved responsiveness among my taxa (Figure 3.6). Not all of these groups of genes are considered classically immune. Arginine N-methyltransferase 3 (*Art3*) is part of an ancient family of protein arginine methytransferases (PRMT) that are present in life from yeast to humans (Krause et al. 2007). *Art3*, as its name suggests, confers epigenetic modifications by transfering methyl groups to arginine residues and are involved in diverse RNA processing and transcriptional regulation in mammals (Bedford and Clarke 2009). Its possible role in an immune response is less clear: in *Aedes aegypti*, an *Art3* ortholog was up-regulated by *Wolbachia*, facilitating bacterial colonisation, but *Art3* expression does not have any effect on Dengue virus titre (G. Zhang et al. 2014). Given the epigenetic role of PRMTs, one can imagine differential expression of this modulating protein could have diverse targets, including regulating immune-associated genes to orchestrate immune responses.

Ninjurins are another family of conserved proteins found in both vertebrates and invertebrates (Shuning Zhang et al. 2006). These cell adhesion proteins were first discovered whilst studying nerve injury (Araki and Milbrandt 1996). *Ninjurin-A* is activated by the Toll pathway and is involved in non-apoptic cell death in *D. melanogaster* (Broderick et al. 2012). *Ninjurin-2* is inconsistently differentially expressed in infected honeybees (Doublet et al. 2017). As injury would have been present in all of my administered treatments as the cuticle was pierced, it can only

be speculated that this gene is up-regulated in response to injury, not unlike how it responds to neuronal injury in vertebrates. Although both *Art3* and *ninjurin-1* were mostly up-regulated relative to naive conditions, it was only in the Gram positive treatment that this up-regulation was significant, suggesting that, whatever roles they play, they are potentially more pronounced in response to Gram positive bacteria.

The *toll* orthogroup consists of the classic PRR that activates the Toll signalling immune pathway (Hillyer 2016; W. Zhang et al. 2021). While *toll* signalling is commonly thought to be primarily a response to Gram positive bacterial or fungal recognition, I find up-regulation across the bacterial treatments. This, however, is largely consistent with the literature which describes extensive cross-talk between the Toll and IMD pathways and the activation of *toll* by both β-glucan and peptidoglycan receptor proteins (Edosa et al. 2020; Park et al. 2019; Yu et al. 2010; Yu et al. 2020).

There is also the possibility that the wounding stimulus, consistent across all challenges, introduced additional cues of infection leading to *toll* signalling. This is supported by the difference in expression patterns between the lab-reared and field-caught animals, the former of which have clear increased *toll* expression in the bacterial conditions relative to sterile wounding, while wild collected animals had more uniform response to all treatments. The NF-κ-B inhibitor *cactus* was also significantly differentially expressed in bacterial exposures. The *cactus* proteins play a negative regulatory role in the Toll pathway, keeping the response in check and reducing risk of harm to the host (W. Zhang et al. 2021). Both of these groups of genes being up-regulated in the same conditions suggest that the *toll* signalling is being both activated and regulated in response to my challenges.

Leukocyte elastase inhibitors (LEI) are serpins, but their exact role in immunity is not clear. In the ant *Camponotus floridanus*, LEI was up-regulated in response to pathogen challenge (Ratzka et al. 2011). The closest homolog hit to *A. mellifera* LEI (*LOC100577408*) in *D. melanogaster* was Serpin 28Dc (*SPN28Dc*). *SPN28Dc* is an immune response-associated serpin involved in triggering PPO cascade in response to wounding in *Drosophila* (Reichhart et al. 2011; Scherfer et al. 2008). Considering

that all my treatments include a wound, it is plausible that the Hymenopteran LEI plays similar roles.

3.5.5 Conclusion

I find overwhelmingly species-specific differential expression consistent with previous work demonstrating taxon- and even challenge-specific immune responses (Doublet et al. 2017; Sackton 2019; Troha et al. 2018). To improve the likelihood of detecting shared immune responses, I attempted to standardise as much as possible the protocols of animal care, exposure, and sampling. However, some species were not able to be lab-reared or age controlled. Despite these limitations, I do still detect signals of common innate immune responses across the four Hymenopteran species examined here. I show that these shared immune genes are part of both known and novel immune pathways including the Toll pathway (*toll*, *cactus*, *serpin5/88Ea*, *βGBP-1*, MSP), the prophenoloxidase (PPO) and melanisation responses (MSP, *βGBP-1* and putatively LEI), stress resistance (L(2)EFL) and possibly nutritional and social immunity (*malvolio*). I also present a number of candidates that warrant further investigation, from the uncharacterised OG0002961 group of genes, to FHC, *Art3* and *ninjurin-1*. I propose that future work expand these assays by recruiting more diverse species to further elucidate our understanding of Hymenopteran immunity.

Contributions

I would like to highlight and thank the following for their contributions to the work included in this chapter: Tim Sackton and Rob Brucker, whose input helped my supervisor Seth Barribeau design this experimental protocal; Sandra Rehan and Katherina Odanaka who collected and challenged the *Ceratina australensis* samples; and Sierian Sumner and her postdoc Robin Southon who collected and challenged the *Polistes lanio* samples.

Chapter 4

The influence of social lifestyles on host-microbe symbioses in the bees

---

4.1 Abstract

Microbiomes are increasingly recognised as critical for the health of an organism. In eusocial insect societies, frequent social interactions allow for high fidelity transmission of microbes across generations, leading to closer host-microbe coevolution. The microbial communities of bees with different social lifestyles are less well studied, and few comparisons have been made between taxa that vary in social structure. To address this gap, I leveraged a cloud-computing resource and publicly available transcriptomic data to conduct a survey of microbial diversity in bee samples from a variety of social lifestyles and taxa. This method was able to consistently recover the core microbes of well studied corbiculates, supporting its ability to accurately characterise microbial communities. I find that the bacterial communities of bees are influenced by host location, phylogeny, and social lifestyle, though no effect was found for fungal or viral microbial communities. Bee genera with more complex societies tend to harbour more diverse microbes, with *Wolbachia* detected more commonly in solitary tribes. I present the first description of the microbiota of Euglossine bees and find that they do not share the "corbiculate" core microbiome. Notably, I find that

bacteria with known anti-pathogenic properties are present across social bee genera, suggesting that symbioses that enhance host immunity are important with higher sociality. My approach provides an inexpensive means of exploring microbiomes of a given taxa and identifying avenues for further research. These findings contribute to our understanding of the relationships between bees and their associated microbial communities, highlighting the importance of considering microbiome dynamics in investigations of bee health.

4.2 Introduction

In the insect world, microbial symbionts can play a major role in many biological processes (Munoz-Benavent et al. 2021), including reproduction (Bourtzis et al. 1996; Singh and Linksvayer 2020; Werren et al. 2008), nutrition (Andersen et al. 2012; Cheng et al. 2019) and pathogen defense (Bian et al. 2010; Benoit et al. 2017; Duplouy et al. 2015). For social insects, where consistent social contact between conspecifics allows for high-fidelity vertical transmission of microbial communities, these symbionts can be passed on for generations, allowing for coevolution of microbiome and host (Dietrich et al. 2014; Lombardo 2008; Kwong et al. 2017a; Sanders et al. 2014; Zhang and Zheng 2022). This has been demonstrated in the obligately eusocial corbiculate bees, which all share a core set of bacterial microbes (Koch and Schmid-Hempel 2011a; Koch et al. 2013; Kwong and Moran 2016b; Kwong et al. 2017a; Lim et al. 2015; Moran et al. 2012). The members of this conserved bacterial complement are important for the health of their hosts, particularly in terms of disease resistance (Anderson et al. 2014; Bonilla-Rosso and Engel 2018; Koch and Schmid-Hempel 2012; Miller et al. 2021; Vásquez et al. 2012).

However, there are very few bee microbial studies outside of these eusocial corbiculates (Handy et al. 2022; Kapheim et al. 2021; McFrederick et al. 2012; McFrederick et al. 2014; McFrederick et al. 2017), meaning the microbiomes of most less popular bee species remain a mystery. One of the current approaches of characterising the microbiome of a host is to use metagenomic Next Generation

Sequencing (mNGS), where all DNA (or RNA) from a given environment - i.e. an insect gut - is sequenced and the microbial community characterised. While the cost of producing NGS data has dramatically reduced over recent years, it remains reasonably expensive, taking into account sample extraction, library production, sequencing costs, and having the appropriate informatics infrastructure in order to store, process and analyse data (Krampis and Wultsch 2015).

One attractive solution for some analyses is to use cloud-computing resources (Krampis and Wultsch 2015). CZID.org, for example, is an approachable, open source cloud-based service which can provide microbial identification for many different sample types and host species (Kalantar et al. 2020). Here I use this approach to examine NGS datasets from 18 bee genera spanning 100 million years of divergence (Figure 4.1, Peters et al. 2017) that vary in their social structure, ranging from solitary to obligately eusocial. As one of the purposes of this analysis was to assess any differences in microbial composition potentially caused by consistent social interaction, I simplified the many different distinctions in social structure found in the literature to: 1) solitary, where species do not provide any brood care and associate with conspecifics only for mating; 2) facultatively eusocial, which included any species that had considerable contact with conspecifics (i.e. communal nesting) and some brood care (primitively eusocial) but where individuals can and do live solitarily and 3) obligately eusocial species that only ever exhibit eusocial behaviours and solitary living is impossible (Figure 4.1). I used this framework to systematically test whether social structure, location or bee taxa affects microbial composition across the bees.

4.3 Materials and methods

4.3.1 Sample selection

I analysed sequence data sourced from NCBI's Sequence Reads Archive (SRA, Katz et al. 2022; Kodama et al. 2012; Leinonen et al. 2010), accessed September 2022. I

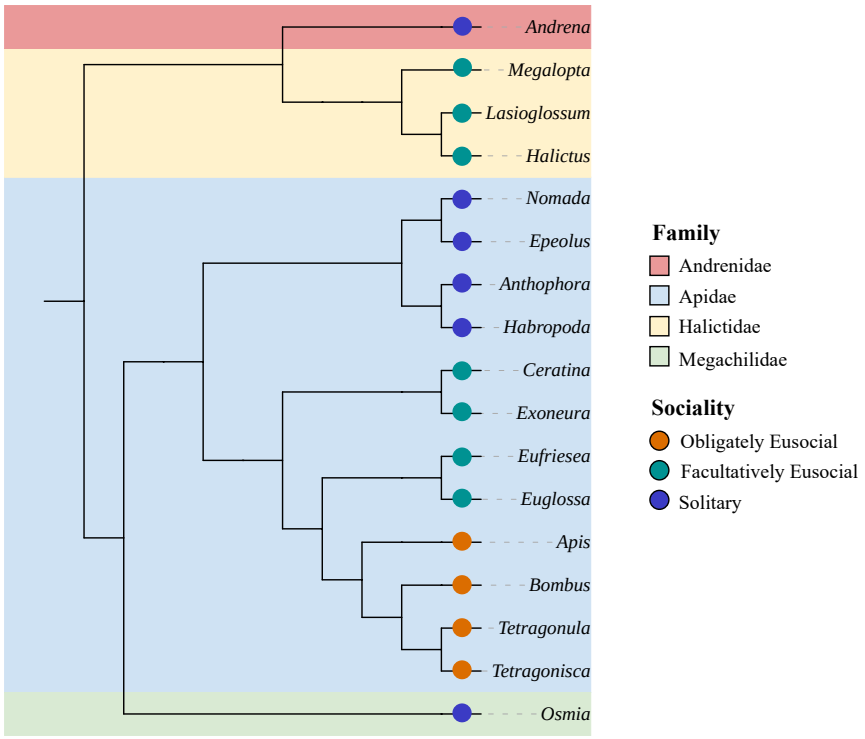

Figure 4.1: Phylogeny of the genera included in these analyses coloured by family, with the sociality of each genus specified by a coloured circle. The corbiculate bees are marked within a black lined box. This tree is based on accepted topology in the literature (Bossert et al. 2019; da Silva, Jack 2021; Gibbs et al. 2012; Husemann et al. 2021; Kapheim et al. 2019; H. Lu et al. 2021). Branch lengths are not indicative of evolutionary time. Figure produced using iTol's web tool (Letunic and Bork 2021).

included all available RNA-Seq adult bee samples that included the animal's abdomen (including pooled individuals) and I excluded projects that exclusively sequenced any other part (e.g. antennae, brain, ovaries). I only included unaltered control specimens (i.e. no treatment or stressor introduced/administered) to ensure that the microbial composition was as natural as possible.

4.3.2 Processing, mapping and uploading reads

All sequence data (fastq format, Table 4.1, see **S4.1**) were downloaded and unpacked from the SRA using `prefetch` and `fasterq-dump` from the SRA-toolkit (version 3.0.0). From here I split the pipeline: files sequencing the European honeybee *Apis mellifera* were uploaded directly to CZID.org using the command-line interface (version 4.1.2), and non-*A. mellifera* sequences were retained for further processing.

CZID (Chan Zuckerberg ID, previously known as IDSeq, Kalantar et al. 2020) is a cloud-based, open-source platform that maps input sequence files against a chosen species genome and then aligns any unmapped reads to NCBI databases in order to detect non-host sequences.

Briefly, the CZID pipeline (Kalantar et al. 2020) used in this analysis can be summarised in the following steps. Firstly, a genome and blank sample is chosen. The former is what input reads are mapped against, the latter is used to calculate the likelihood of alignment hits occurring due to contamination. The input sequences are validated before the first round of mapping reads against the chosen genome. The resultant unmapped reads are then processed as to remove adaptor sequences, duplicated or low quality reads. These reads are then mapped again using a different genome mapper and, finally, unmapped reads are sub-sampled and remaining reads are aligned against the NCBI nucleotide (NT) and non-redundant protein (NR) sequence databases. In each of these non-host taxa alignment "hits" the number of reads are recorded and these counts can be considered as representative of microbial taxa presence and abundance. The pipeline output is a CZID taxon report with all taxa hits and accompanying measurements, such as number of aligned reads, evalues and Z-scores (used to determine likelihood of a read being contamination).

The genome that original input sequences will be mapped against is selected from a pre-determined list and at the time of the analysis (October 2022) the host genome option "Bee" included only the honeybee, *A. mellifera*, genome. Therefore, non-*A. mellifera* runs required a number of pre-processing steps. First, each sample was assigned the phylogenetically closest reference genome (see **Appendix** 6.4: Table 6.10). These sequence files were then mapped against each respective genome using `STAR` (version 2.7.10a, Dobin et al. 2013; Dobin and Gingeras 2016). Every sample that achieved $> 50\%$ of reads successfully mapping to the reference genome proceeded to the next step. For the samples that had $\leq 50\%$ reads fail to map because they were 'too short', I repeated the mapping with slightly relaxed parameters (`--outFilterScoreMinOverLread` 0.3 `--outFilterMatchNminOverLread`

0.3). This was needed when the species was comparatively phylogenetically distant from the nearest available genome. Regardless of the success of the second mapping run, all unmapped sequence files were then uploaded to CZID.org for taxonomic assignment using pipeline version 7.1.

Table 4.1: List of host species with associated NCBI projects and references when available. For species that were sourced from more than one project, the number of samples taken from each project is indicated in parentheses next to project accession IDs. Unaccounted *Apis mellifera*, *Bombus terrestris* and *Ceratina australensis* samples were sourced from my own unpublished experiments. See **S4.1** for further details.

Species	**n**	**Project(s)**	**Reference(s)**
Andrena spp.	4	PRJNA687318	Daughenbaugh et al. 2021
Andrena camellia	4	PRJNA510543	
Andrena cineraria	1	PRJNA411946	Schoonvaere et al. 2018
Andrena fulva	1	PRJNA411946	Schoonvaere et al. 2018
Andrena haemorrhoa	2	PRJNA411946	Schoonvaere et al. 2018
Andrena vaga	1	PRJNA318490	Schoonvaere et al. 2016
Anthophora plumipes	1	PRJNA252326	Peters et al. 2017
Apis cerana	5	PRJNA235974(1), PRJNA562784 (4)	Fan et al. 2022; Park et al. 2015
Apis mellifera	87	PRJNA274674 (3), PRJNA357165 (6), PRJNA357523 (7), PRJNA380316 (3), PRJNA495845 (9), PRJNA510543 (3), PRJNA531527 (4), PRJNA681941 (6), PRJNA687066 (6), PRJNA754836 (6), PRJNA793424 (5), PRJNA820512 (23)	Brettell et al. 2019; Brettell et al. 2020; Daughenbaugh et al. 2021; Galbraith et al. 2015; Lester et al. 2022; Melicher et al. 2019; Remnant et al. 2017; Roberts et al. 2017; X. Wang et al. 2021; Y.-Y. Wu et al. 2017
Bombus spp.	1	PRJNA704259	Pascall et al. 2021
Bombus breviceps	1	PRJNA659133	Sun et al. 2021
Bombus confusus	1	PRJNA659133	Sun et al. 2021
Bombus consobrinus	1	PRJNA659133	Sun et al. 2021

Continued on next page

Table 4.1: continued

Species	n	Project(s)	Reference(s)
Bombus difficillimus	1	PRJNA659133	Sun et al. 2021
Bombus haemorrhoidalis	1	PRJNA659133	Sun et al. 2021
Bombus ignitus	1	PRJNA659133	Sun et al. 2021
Bombus lucorum	2	PRJNA704259	Pascall et al. 2021
Bombus opulentus	1	PRJNA659133	Sun et al. 2021
Bombus pascuorum	9	PRJEB43529 (1), PRJNA318490 (1), PRJNA411946 (5), PRJNA704259 (2)	Darwin Tree of Life Project Consortium 2022; Pascall et al. 2021; Schoonvaere et al. 2016; Schoonvaere et al. 2018
Bombus picipes	1	PRJNA659133	Sun et al. 2021
Bombus pyrosoma	7	PRJNA646593 (1), PRJNA646602 (1), PRJNA646687 (1), PRJNA646806 (1), PRJNA646816 (1), PRJNA646831 (1), PRJNA659133 (1)	Y. Liu et al. 2020; Sun et al. 2021
Bombus rupestris	1	PRJNA252285	Peters et al. 2017
Bombus sibiricus	1	PRJNA659133	Peters et al. 2017
Bombus soroeensis	1	PRJNA659133	Sun et al. 2021
Bombus superbus	1	PRJNA659133	Sun et al. 2021
Bombus terrestris	20	PRJNA295976 (5), PRJNA318490 (1), PRJNA411946 (3), PRJNA615177 (6), PRJNA704259 (2)	Amsalem et al. 2015; Araujo and Arias 2021; Pascall et al. 2021; Schoonvaere et al. 2016; Schoonvaere et al. 2018
Bombus terricola	12	PRJNA730495	Tsvetkov et al. 2021
Bombus turneri	1	PRJNA659133	Sun et al. 2021
Bombus waltoni	1	PRJNA659133	Sun et al. 2021
Ceratina australensis	5	PRJNA302035 (2)	Rehan et al. 2018
Dufourea novaeangliae	1	PRJNA279825	Kapheim et al. 2015
Epeolus variegatus	1	PRJNA252262	Peters et al. 2017
Eufriesea mexicana	1	PRJNA279814	Kapheim et al. 2015
Euglossa dilemma	7	PRJNA252310 (1), PRJNA636137 (6)	Peters et al. 2017; Séguret et al. 2021
Euglossa viridissima	20	PRJNA636137	Séguret et al. 2021

Continued on next page

Table 4.1: continued

Species	n	Project(s)	Reference(s)
Euglossa viridissima	20	PRJNA636137	Séguret et al. 2021
Exoneura spp.	1	PRJNA687066	Brettell et al. 2020
Halictus sexcinctus	1	PRJNA374528	Ballenghien et al. 2017
Lasioglossum spp.	2	PRJNA687066	Brettell et al. 2020
Megalopta genalis	22	PRJNA331103	B. M. Jones et al. 2017
Nomada lathburiana	1	PRJNA252330	Peters et al. 2017
Osmia bicornis	8	PRJNA285788 (7), PRJNA411946 (1)	Beadle et al. 2019; Schoonvaere et al. 2018
Osmia cornuta	4	PRJNA318490 (1), PRJNA411946 (3)	Schoonvaere et al. 2016; Schoonvaere et al. 2018
Tetragonisca angustula	6	PRJNA615177	Araujo and Arias 2021
Tetragonula carbonaria	2	PRJNA687066	Brettell et al. 2019

4.3.3 Taxonomy

All taxonomic classifications used in this analysis were sourced from the NCBI Taxonomy (taxonomy dump file from NCBI ftp service Federhen 2012; Schoch et al. 2020, accessed 18th October 2022). A single manual change was made: to distinguish the *Lactobacillus Firm-5* as a separate genus to *Lactobacillus*, as this taxonomic cluster has repeatedly been found to be an important member of the corbiculate bee microbiome (Kwong et al. 2017a; Martinson et al. 2011; Vásquez et al. 2012).

CZID also uses this resource as the basis of its taxon reports, but, as it is only updated periodically, there were some minor differences between taxa identified as hits by CZID and corresponding classifications in the NCBI taxonomy dump file. In these instances, I updated the taxon reports to reflect the more recent classifications (NCBI). For all analyses, I only used genus-level CZID results (i.e. the evalue, aggregate score, read count and reads per million [rPM]) as species information was not available for all taxa. To collapse species to the genus level I took the minimum,

maximum and sums of the evalue, aggregate score and read counts/rPM, for all species within a genus. To control for potential contamination, CZID uses a "blank" as background to compute a taxon level z-score which reflects the likelihood of a taxonomic hit being a contaminant. As these experiments are from many different laboratories using different reagent kits throughout extraction and sequencing, I selected a generic water as the blank sample as it is likely to be analogous to other molecular grade waters used in sample preparation (specifically, "EARLI Novaseq Water Control").

4.3.4 Generating community count tables

Each CZID taxon report file is produced individually per host sample. Each report file was checked for taxa that matched to non-microbial sources - such as the host, other invertebrates or plants - and removed when found. These files were then iterated through and non-host taxon hits were filtered according to the following criteria: 1) read counts were present above a level of 5 reads per million, 2) alignment length was larger than 50 nucleotides, 3) evalue was below $1e-6$, 4) CZID aggregate and z-scores were above 0, and 5) alignment percent identity was above 90%. This process was run separately for bacteria, eukaryote, and viral taxa hit sequences. For prokaryotic and eukaryotic taxa, the above filters were assigned to the taxa hits mapped against the NT database; the viral taxa were assessed against the NR database results. This is necessary as viruses evolve so rapidly that they can fail to map to the NT database but map perfectly well against the more conserved NR database. Viral taxa were analysed at family level, with bacterial and eukaryote taxa at genus level. Results of each host sample were combined into a single counts table per microbial classification (bacteria, eukaryotes, and viruses).

As some of these sample libraries were prepared with the aim of maximising eukaryotic read yield without consideration for, or by directly reducing, prokaryotic reads, I also opted to assess differences in this approach's ability to detect unique bacterial genera across different library preparations. Using the library preparations

as reported for each project in the SRA (see **S4.1**), I used the Kruskal-Wallis rank sum test to check for significant difference between library preparations in relation to number of unique detected prokaryotic genera. Pairwise comparisons were undertaken using Dunn's test of multiple comparisons with Bonferroni correction for multiple testing. Only library preparations that were reported for at least three samples were considered. All samples where library preparation was not available were classified as "Not specified".

4.3.5 Beta diversity (dissimilarity) analyses

Count tables were further reduced by removing host samples that had fewer than 100 non-host reads total and microbial taxa that were present in less than 5% of the remaining samples. As sample phylogeny was to be considered in microbial composition, I restricted sample sets to taxa that contained at least four samples to allow for centroid calculation. Host taxa with less than this were removed. In the bacterial analysis, this could be done to the level of tribe, and in the other two analyses, family.

Beta diversity was calculated with `vegan` (version 2.6-4, Dixon 2003) in R (version 4.2.2, R Core Team 2020) and its associated functions. Bray-Curtis dissimilarity matrices were calculated for each microbial category using the function `avgdist` with 10,000 iterations. Rarefaction for each matrix was set to use the lowest number of reads from the smallest sample grouping of sociality - solitary - in order to retain as many samples of that grouping as possible. This read limit was therefore different for each of the three matrices: bacteria $n = 323$, eukaryotes $n = 171$, viruses $n = 111$. Samples with total reads less than this number were discarded. For the virus analysis, two further samples were removed to ensure there were no singletons within social lifestyle, continent or host family factor levels. Rarefied reads were used to make 10000 distance matrices and the final matrix consisted of the average distances computed across these iterations.

Non-metric multidimensional scaling (NMDS) was used to visualise dissimilarities,

computed by `metaMDS`. To assess whether variables of interest - social lifestyle, phylogeny, location - significantly affected community composition I performed permutational multivariate analyses (PERMANOVA) using `adonis2` with 9,999 permutations. Each factor was checked for homogeneity of group dispersion using `betadisper` to compute average distances around the median and ANOVA to test significance of any difference between groups.

4.3.6 Predicting microbial complements

I assessed filtered count data for each microbial grouping to determine prevalence of microbial taxa per host species. Bacterial data was subject to further scrutiny where each tribe of samples was assessed for average relative abundance and prevalence of all detected prokaryotic species. Those at above 50% prevalence and 0.01% average relative abundance per tribe were considered potential members of conserved tribe-level community, termed here as "associate" species. Overlaps of prokaryotic species by sample tribe, family and sociality was also considered. Finally, hosts were checked specifically to see if they contained any of the core phylotypes found to be associated with corbiculate bees in previous studies. Prevalence was calculated per tribe for the corbiculates (Apini, Bombini, Meliponini and Euglossini), with non-corbiculates then being ordered by sociality.

4.4 Results

4.4.1 Sample selection and CZID pipeline

There were initially 285 bee samples that met the selection requirements for download from the SRA. After filtering out samples that had too few counts after host mapping (in non-*A. mellifera* samples), the CZID pipeline, and further filtering steps, there were 254 samples remaining, containing bee tissue from 4 phylogenetic families (Figure 4.1), 14 tribes, 18 genera and 45 species from experiments across six continents (Table 4.1, see **S4.1**). Non-*A. mellifera* samples were mapped against 32 host genome

assemblies (see **Appendix** 6.4: Table 6.10) before being uploaded to CZID.org (Kalantar et al. 2020). There were considerably more *Apis* and *Bombus* samples available and included (92 and 65 samples respectively) and 79.9% of all samples were from the Apidae family, particularly from corbiculate species. 165 samples are obligately eusocial, 59 facultatively eusocial, and 30 solitary. All samples successfully ran through the CZID pipeline (version 7.1), with 97% passing quality control with more than 50% of input reads (see **S4.2**).

4.4.2 Differences in microbial composition

For the viruses, there was no significant effect of sociality, host family or continent where the sample was collected on the data (Figure 4.2:A,D,G, see **S4.6**). In

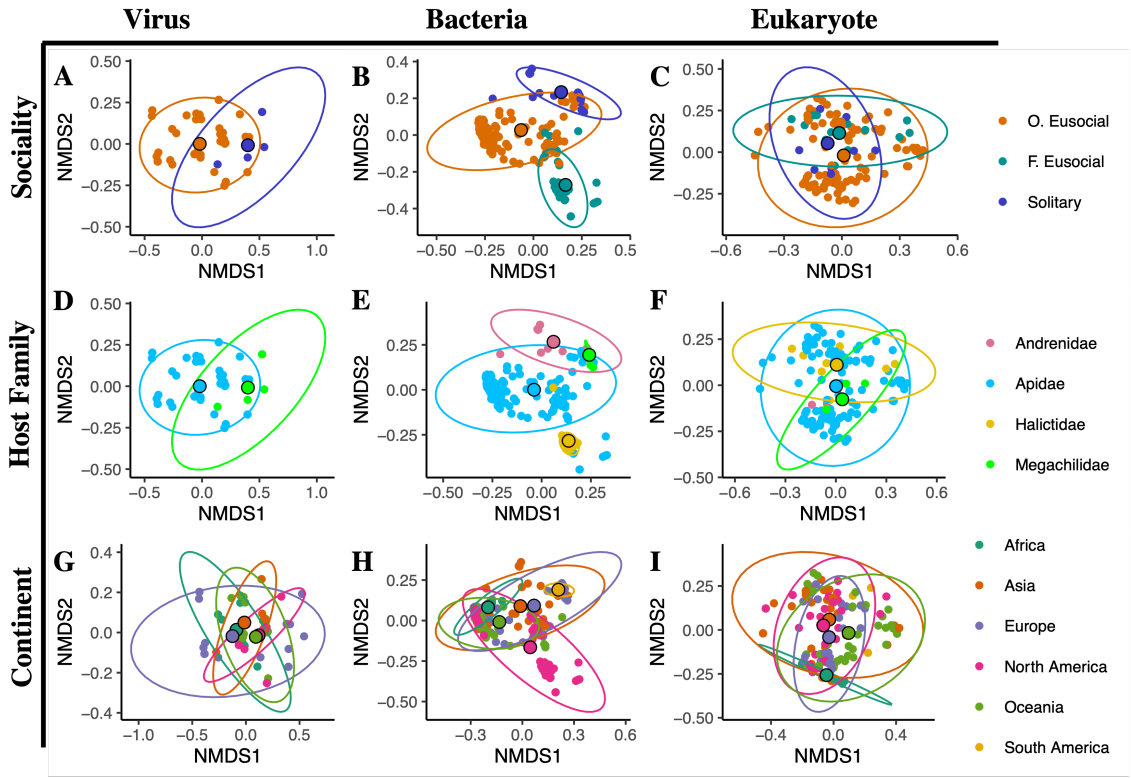

Figure 4.2: NMDS plots of Bray-Curtis dissimilarity matrices computed separately for virus (column 1), bacterial (column 2), and eukaryote (column 3) reads. Three factors were tested to assess influence on composition: sociality (row 1), host family (row 2) and continent where the samples were collected according to NCBI SRA records (row 3). Centroids for each factor level are shown larger and bordered in black. Axes may differ to incorporate full ellipses.

eukaryotes (Figure 4.2:C,F,I), sociality and continent were statistically significant factors (sociality: pseudo-F = 2.271, p = 0.001; continent: pseudo-F = 1.794, p = 0.001), but both are overdispersed, suggesting caution in interpreting these results (see **S4.6**).

Sociality significantly influences bacterial composition (Figure 4.2:B), has homogeneous dispersion (see **S4.6**) and significantly influences the composition of the distance matrix (pseudo-F = 2.884, p = 0.001). This was mostly driven by the differences between obligately and facultatively eusocial samples (Pairwise PERMANOVA: $p = 0.0195$, Benjamini-Hochburg correction, see **Appendix** 6.4: Table 6.11). Host family and continent (Figure 4.2:E,H) both also significantly affected bacterial composition (pseudo-F = 4.318, $p = 1e-04$ and pseudo-F = 2.361, $p = 1e-04$, respectively), and are unaffected by heterogeneous dispersion (see **Appendix** 6.4: Table 6.11 for pairwise PERMANOVA).

4.4.3 Detected microbial complements

Bacterial community

For the bacterial data there were sufficient reads in 227 samples from 10 bee genera resulting in the detection of 65 prokaryotic taxa (Figure 4.3, see **S4.3**). There were few significant variations between the reported methods of library preparation, as indicated in the SRA metadata for each sample, and the number of detected prokaryotic genera (see **Appendix** 6.4: Figure 6.2, also Table 6.12). The most taxa-rich host family was Apidae, which had unique taxa, while all taxa detected in other families were also present in Apidae (see **Appendix** 6.4: Figure 6.3). There were no bacterial taxa found only in solitary hosts, whereas there were 1 and 11 taxa unique to facultatively and obligately eusocial hosts, respectively. The former was *Asticcacaulis*, an associate bacterial taxa of Euglossini samples (Table 4.2), and the latter consisted of *Lactobacillus: Firm-5*, *Bartonella*, *Apibacter*, *Alcaligenes*, *Brevibacterium*, *Citrobacter*, *Deinoccocus*, *Enterobacter*, *Orbus*, *Prevotella* and *Shigella*. The majority of detected taxa belong to the Proteobacteria phylum.

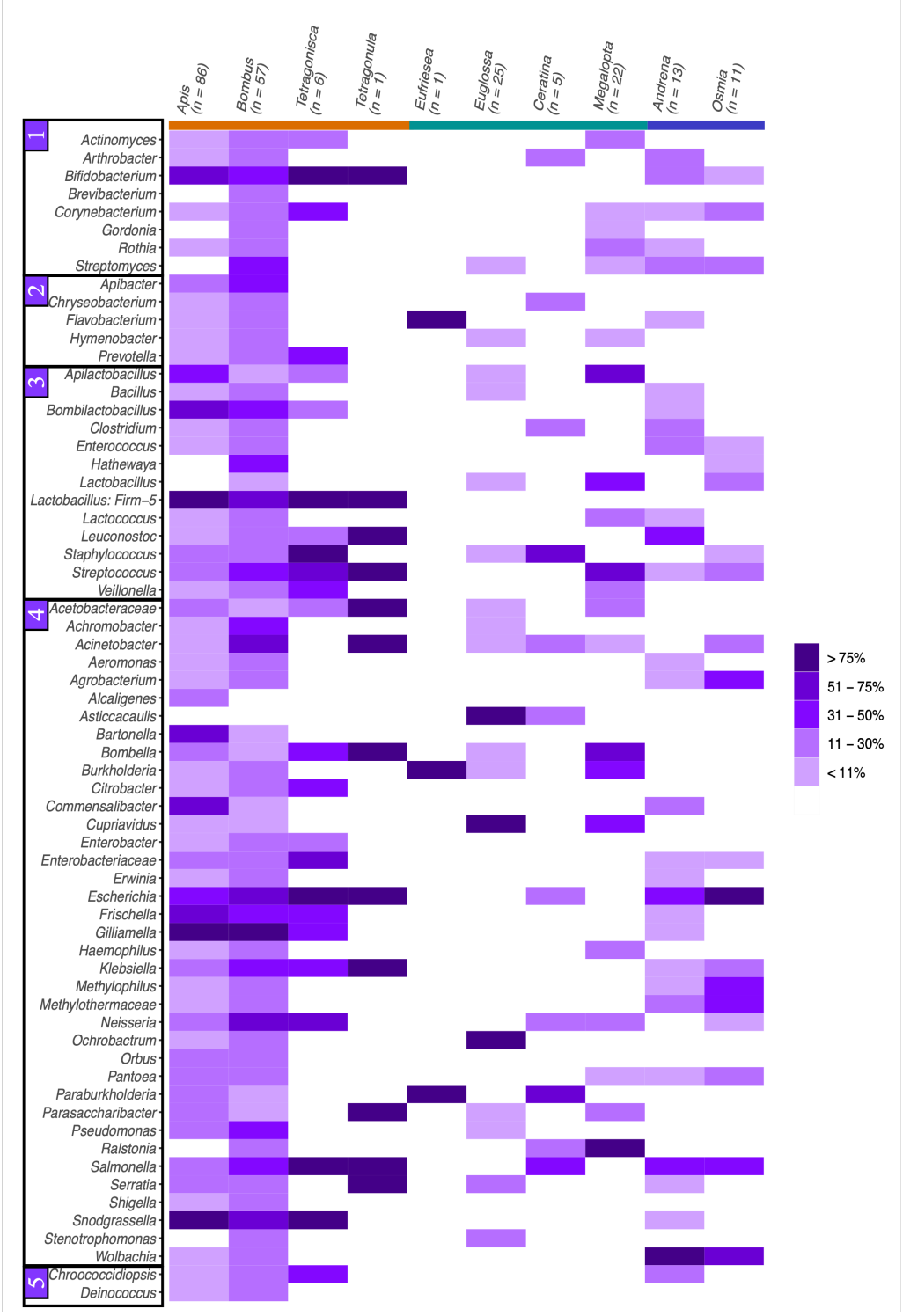

Figure 4.3: Heatmap of bacterial prevalence in each genus of host. Bacterial taxa are ordered 1) Actinobacteria, 2) Bacteroidota, 3) Firmicutes, 4) Proteobacteria and 5) other. Host genera are coloured by sociality: orange = obligately eusocial, green = facultatively eusocial, blue = solitary.

Eukaryotic and viral taxa

There were considerably fewer samples available for determining eukaryote and virus composition after filtering steps (see **S4.4-4.5**). In 158 samples, I identified 32 eukaryotic taxa, including 24 fungi and five genera from the parasitic family Trypanosomatidae (see **Appendix** 6.4: Figure 6.4). The two fungal genera *Alternaria* and *Aspergillus* were detected in the majority of species, appearing in 13 and 11 out of 17 species, respectively. 12 viral families - six of which from the phylum Pisuviricota - were found across 88 host samples (see **Appendix** 6.4: Figure 6.5).

Tribe-bacterial associates

In the more commonly studied corbiculate tribes - Apini, Bombini and Meliponini - I find at least two previously described "corbiculate" core phylotypes as associate taxa (Table 4.2). All eight of the taxa associated with Apini are included in the core phylotypes. All three of these tribes share an association with *Snodgrassella*, yet there is no overlap between associate taxa of these three and the other corbiculate tribe, Euglossini. *Wolbachia* is an associate of the two solitary tribes included in this analysis.

Table 4.2: Associate bacterial taxa found at above 50% prevalence and 0.01% relative abundance per tribe. Tribe cells are coloured according to host family: blue for Apidae, red for Andrenidae, yellow for Halictidae, green for Megachilidae. "Corbiculate" core bacterial taxa are indicated with *. Only tribes included in the bacterial dissimilarity matrix were assessed (see **S4.1).**

Tribe	**n**	**Associate Taxa**
Andrenini	13	*Wolbachia*
Apini	86	*Bartonella**, *Bifidobacterium **, *Bombilactobacillus**, *Commensalibacter**, *Frischella**, *Gilliamella**, *Lactobacillus: Firm-5**, *Snodgrassella**
Augochlorini	22	*Apilactobacillus**, *Bombella**, *Ralstonia*, *Streptococcus*
Bombini	57	*Acinetobacter*, *Escherichia*, *Gilliamella**, *Lactobacillus: Firm-5**, *Snodgrassella**
Ceratini	5	*Paraburkholderia*, *Staphylococcus*
Euglossini	26	*Asticcacaulis*, *Cupriavidus*, *Ochrobactrum*
Meliponini	7	*Bifidobacterium**, *Escherichia*, *Snodgrassella**, *Staphylococcus*
Osmiini	11	*Escherichia*, *Wolbachia*

Corbiculate core taxa

Repeating what has been described previously, I find that the "corbiculate" core bacterial taxa are widely prevalent in the three previously studied obligately eusocial tribes: Apini, Bombini and Meliponini (Figure 4.4, see **Appendix** 6.4: Table 6.13). This pattern was not repeated in Euglossini, however, where only *Apilactobacillus* was detected at low average relative abundance and prevalence. *Apilactobacillus* was interestingly found at considerable prevalence in facultatively eusocial host species, specifically in the *Megalopta* genus, where it was detected in 15/22 available samples.

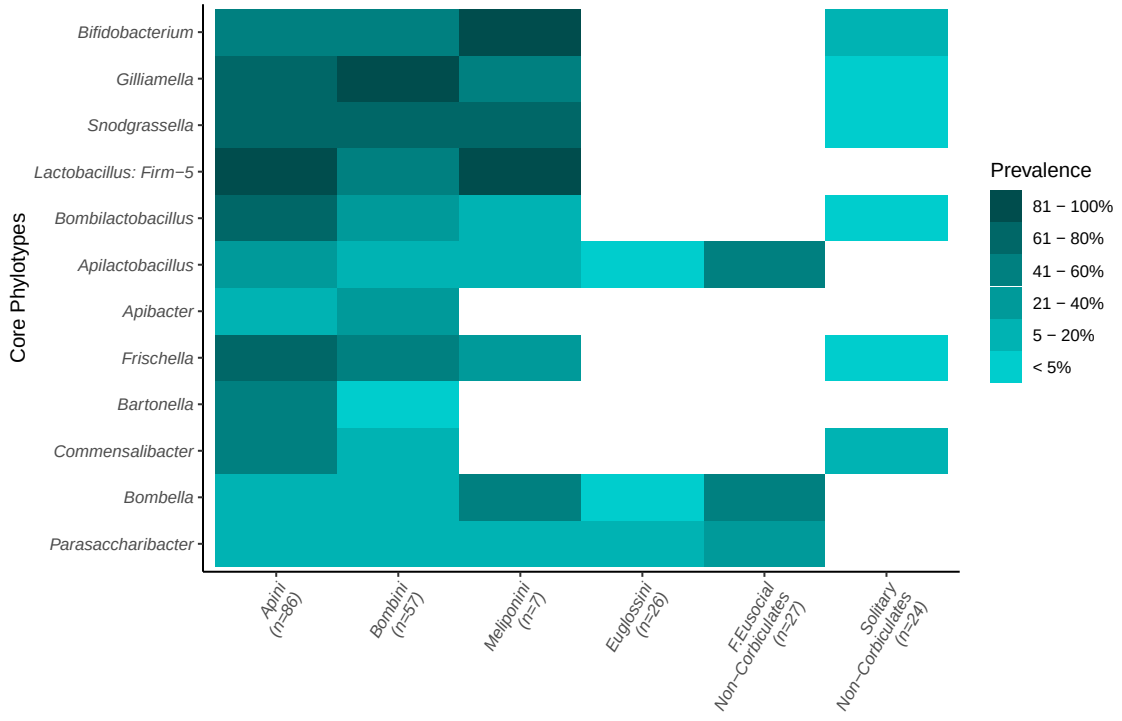

Figure 4.4: Prevalence of different microbial taxa previously described in the literature as part of the "corbiculate" core bacteria across samples. Darker tiles indicate higher prevalence.

Other bacterial phylotypes were detected in three solitary samples: *Bifidobacterium* was detected in one individual *Andrena haemorrhoa* sample (SRR6148367), an individual *Osmia cornuta* (SRR6148371) and in a sample consisting of pooled *Andrena* individuals of different species (SRR13404633). In the latter, *Gilliamella*, *Snodgrassela*, *Bombilactobacillus* and *Frischella* were also detected. *Bombiscardovia*

and *Candidatus Schmidhempelia*, both taxa previously found associated with *Bombus* bees, were not detected in the analysis after filtering. *Apilactobacillus*, *Bombella* and *Parasaccharibacter* were at considerable prevalence in the facultatively eusocial bees. These values are driven largely by *Megalopta* samples (Figure 4.3).

4.5 Discussion

4.5.1 Bacterial complement affected by location, phylogeny and sociality

Though no firm relationships were detected in viral or eukaryotic community composition, I find bacterial communities to be significantly affected by social lifestyle, family and collection location of the bee (Figure 4.2:B,E,H, see **S4.6**). Location and phylogeny has been found to be significant drivers of bee bacterial communities elsewhere, but there isn't always consensus on which is more important. While some studies can identify communities to specific subfamilies or even species (Dew et al. 2020; Kwong and Moran 2015; Kwong et al. 2017a), others find location to be more informative (Kapheim et al. 2021; Keller et al. 2013; McFrederick and Rehan 2016; McFrederick et al. 2017; McFrederick and Rehan 2019), though often both play a significant role (McFrederick et al. 2012; Shell and Rehan 2022).

It is likely that the contribution of either factor is further determined by the social lifestyle of the bee: social living allows for transmission of symbiont species in eusocial insect societies, where this vertical transmission route allows for coevolution of unique and detectable host-microbe associations (Dietrich et al. 2014; Kwong et al. 2017a; Lombardo 2008; Sanders et al. 2014; Zhang and Zheng 2022). Solitary animals, on the other hand, are likely to have less stable communities that are largely acquired from the immediate environment (Voulgari-Kokota et al. 2019). For some of the obligately eusocial samples this can be seen in my data: when the bacterial community NMDS plots are coloured by tribe, there is a clear cluster of Apini samples to the left of the NMDS1 (Figure 4.5), despite the fact that these samples came from 11 different countries across five continents (Table 4.1, see **S4.1**).

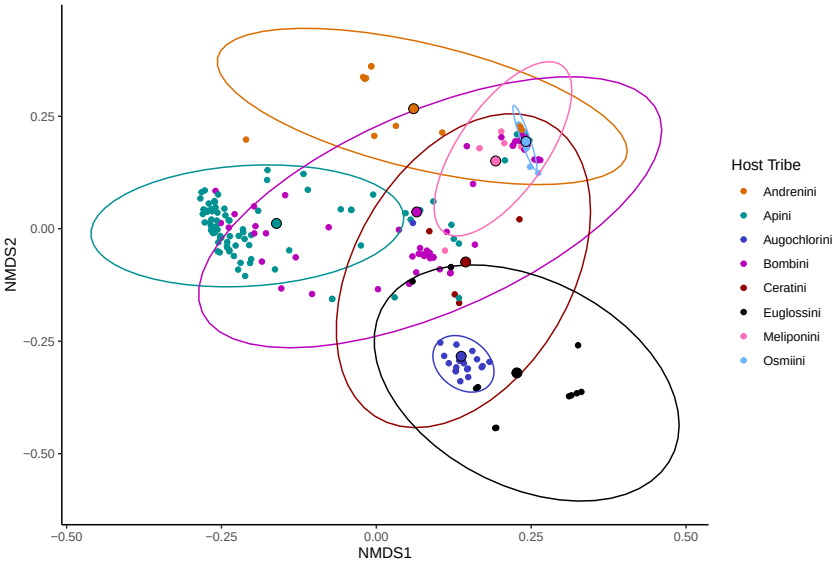

Figure 4.5: NMDS plot of Bray-Curtis dissimilarity matrix computed for bacterial read counts shows clustering of samples by host tribe. Centroids for each factor level are shown larger and bordered in black.

The limited availability of samples from solitary species does, in turn, somewhat limit the ability to untangle the microbial community composition of solitary bees. Specifically, the samples I analysed from solitary tribes Andrenini and Osmiini were primarily derived from a handful of studies (see **S4.1**). However, if solitary species have microbiomes that are predominantly environmentally acquired and lack the consistent vertical transmission of eusocial bees, then they should be more variable and show greater dispersion around the median than more social groups. I do find this (see **S4.6**), but the differences in variance is small and non-significant. Future work that includes many more solitary samples would be able to better test whether solitary species have more variable microbial communities than the well characterised and more strongly vertically transmitted social microbiomes.

4.5.2 Social lifestyle impacts number and type of associate taxa

Tribes made of obligately eusocial species have the most associate microbe species in this analysis (Table 4.2, Figure 4.6:A), with Apini, Bombini and Meliponini being associated with eight, five and four bacterial genera respectively. Of these, at least two bacterial taxa were from the identified "corbiculate" core per tribe (Figure 4.4). This again lends weight to the hypothesis that vertical transmission leads to more stable communities in the social bees, allowing for the establishment of multiple fixed

associations. I also detect more associated bacterial genera with increasing number of samples (Figure 4.6:B), though Meliponini has double the identified associate taxa from fairly few samples relative to the solitary tribes.

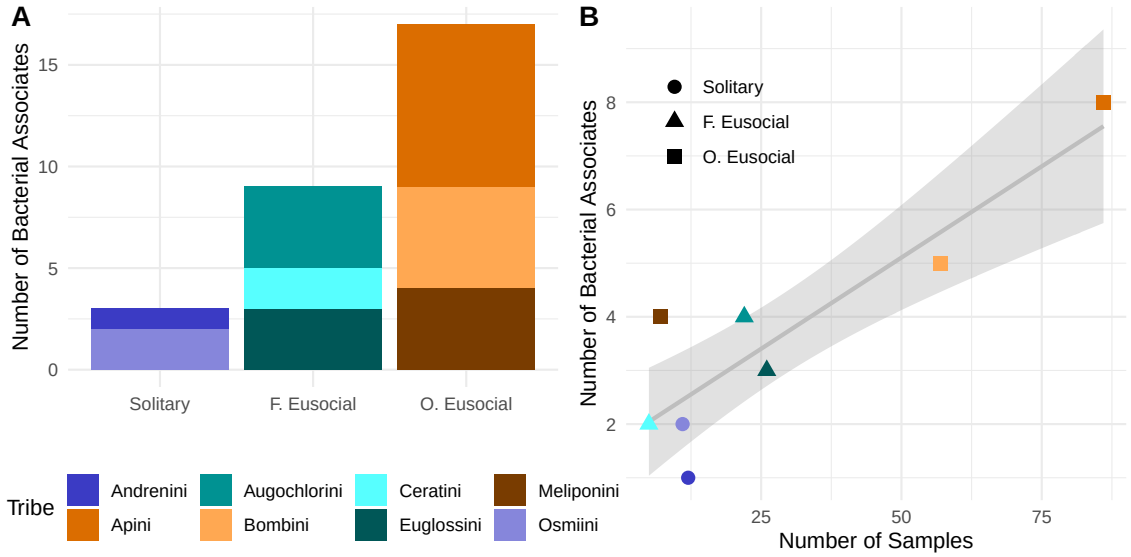

Figure 4.6: Number of bacterial associates per tribe increase with both A) sociality and B) number of samples. Meliponini has four associates based off only seven samples and potentially challenges this trend. Line in plot B represents fit estimated by GLM with standard error shaded in grey.

I find *Wolbachia* associated with the two solitary tribes, Andrenini and Osmiini. *Wolbachia* was also detected at low prevalence in Apini and Bombini (Figure 4.3), but at comparably low average relative abundance (see **Appendix** 6.4: Table 6.14). *Wolbachia* is an extremely successful insect endosymbiont, estimated to be present in as much as 52% of all insect species (Weinert et al. 2015). This endosymbiont is capable of manipulating the reproduction of its host in order to spread throughout populations, most famously by inducing cytoplasmic incompatibility (Bourtzis et al. 1996; Werren et al. 2008), and has been proposed to be a potential factor behind *Andrena* diversification (McLaughlin et al. 2023). In the bees, increased *Wolbachia* prevalence and diversity associated with solitary over social species has been described before (Gerth et al. 2011; Gerth et al. 2015; Ramalho et al. 2021; Saeed and White 2015), though the reasons for this remain speculative.

As *Wolbachia* is maternally inherited, it may be that obligately eusocial societies

that consist of many sterile or reproductively-constrained females would be considered an evolutionary dead-end for the symbiont, if it were not established that *Wolbachia* persists in high prevalence in a number of eusocial ant species (Ramalho et al. 2018; Ramalho et al. 2021; Russell 2012). It has been previously proposed that this disparity in *Wolbachia* presence between social and solitary bees occurs either due to solitary individuals having a greater number of interactions with other potentially infected taxa, or that social species have a more limited number of ecological environments within which they forage and live (Ramalho et al. 2021). Further to this, other factors of social bee lifestyle, such as living in colony structures that are somewhat removed from the environment or else the use of social immunity behaviours, such as adding antibacterial substances to honey before feeding, reduce risk of infection (Kwakman et al. 2011).

I postulate that perhaps this is also to do with obligately eusocial bees having these evolutionary long-term and stable host-microbe relationships that solitary insects are not able to achieve with their relative lack of social and inter-generational interaction. Perhaps *Wolbachia* fails to persist in social bees because the established community protects against it, at least in the case of the most social corbiculates. This phenomenon is termed "colonisation resistance" (Lawley and Walker 2013), and many features of the social bee core microbes already identified could play a part, such as priming the host immune system (Horak et al. 2020; Kwong et al. 2017b; Lang et al. 2022; Näpflin and Schmid-Hempel 2016) or the occurrence of direct antagonistic microbe-invader interactions (Dyrhage et al. 2022; Endo et al. 2012; Endo and Salminen 2013; Koch and Schmid-Hempel 2012; Steele et al. 2017; Vásquez et al. 2012). Solitary bees - such as *Andrena* species (McLaughlin et al. 2023) - missing these interconnected communities therefore would lack the protection they confer and may become vulnerable to *Wolbachia* driven reproductive manipulation. On the other hand, though often parasitic, *Wolbachia* can be advantageous to hosts conferring nutritional or fecundity benefits (Andersen et al. 2012; Cheng et al. 2019; Singh and Linksvayer 2020) or else resistance to viral or parasitic infection (Bian et al.

2010; Cogni et al. 2021; Duplouy et al. 2015; Pimentel et al. 2021; Van Den Hurk et al. 2012). Future work testing whether *Wolbachia* are beneficial or virulent symbionts in solitary bee species would be most welcome.

4.5.3 Potential first members of shared orchid bee microbiome identified

Despite being an important group of pollinators, the orchid bees (Euglossini) remain the least studied group of corbiculate bees and, to the best of my knowledge, the microbiomes are undescribed. Two of the three orchid bee species included in this analysis - *Euglossa dilemma* and *E. viridissima* - exhibit some primitively eusocial behaviour, where a mother foundress and a subordinate daughter (sometimes two) administer brood care (Cocom Pech et al. 2008; Saleh et al. 2022). In these instances, there is the opportunity of vertically transmitted microbes becoming established across generations, though the fact that some daughters leave the nest after eclosure would suggest these relationships could be less stable than those in obligately eusocial corbiculates. In this analysis - looking at 26 orchid bee samples - I find three Euglossini associate microbial taxa: *Asticcacaulis*, *Cupriavidus* and *Ochrobactrum* which represent the first description of the microbiota of this important group of corbiculate bees.

Asticcacaulis are Gram-negative bacteria that are constituents of microbial communities in freshwater, bark, and soil environments (Aschenbrenner et al. 2017; Ishizawa et al. 2019; Xie et al. 2015; Shouke Zhang et al. 2022). Members of this genera also have potentially symbiotic relationships with plants (Jha et al. 2020; Rajkumar et al. 2009), arachnids (Hu et al. 2019; Zhu et al. 2020) and Hemipteran insects (Cooper et al. 2017). As a plant endophyte, the genus may play a role in phosphate solubilisation, though this is yet to be definitively confirmed (Rajkumar et al. 2009). Its role as an insect endosymbiont is not well understood and may simply be a result of its widespread presence in the environment.

Cupriavidus species are known for their ability to tolerate and utilize a wide range of compounds and toxic pollutants, making them important in the bioremediation of

contaminated environments (Malik et al. 2021; Sohn et al. 2021). This genus includes a well-characterised plant symbiont, *Cupriavidus taiwanensis* (previously *Ralstonia*, W.-M. Chen et al. 2003) and has otherwise been detected in the microbiome of mosquitoes (Mancini et al. 2018) and a beetle (Garcia et al. 2014). *Cupriavidus* is a member of the Burkholderiaceae family, members of which are known to possess genomic islands that may increase its propensity for forming symbiotic relationships with insects (Stillson et al. 2022). Despite this, and despite its apparent detection in other insects, *Cupriavidus* genomes assessed to date do not contain this "symbiosis" island and attempts to have other genera colonise beetle samples experimentally had uncertain results (Acevedo et al. 2021). There is, however, evidence of these symbiosis islands transferring between different Burkholderiaceae genera via horizontal transfer (Stillson et al. 2022). Perhaps the species I detect at genus-level here may well contain this island or similar upon more thorough sequencing. Notably, this island was detected in species of *Paraburkholderia*, which I find associated with Ceratini.

The final detected associate taxa, *Ochrobactrum*, includes genera that are known to be beneficial symbionts in plants (Babalola 2010; Balachandar et al. 2007). It has also been identified as an endosymbiont in snails (Dar et al. 2015), root-feeding beetles (Huang et al. 2012) and termites (Tsegaye et al. 2019; Wenzel et al. 2002). Within these relationships, *Ochrobactrum* plays a cellulytic role, helping to break down cellulose and depolymerise lignin. *Euglossa* species enlarge tree cavities for nesting (Dressler 1982) and perhaps to collect resin, which is an important nest building material (Cameron 2004). For these purposes, perhaps *Ochrobactrum*'s cellulytic abilities is advantageous. This work is the first attempt at characterising the microbiome of orchid bees, and further experimental work is required to confirm these relationships and elucidate potential functions.

4.5.4 "Corbiculate" core microbes may be specific to the obligately eusocial clade

When it comes to bee microbiota, the literature overwhelmingly studies the obligately eusocial corbiculate bees, Apini, Bombini and Meliponini. This attention has identified a corbiculate core microbiome that is shared amongst these tribes (Kwong et al. 2017a). These previous studies, however, do not include Euglossini, the fourth corbiculate tribe. Here I find that these core microbial taxa are not found in the same composition or prevalence as they are in the other corbiculates (Figure 4.4, see **Appendix** 6.4: Table 6.13). Perhaps this "corbiculate" core community is a misnomer, and that what had been previously described were communities shared only between the obligately eusocial corbiculates.

There are phylogenetic implications of this insight. While the phylogeny of the corbiculates has historically been controversial, most analyses today place Euglossini as the outgroup to the other three tribes (Bossert et al. 2019; Engel and Rasmussen 2021). Potentially, then, this core microbiome shared between Apini, Bombini and Meliponini may be as ancient as their last common ancestor (LCA), and was composed after the split between the orchid bees and other corbiculates. It could therefore be argued that this LCA would have likely been obligately eusocial, allowing these bacterial communities to establish stably enough to be passed on to three different lineages through $\sim$ 55 million years of host diversification (Peters et al. 2017).

It is worth noting that, in this analysis, our Euglossine sample size was limited ($n = 26$), and mostly consisted of *Euglossa* samples. Larger sample sizes and more species may reveal a more complicated picture of Euglossine species presenting with some or all of the "corbiculate" core microbes. However, the sample size for the Meliponini bees in this analysis was considerably smaller ($n = 7$), and yet this core community was detectable. Further microbiological investigation into the Euglossine bees would be helpful to confirm my findings.

4.5.5 Bacteria with anti-pathogen potential persist across bee taxa

Though the "corbiculate" core community was not similar between Euglossini and the classic corbiculate tribes, there were shared microbial taxa. *Apilactobacillus* and *Bombella / Parasaccharibacter* - likely to actually be one genus (Smith et al. 2021) and referred to hereafter as *Bombella* - were detected not just in the orchid bees but also at considerable prevalence in *Megalopta* (Figure 4.4, see **Appendix** 6.4: Table 6.13). Similarly, both *Apilactobacillus* and *Bombella* were detected in five and six of the ten species included in the bacterial analysis, respectively (Figure 4.3), though not in either of the solitary genera.

One of the reasons why these two bacterial groups are so successful at establishing in such diverse bee taxa may be their roles as anti-pathogen symbionts. *Bombella*, for example, has anti-fungal properties (Miller et al. 2021) and are found frequently in honeybee larvae and food stores, two components of the colony which are especially vulnerable to fungal infection (Anderson et al. 2014). This would also be an advantage to any host that stores pollen, and could help explain its presence in most of the social species in this analysis. *Apilactobacillus* increases individual resistance to a number of pathogens including *Paenibacillus larvae* (American foulbrood, Butler et al. 2013; Forsgren et al. 2010; Kačániová et al. 2020; Kiran et al. 2022), the microsporidian *Nosema* (Arredondo et al. 2018), fungal infection (Iorizzo et al. 2020) and *Melissococcus plutonius* (European foulbrood, Endo et al. 2012; Endo and Salminen 2013; Vásquez et al. 2012; Zendo et al. 2020). It is also prevalent in the floral environment, suggesting an intuitive route for transmission between different bee species visiting the same flowers (Anderson et al. 2013; Tamarit et al. 2015).

Many other members of the "corbiculate" core community may also confer resistance to common bee pathogens. *Snodgrassella* increases honeybee resistance to *Serratia marcescens* infection (Horak et al. 2020), in bumblebees *Gilliamella* and *Apibacter* suppress trypanosomatid *Crithidia* species (Cariveau et al. 2014; Mockler et al. 2018), and members of *Lactobacillus: Firm-5* inhibit *P. larvae* and *M. plutonius* growth (Killer et al. 2014a) and *C. bombi* infection in bumblebees

(Mockler et al. 2018). The mechanisms of this protection could be host moderated, e.g. by increasing the expression of immune-associated genes (Horak et al. 2020; Kwong et al. 2017b), allowing for immune priming (Milutinović et al. 2016; Sadd and Schmid-Hempel 2006), or symbiont moderated, e.g. by creating a physical barrier to pathogen colonisation (Martinson et al. 2012; Kwong and Moran 2013) or producing anti-pathogen molecules (Dyrhage et al. 2022; Endo et al. 2012; Endo and Salminen 2013; Koch and Schmid-Hempel 2012; Steele et al. 2017; Vásquez et al. 2012).

The preponderance of anti-pathogen effects by bee associated microbes may be linked to the immune gene architecture of bees. When the honeybee genome was first sequenced (Honeybee Genome Sequencing Consortium and others 2006), one of the curious features was the relative lack of immune genes compared to other insect models (Evans et al. 2006). This was surprising for the honeybee, a eusocial insect that lives in societies of thousands of genetically similar individuals that are thus vulnerable to pathogen spread. Initially, this disparity was explained by the unique benefits of social immunity - a suite of behaviours that social animals use to help prevent and slow disease transmission, such as allogrooming and expulsion of the sick (Cremer et al. 2007; Cremer et al. 2018; Dolezal and Toth 2014; Wilson-Rich et al. 2009) - leading to relaxed selection on individual immunity and, eventually, gene loss. However, as more bee genomes became available, it became clear that this depauperate immune gene repertoire predated sociality within the bees (Barribeau et al. 2015).

This restricted immune genetic architecture could perhaps be why *Apilactobacillus* is often found outside of the classic corbiculate bees, as is found in this analysis and elsewhere (Handy et al. 2022). In *Apilactobacillus kunkeei*, a plasmid causes one strain's antibacterial effects against *M. plutonius* (Endo and Salminen 2013; Zendo et al. 2020). Upon further investigation, more plasmids putatively encoding antibiotic compounds were discovered in other strains (Dyrhage et al. 2022). Similarly, *Apilactobacillus kunkeei* is usually found as multiple strains within hosts where transfer of mobile genetic elements are common (Tamarit et al. 2015). These features

allow for the rapid evolution of *Apilactobacillus* and may represent an example of an extended immune phenotype where the genetic potential of *Apilactobacillus* - and, perhaps, many other strains of bee-associated taxa - compensates for the relatively restricted host immune genetic potential. It is also possible that similar extended immunity phenotypes are occurring in the solitary bees - for example, the putative antiviral capability of *Wolbachia* - but these would require further investigation. It is likely that the relative lack of social-contact driven vertical transmission within the solitary species means that such relationships, when they occur, may be much more specific to solitary taxa and less permanent than what has been found in more social bees. Perhaps there are other species that, like *Wolbachia*, have evolved mechanisms to ensure high-fidelity vertical transmission without the need for consistent social interactions.

4.5.6 Mining RNA-Seq samples recapitulates experimental findings in obligately eusocial corbiculates

The composition of the "corbiculate" core microbiome has been well characterised (Engel et al. 2012; Engel and Moran 2013; Koch and Schmid-Hempel 2011a; Koch et al. 2013; Kwong and Moran 2016b; Kwong et al. 2017a; Lim et al. 2015; Moran et al. 2012), making it a good yardstick against which I could assess the efficacy of using this pipeline to detect microbial communities. Out of the 14 microbes I opted to include as members of this core set, 12 were detected - the supposedly *Bombus*-specific *Bombiscardovia* and *Candidatus Schmidhempelia* were not detected in any samples after filtering. Having several samples per host taxa obviously improves the reliability of any detected compositions or associations, though it should be reiterated that the core microbes were recapitulated in Meliponini samples despite the relative lack of individual samples (Figure 4.4). I also detected the disparity in *Wolbachia* presence and abundance between social and solitary bees (Table 4.2), as previously described (Gerth et al. 2011; Gerth et al. 2015; Saeed and White 2015; Ramalho et al. 2021). It is important to acknowledge that the method of

library selection can influence the detection of bacterial taxa (see **Appendix** 6.4: Figure 6.2 and Table 6.12). The majority of samples in this analysis reportedly employed library selection methods, such as poly-A enrichment (see **S4.1**), which reduce the presence of non-eukaryotic RNA in the sample (Cui et al. 2010). Though this restricts my ability to comment on differences in absolute abundances, this approach has consistently found most of the key bacterial taxa that are expected and thus represents a useful tool to estimate community composition.

While useful, this approach did not reach 100% detection of predicted microbes, and, thus, some individual microbes are potentially being missed. Despite this limitation, I was still able to unveil a number of interesting avenues for further research based on existing sequencing data. These results reveal a series of future questions that would be exciting to explore. For example, do Euglossini species lack the classic "corbiculate" core shared amongst its relatives (Figure 4.4)? Does it instead have specialised associations with microbes not detected elsewhere (Table 4.2)? Would the pattern of increasing numbers of bacterial associates with increasing social complexity hold when more solitary species are included (Figure 4.6)? What are the phylogenetic relationships of cross-host species bacteria such as *Apilactobacillus*? Do obligately eusocial hosts with long-standing microbial relationships act as evolutionary reservoirs for bee symbionts, and how important are flowers and other factors in the epidemiology of gut microbial communities?

4.5.7 Conclusion

By leveraging existing RNA sequencing datasets, I was able to test whether microbial communities are affected by social structure, geography, or phylogeny. I found that bacterial community composition is significantly affected by the social lifestyle, collection location and phylogeny of the host (Figure 4.2, 4.5, see **S4.6**; also **Appendix** 6.4: Table 6.11). In the eukaryotic and viral analyses, however, I failed to detect any factor contributing to community composition that wasn't affected by heterogeneous dispersion (see **S4.6**). It appears that as the complexity of social

lifestyle increases, so too does the number of bacterial associates (Table 4.2, Figure 4.6). This may be expected as prolonged social contact between host generations allows for more reliable vertical transmission and coevolution of host and symbiont. I also provide, for the first time, an initial description of the microbial community of the Euglossine bees, a complement that does not align with the regimented core microbes of their sister corbiculates (Figure 4.4). The anti-pathogen potential of microbial symbionts is massive, which may be how bees compensate for their own restricted immune gene arsenal. This survey has shone a light on many avenues that require future research and also highlighted the need to further investigate bees of varying social lifestyles from outside the classic corbiculates. Hopefully further work into the complicated, genetically mobile world of bee symbionts will further illuminate host-microbe complexities and their role in optimising bee health.

Chapter 5

General Discussion

5.1 Summary of findings

5.1.1 Thesis aims

Bees are highly successful organisms consisting of over 20,000 species spread globally (Ascher and Pickering 2020). Despite having a relatively limited immune gene repertoire, compared to other insects, they have thrived and diversified (Barribeau et al. 2015; Evans et al. 2006). While social immunity may compensate for this apparent immunological deficiency in the eusocial bees (Cremer et al. 2007), it remains unclear how non-eusocial bees, which greatly outnumber eusocial species (Batra 1984), fortify their health and immunity. The question then becomes, how do bees, across different social lifestyles and phylogenetic lineages, make up for their apparent lack of immunological potential? Is it that there are bee- or Hymenoptera-specific immune response genes, overlooked by gene annotation models based on Dipteran insects, like the candidate immune genes suggested by previous studies (Alaux et al. 2010; Doublet et al. 2017; Richard et al. 2012)? Or does a bee's microbiota play an important role in the host's resistance to pathogenic threat, as has been described in some eusocial bees (Engel et al. 2016; Koch and Schmid-Hempel 2012; Raymann et al. 2017)? The aim of this thesis was to begin to investigate these avenues by 1) assessing the genomic architecture of non-canonical candidate

immune genes relative to immune / background genes in an evolutionary framework; 2) triggering the immune response in a diverse set of Hymenoptera to elucidate potential shared or specific responses and 3) assessing the microbial composition of bees from across different lifestyles and considering the immune supporting potential of these communities. This section will consider what was found throughout this thesis and the implications for our understanding of how bees manage pathogen risk across variable social structures.

5.1.2 Conserved immunological function of candidate genes across divergent bee species

First, it should be noted that the bee candidate immune genes were first identified in *A. mellifera* transcriptomic investigations, a species that exists at the extreme end of eusociality. Perhaps at least some of what has been detected in this analysis - and in other honeybee experiments like it - are the genetic underpinnings of social immune behaviours, which do not exist in solitary species. So what evidence have I that these genes are indeed immune-associated, and can this be applied outside of the social bees?

The non-canonical genes share evidence of considerable genomic change with immune genes, particularly receptor genes, such as elevated dN/dS and GC content. GC-biased gene conversion (gBGC) appears to be an important evolutionary driver in the honeybee (Wallberg et al. 2015) and is likely to also be important in other post-origin of sociality bee species that show differences in GC content between genes exhibiting and not exhibiting signals of positive selection. Though recombination, and gBGC, can theoretically mimic the effect of positive selection whilst being neutral or even deleterious (Galtier and Duret 2007), it appears that, at least in the honeybee, it is indicative of pervasive selection sweeps (Wallberg et al. 2015). This is likely a consequence of the low effective population (Ne) of eusocial species, where elevated recombination has potentially been selected for in order to reduce Hill-Robertson interference and restore the ability of selection to act upon advantageous alleles

(Hill and Robertson 1966; Keightley and Otto 2006; Kent and Zayed 2013). In the eusocial bees, recombination is likely a driver of gene diversification (Fischer and Schmid-Hempel 2005; C. F. Kent et al. 2012; Kerstes et al. 2012; H. Liu et al. 2015).

Sharing genomic features with canonically immune genes suggests that these candidate genes may indeed have undergone similar selection pressures, perhaps because of a shared role in immune function. Additionally, this significant difference between the non-canonical and immune receptor genes and background genes occurs regardless of whether the species considered is solitary, or else falls past the origin or elaboration of sociality in the included phylogeny.

My transcriptomic analysis provides additional evidence for both the immunological role of these genes and their consistent function across organisms with different social lifestyles. Specifically, I found that upon immune activation, these candidate genes were induced at a higher rate than genes of other functions, including canonical immune genes, which further supports their putative immunological function. A small number of these were expressed uniformly in the different challenges amongst three bee species separated by about ~ 100 MY divergence (Kumar et al. 2022, accessed March 2023), suggesting a potentially ancient role in the bee immune response that could prospectively predate the origins of sociality in either subfamily (da Silva, Jack 2021; Kapheim et al. 2015; Rehan and Toth 2015; Rehan et al. 2016).

A limitation of this work that should be considered is the restricted number of samples available. The original experimental design for this project was to have 10 samples per treatment per species, but, due to unfortunate and unforeseen circumstances, the resources available for sequencing were drastically reduced. As collaborators had dedicated their time to donate samples, it was decided that statistical power would be somewhat sacrificed in order to include all four Hymenopteran species. This was perhaps a mistake - limiting the experiment to the bees or else the two species I personally administered challenges to may have been a better approach, conserving the other samples until funding became available. Having multiple genotypes for the lab-reared bees would have also been advantageous: this

was available for the bumblebees, which came from three colonies, but, again, when resources became limited the decision was made to reduce experimental noise at the possible expense of biological variation.

There was also the issue of the challenges themselves - for example, perhaps the challenges should have had dosages dependent on the size of the animal. This protocol worked very well in honeybees and - slightly less so - in bumblebees, though did not work well with the wasps. *Polistes lanio* are large insects, and it could be that more concentrated dosages would be needed to elucidate a clear response. Else there was just too much transcriptional noise with the field-caught animals. However, despite these drawbacks, the experiment did home in on some orthologous genes and gene families that were expressed in uniformity across the four species, and especially in the three bees. These data are still valuable in adding to the larger picture of immunity in bees and wasps.

5.1.3 The hologenome framework: how microbiomes must be considered in understanding bee disease resistance across social lifestyles

The importance of the microbiome in bee immunity was perhaps first most clearly demonstrated in the bumblebee, where the microbial community was a more important driver of disease resistance than either host or parasite genotype (Koch and Schmid-Hempel 2011b; Koch and Schmid-Hempel 2012). The authors suggested that the microbiome of the bee should be considered as an "extended immune phenotype", a non-host component of bee immunity. Another way to consider this is using a hologenome framework (Zilber-Rosenberg and Rosenberg 2008), wherein the genomes of both host and microbiota are considered together as a single unit evolving in tandem in response to environmental pressures, including pathogen exposure. The importance of microbial players in determining the health of bees that have been studied to date is clear (Engel et al. 2016; Raymann and Moran 2018), though there remains a relative paucity of knowledge regarding solitary bees in this area (Voulgari-Kokota et al. 2019).

The obligately eusocial corbiculate microbiomes are all dominated by a few shared core microbes (Kwong et al. 2017a), most of which I was able to detect in the obligately eusocial corbiculate tribes in Chapter 4. This implies that the core members of this community and their hosts have been coevolving at least as long as the split between Apini, Bombini and Meliponini, as old as $\sim$ 55 MY (Peters et al. 2017). The maintenance of this ancient relationship can occur due to the extensive and consistent social contact between conspecifics, allowing for the stable vertical transmission of microbial communities across generations. According to the hologenome framework, we can anticipate that these communities have undergone similar selection pressures as their hosts, and that some of the established microbial symbionts may provide resistance to invading microbes. This is because maintaining an equilibrium in the established community benefits both the host and the symbiont. Indeed, many of these core microbes have been found to confer resistance to bee pathogens (Table 1.3).

This is a relatively straightforward system in the obligately eusocial bees: social contact allows for stable and consistent vertical transmission; the host and symbionts can be considered as a single evolutionary unit facing the same external pressures and coevolving in kind; identified symbiont microbes and their antagonistic interactions with potential pathogens have been demonstrated (reviewed in Engel et al. 2016; Raymann and Moran 2018). But what about other bees from other taxa and social lifestyles? To my knowledge, I presented the first investigation into the "other" corbiculate tribe, Euglossini, and revealed that this "corbiculate" core microbial community was not shared outside of the obligately eusocial tribes. I identified three possible associated symbionts of my Euglossini samples, but their role in the health of their hosts, especially in anti-pathogen defense, is unknown.

The orchid bees I included - mostly *Euglossa* - are primitively eusocial, with one or two daughters sometimes remaining with a foundress to provide brood care (Cocom Pech et al. 2008; Saleh et al. 2022), meaning social contact-mediated bacterial transfer is still likely an important driver of community composition. Solitary bees,

where lifetime contact with conspecifics can be limited to a single act of mating, are expected to acquire their microbiomes from their environment. This has been demonstrated in studies that focused on wholly solitary bees (Kapheim et al. 2021; Keller et al. 2013; McFrederick et al. 2017), but the picture becomes less clear when bees exhibiting social lifestyles between solitary and eusocial are considered (Dew et al. 2020; McFrederick et al. 2012; McFrederick and Rehan 2016; McFrederick and Rehan 2019; Shell and Rehan 2022). I found that both the geographic location (continent) and bee taxa (phylogeny) played important roles in determining the composition of bacterial communities, along with the sociality of the host bee.

Though socially mediated microbial spread may be lower in solitary bees, there are other microbial mechanisms to ensure intergenerational transmission. *Wolbachia*, for example, use cytoplasmic incompatibility to manipulate host reproduction and ensure its propagation via maternal inheritance (Bourtzis et al. 1996; Werren et al. 2008). Interestingly, I found *Wolbachia* associated with the two solitary tribes in my analysis, though its presence in social species was low to absent. This relationship may be purely manipulation on the side of *Wolbachia*, or the symbiont may be conferring some benefits to host health. For example, some *Wolbachia* strains have been suggested to exhibit antiviral properties (Bian et al. 2010; Cogni et al. 2021; Duplouy et al. 2015; Pimentel et al. 2021; Van Den Hurk et al. 2012). However, the potential benefits of this relationship, as well as other possible microbial associates, for enhancing the pathogen resistance ability of solitary bees have yet to be explored.

5.2 Outstanding questions for future work

5.2.1 How can the interplay between selection and recombination be disentangled to understand the evolution of immunity in social bees?

An interesting feature of eusocial bee genomes is their dramatically greater rates of recombination relative to solitary bees and other organisms (Beye et al. 2006; J. C. Jones et al. 2019; Kawakami et al. 2019; Waiker et al. 2021; Wallberg et al.

2015; Wilfert et al. 2007). Studies indicate that genomes with high recombination rates tend to accumulate GC-biased nucleotide substitutions over time, resulting in a positive association between GC content and recombination (Galtier and Duret 2007). In Chapter 2, I showed how positively selected genes tended to have significantly greater GC content than those not considered under positive selection, but only after the advent of sociality. So why is this apparently only occurring in social bees?

Likely this disparity reflects the relationship between eusociality, low Ne, recombination and selection (Kent and Zayed 2013). Eusocial insects have low Ne due to their extreme reproductive skew (Graur 1985; Romiguier et al. 2014). Low Ne can lead to Hill-Robertson interference (Hill and Robertson 1966), where deleterious alleles can be linked to advantageous genes as rates of linkage disequilibrium increases (Keightley and Otto 2006). Over time this can lead to increased genetic load but high recombination rate can reduce interference between advantageous and deleterious alleles and thus begin to rescue the efficacy of selection and overall fitness (Hartfield and Keightley 2012; Kent and Zayed 2013; Webster and Hurst 2012). It is therefore likely that genes that lead to elevated recombination rate then become advantageous in eusocial taxa with low Ne, and are then selected for (Kent and Zayed 2013), leading, over time, to the extreme recombination rates characteristic of eusocial bees (Beye et al. 2006; J. C. Jones et al. 2019; Kawakami et al. 2019; Waiker et al. 2021; Wallberg et al. 2015; Wilfert et al. 2007).

There are indeed indications in the honeybee that GC-biased gene conversion (gBGC), likely a result of recombination, has played a significant role in shaping genome composition, possibly through selective sweeps (Wallberg et al. 2015). In the context of social bees, it has been proposed that recombination has contributed to diversification of genes involved in worker caste differentiation (C. F. Kent et al. 2012; H. Liu et al. 2015) or immunity (Fischer and Schmid-Hempel 2005; Kerstes et al. 2012). My work appears to lend support to the latter, as I find both canonical immune (especially receptors) and non-canonical immune genes exhibiting higher evolutionary rate and GC content.

Further work that could provide support for this hypothesis would involve utilising population data from both social and solitary species already discussed in this thesis. From this Ne and recombination rates may be estimated, and recombination maps of whole genomes be produced. This could then be used to test the hypothesis that species exhibiting more complex eusocial life traits have lower Ne and higher rates of recombination. Recombination maps would then allow to test if these crossover events have occurred more often in genes of canonical or implied immune function than genes of other functions, and how strongly GC-content and recombination is correlated outside of the honeybee. An increased frequency of recombination events occurring in immune associated regions would support the hypothesis that in social insects recombination is a driving force behind immune gene diversification (Fischer and Schmid-Hempel 2005; Kerstes et al. 2012).

Additionally, Kapheim et al. 2015 found that eusocial bees have drastically reduced numbers of transposable elements (TE) relative to simple social and solitary bees, which may be due to the difference in recombination rates. While the presence of TEs is generally inversely correlated with local genomic recombination rate (T. V. Kent et al. 2017), it is uncertain whether this reduction in TEs was a direct response by the host genome to suppress TEs or (more likely) an incidental effect of high recombination.

J. C. Jones et al. 2019 compared evidence of recombination between *A. mellifera* and the solitary *Megachile rotundata* and found that there was evidence of gBGC occurring in the latter, though at a lower rate than in honeybees. The authors' conclusion was that sociality in insects might have required the selection of modifiers that raised recombination rates throughout the genome, but that these modifiers already affected recombination in both solitary and social bees. They refuted previous work that linked recombination to genes of particular functions and found no correlation between protein evolutionary and recombination rates. It is worth mentioning that during the evaluation of this correlation, however, the breakdown of genes into different functions was not considered. The fact that I, and others,

found suggestion of elevated dN/dS rate and recombination in regions associated with sociality or immunity (C. F. Kent et al. 2012; H. Liu et al. 2015; Wallberg et al. 2015) on the other hand, suggests an important and perhaps targeted role of recombination in the evolution of social bee genomes (Wallberg et al. 2015).

There is, overall, constrained adaptive evolution occurring in single-copy orthologs of immune genes in the bees I assessed. This is not dissimilar to what has been found in other social insects (Meusemann et al. 2020). This apparent lack of change at the level of the coding gene may be missing important parts of the genetic underpinning of the immune response, however. Work comparing genomic architecture of bees of different social lifestyles across the bee phylogeny has found that there is a positive correlation between complexity of gene regulation and sociality (Kapheim et al. 2015; Shell et al. 2021). Other studies have provided additional support for this notion by identifying higher levels of selection in non-coding regions, which may have regulatory functions, than in the coding regions of bee genomes (Rubin et al. 2019). With this in mind, future analyses should consider changes in potential regulators of immune genes when elucidating the evolution of the immune response in social taxa.

My selection analysis worked with 11 bee genomes, which was the extent of what was available and of necessary quality at the opening of this PhD. At the time of writing this chapter (March 2023), there are 198 genome assemblies available for Apoidea via NCBI (Sayers et al. 2023), consisting of bees of many diverse social life histories. This is an ample opportunity for large-scale analyses of genome architecture, including assessing patterns of selection in coding and non-coding regions, gene family changes, and recombination in immune-associated regions. Such analyses can shed light on the genetic mechanisms underlying the evolution of immunity in step with sociality.

5.2.2 What candidate genes can be considered functionally immune?

Transcriptional immune assays throughout the insect orders are identifying non-canonical genes being recruited into immune responses (reviewed in Sackton 2019).

Though RNA-Seq experiments such as the one included in this thesis can help allude to the function of particular genes or pathways, there is more that must be done to confirm the potential immunological function of a candidate gene. First, acknowledging a limitation of my work, large numbers of samples should be used to increase power and reliability of assertions. Once particular candidates - such as the eight non-canonical genes I found up-regulated throughout treatments across my three bees - are identified, functionality can be inferred using direct approaches, such as RNA interference (RNAi). For example, this was the approach used in Nelson et al. 2007, which found reducing levels of vitellogenin (*Vg*) caused worker bees to switch to foraging behaviour preemptively.

Studies could also use other advanced gene modification methods, such as knocking out a gene of interest using CRISPR/Cas9 or else producing transgenic bee lines with altered gene function via transposon-mediated transgenesis (reviewed in Kohno and Kubo 2019). Phenotypes of altered transgenic or genome edited bees can be assessed both prior to and after immune challenge, perhaps comparing between both heat-killed and live infections. Combining different approaches and treatments in this way could help elucidate the roles these candidate immune genes play in bee immunity, if any at all.

It is also important to consider the social lifestyle of the bee being studied when interpreting the roles played by candidate genes. When a candidate gene is found to be immunological in one species, other related species from different social lifestyles should be assessed to see if the response is conserved within the bees or else likely to have evolved as part of the immune response because of a transition to social living. To discuss the importance of this we can consider a non-canonical immune gene that was universally differentially expressed in my transcriptomic analysis: *malvolio*.

Malvolio, like *Vg*, has been implicated in causing a shift to foraging behaviour in infected honeybees, which removes the threat of infection away from the colony heart (Alaux et al. 2010; Antonio et al. 2008; Ben-Shahar et al. 2004). This looks to be triggered by *malvolio* up-regulation, which was detected across treatments

in my transcriptomic assay (see Chapter 3). *Malvolio* is likely a manganese/iron transporter (Orgad et al. 1998), and its primary role in immune response may be to sequester ion resources away from potential invaders that require either of these ions to proliferate (Colomer-Winter et al. 2018; Corbin et al. 2008; Hood and Skaar 2012; Hrdina and Iatsenko 2021; Iatsenko et al. 2020; Kehl-Fie et al. 2011; Porcheron et al. 2013). That this is the fundamental role of *malvolio*, rather than its implications in shifting worker honeybee behaviour, is supported by it also being up-regulated in *C. australensis*, an incipiently social bee that exhibits both solitary and social nesting practices (Rehan et al. 2018) upon immune challenge.

This has implications for the evolution of *malvolio* becoming a component of social immunity. Suppose that *malvolio*'s fundamental role in the immune response is to sequester manganese ions out of the haemolymph and into cells. In this scenario, an infected bee will increase *malvolio* expression which would cause an up-regulation in manganese transport (Ben-Shahar et al. 2004). An increase in intracellular manganese transport increases sucrose responsiveness, which, in honeybees, is associated with foraging behaviour (Thamm and Scheiner 2014). Therefore, the shift to foraging behaviour initially occurs as a consequence of the ion sequestration, and secondarily comes to play a role in social immunity where increased precociousness of infected foragers reduces risk to queens or brood (Cremer et al. 2007).

This could be an example of the same gene having different effects in bee taxa when considered in the context of sociality, though further testing in non-honeybee species to assess its regulation and effect on behaviour is required. It is likely that many of the other strong candidates identified from the work in this thesis have similarly complex effects that warrant further investigation, always keeping in mind the importance of assessing bees from different social lifestyles in order to get the entire picture of bee health and immunity.

5.2.3 What is the role that symbionts play in non-eusocial bee health?

The genetic potential of a microbiome can be tenfold higher than that of its host (Morgan et al. 2013). Microbes have short generation times, allowing for rapid evolution of traits that are beneficial for survival, such as developing resistance to pathogenic threat. One good system to study this is in the case of the eusocial corbiculates and *Apilactobacillus kunkeei*. *A. kunkeei* was highlighted in a number of studies as it conferred resistance to a number of pathogens (Arredondo et al. 2018; Butler et al. 2013; Endo et al. 2012; Endo and Salminen 2013; Forsgren et al. 2010; Kačániová et al. 2020; Kiran et al. 2022; Vásquez et al. 2012). Closer investigation has highlighted the genetic mobility characteristics of *A. kunkeei*, with antimicrobial properties being found to be conferred by plasmids (Dyrhage et al. 2022; Zendo et al. 2020). The wide variety of anti-pathogenic capabilities exhibited by different strains of *A. kunkeei* can be attributed to this genetic flexibility.

Could it be that the use of microbial employment as immune defense is primarily specific to social bees, where we are observing these microbe-pathogen interactions? Given that social bees are more likely to be exposed to infection and disease transmission within their communities compared to solitary species, perhaps they evolved to depend on their symbionts to compensate for their limited immune gene repertoire. On the contrary, solitary bees do not encounter the same level of threat and may not require an expanded immune gene set. Therefore, the relatively depauperate immune gene complement that is characteristic of bees may be sufficient for the thriving and persistence of solitary bee species.

This hypothesis can be challenged by expanding on the preliminary explorations of solitary microbiota (reviewed in Voulgari-Kokota et al. 2019). While microbial employment as immune defense may be largely specific to social bees, there may be solitary-specific symbioses that provide anti-pathogenic capability. For example, bacteria from the *Paenibacillus* genus were discovered to be associated with wild bee nesting material (Keller et al. 2013). The bacterium *P. larvae* is a causative agent of American foulbrood, which is a disease threat to honeybee colonies, and

yet the larvae in these nests appeared unaffected. Suspecting that these wild bee *Paenibacillus* populations may be acting as disease reservoirs, researchers isolated and sequenced *Paenibacillus* strains found in wild bees - mostly Megachilids - and their nesting materials (Keller et al. 2018). Contrary to expectation, the wild bee *Paenibacillus* strains were not closely related to *P. larvae*, though there were some genomic indicators of a pathogenic role in its evolutionary past. This strain was commonly found in wild bee nests, guts and across cuticula and contained genes of antimicrobial capability. This led the authors to conclude that this *Paenibacillus* strain, MBD-MB06, is likely a symbiont that protects nests and wild bees from fungal infection.

This may be an instance of an established symbiosis between solitary bees and a microbe where it plays a role in the extended immune phenotype of its host. This raises the possibility of many other as yet undiscovered diverse symbioses among solitary bees and across their different geographic regions. Secondly, it highlights the importance of assessing microbes down to the strain level to elucidate function and possible symbiosis or virulence. Finally, this example underscores the importance of studying solitary bees and avoiding assumptions based on research in social bees.

Could these symbioses then be considered as a hologenome over the evolutionary history of solitary bees and their associated microbes, the way we can with the obligately eusocial corbiculates and their core microbial community? Microbes in solitary systems are likely to be transmitted in an unstable manner between generations via maternal salivary secretions in pollen provisions and nesting materials (Keller et al. 2018; McFrederick et al. 2017). Interestingly, assays of nesting materials and larvae thus far have detected some species known to exhibit antimicrobial potential, such as Lactobacilli (McFrederick et al. 2012) and, notably, *A. kunkeei* (McFrederick et al. 2014), but these species are less abundant in adults. Other microbes may use different mechanisms to ensure their intergenerational spread, such as those utilised by *Wolbachia* (Bourtzis et al. 1996; Werren et al. 2008, discussed above). Additionally, it is unclear whether *Wolbachia*'s association with some solitary

bee species is purely a case of symbiotic manipulation of the host, or whether its presence provides any benefits (Gerth et al. 2011; Gerth et al. 2015; Ramalho et al. 2021; Saeed and White 2015). Further investigation into the microbiomes of solitary bees and the roles they play in maintaining bee health is needed to answer these outstanding questions and understand the specific benefits that microbial symbionts provide to solitary bees.

5.2.4 To what extent do the characteristic features of obligately eusocial corbiculate bees result from their social behaviour versus their phylogeny?

Much of what is known about social bees comes from work with honeybees and bumblebees, both of which are eusocial corbiculate bees from the subfamily Apinae. Some of these features have been discussed throughout this thesis, such as their high recombination rates compared to solitary bees and their strong associations with core microbiota. However, it is unclear whether these features are specific to Apinae bees or characteristic of all social bees.

To shed light on this question, we can look at other subfamilies of Apidae, such as Xylocopinae, which also contain species exhibiting both solitary and advanced eusocial lifestyles, as well as many strategies in between (Rehan and Toth 2015; Shell et al. 2021). Repeating these analyses using species from the latter can help determine which features discovered thus far are specific to Apinae species, and which are common to social lifestyle. Work so far in these areas have identified that independent routes to sociality utilise independent genes and pathways, but gene function and patterns of genomic change share features (Kapheim et al. 2015; Shell et al. 2021). These features include increased levels of selection, gene expansions, and an increased number of taxonomically restricted genes.

Studies have also shown that recombination rates are elevated in social Hymenoptera as a whole, including social wasps and ants, (Beye et al. 2006; H. Liu et al. 2015; Sirviö et al. 2006; Anu Sirviö et al. 2011; Wallberg et al. 2015; Wilfert et al. 2007), but what role recombination plays in the eusocial Xylocopinae remains, to my

knowledge, unexplored. Further, what are the microbial communities of these other eusocial species? How do they compare to the classic "corbiculate" core (Kwong et al. 2017a)? Or else do they provide similar functional capabilities, including protection from disease? Overall, investigating social bees from different subfamilies can help us better understand the evolutionary and genetic underpinnings of social behaviour, immunity and health in the bees.

5.3 Concluding remarks

Understanding bee immunity is a challenging task. Bees have a limited repertoire of immune genes compared to other model insects, yet they have successfully evolved into thousands of species. In this thesis, I explored two main avenues to shed light on this phenomenon: the possibility that bees utilise immune genes from non-canonical pathways, and the potential role of microbial symbionts in extending their immune phenotype. Increasing evidence suggests that some of these non-canonical immune genes have functional roles in bee immunity, but their effectiveness likely depends on the pathogen and evolutionary history of the species. The social behaviour of bees also adds another layer of complexity, as eusocial bees have evolved social immunity behaviours to protect their colonies. Evolutionary processes have driven divergent selection in immune-associated genes between solitary and social bees, reflecting the transition from solitary to social living. The microbiome likely also plays a crucial role in bee immune defense, and certain microbes may compensate for the limited genetic immune potential of bees across different social lifestyles. Future research leveraging the growing number of genomic resources and technologies can further advance our understanding of bee innate immunity.

Bibliography

Acevedo, Tarik S et al. (2021). "The importance of environmentally acquired bacterial symbionts for the squash bug (*Anasa tristis*), a significant agricultural pest". In: *Frontiers in Microbiology* 12, p. 719112.

Ahmad, S Tariq, Sean T Sweeney, Jin-A Lee, Neal T Sweeney, and Fen-Biao Gao (2009). "Genetic screen identifies serpin5 as a regulator of the toll pathway and CHMP2B toxicity associated with frontotemporal dementia". In: *Proceedings of the National Academy of Sciences* 106.(29), pp. 12168–12173.

Alaux, Cédric, Christelle Dantec, Hughes Parrinello, and Yves Le Conte (Oct. 2011). "Nutrigenomics in honey bees: digital gene expression analysis of pollen's nutritive effects on healthy and varroa-parasitized bees". In: *BMC genomics* 12.(1), p. 496.

Alaux, Cédric, François Ducloz, Didier Crauser, and Yves Le Conte (2010). "Diet effects on honeybee immunocompetence". In: *Biology letters* 6.(4), pp. 562–565.

Alem, Sylvain et al. (2016). "Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect". In: *PLoS biology* 14.(10), e1002564.

Alexa, A. and J. Rahnenfuhrer (2016). "Gene set enrichment analysis with topGO". In: *R package version 2.30.0.*

Alger, Samantha A, P Alexander Burnham, Humberto F Boncristiani, and Alison K Brody (2019). "RNA virus spillover from managed honeybees (*Apis mellifera*) to wild bumblebees (*Bombus* spp.)" In: *PloS one* 14.(6), e0217822.

Ali Mohammadie Kojour, Maryam, Yeon Soo Han, and Yong Hun Jo (2020). "An overview of insect innate immunity". In: *Entomological research* 50.(6), pp. 282–291.

Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman (1990). "Basic local alignment search tool". In: *Journal of molecular biology* 215.(3), pp. 403–410.

Amsalem, Etya, David A Galbraith, Jonathan Cnaani, Peter EA Teal, and Christina M Grozinger (2015). "Conservation and modification of genetic and physiological toolkits underpinning diapause in bumble bee queens". In: *Molecular ecology* 24.(22), pp. 5596–5615.

Andersen, SB, Mette Boye, DR Nash, and JJ Boomsma (2012). "Dynamic *Wolbachia* prevalence in *Acromyrmex* leaf-cutting ants: potential for a nutritional symbiosis". In: *Journal of evolutionary biology* 25.(7), pp. 1340–1350.

Anderson, Kirk E, Mark J Carroll, TIM Sheehan, Brendon M Mott, Patrick Maes, and Vanessa Corby-Harris (2014). "Hive-stored pollen of honey bees: many lines of evidence are consistent with pollen preservation, not nutrient conversion". In: *Molecular ecology* 23.(23), pp. 5904–5917.

Anderson, Kirk E et al. (2013). "Microbial ecology of the hive and pollination landscape: bacterial associates from floral nectar, the alimentary tract and stored food of honey bees (*Apis mellifera*)". In: *PloS one* 8.(12), e83125.

Andersson, Marlene, Jan Johansson, and Anna Rising (2016). "Silk spinning in silkworms and spiders". In: *International journal of molecular sciences* 17.(8), p. 1290.

Antonio, David Santos Marco, Karina Rosa Guidugli-Lazzarini, Adriana Mendes Do Nascimento, Zilá Luz Paulino Simões, and Klaus Hartfelder (2008). "RNAi-mediated silencing of vitellogenin gene function turns honeybee *Apis mellifera* workers into extremely precocious foragers". In: *Naturwissenschaften* 95.(10), pp. 953–961.

Araki, Toshiyuki and Jeffrey Milbrandt (1996). "Ninjurin, a novel adhesion molecule, is induced by nerve injury and promotes axonal growth". In: *Neuron* 17.(2), pp. 353–361.

Araujo, Natalia de Souza and Maria Cristina Arias (2021). "Gene expression and epigenetics reveal species-specific mechanisms acting upon common molecular pathways in the evolution of task division in bees". In: *Scientific reports* 11.(1), pp. 1–16.

Arredondo, Daniela et al. (2018). "*Lactobacillus kunkeei* strains decreased the infection by honey bee pathogens *Paenibacillus larvae* and *Nosema ceranae*". In: *Beneficial microbes* 9.(2), pp. 279–290.

Arsenault, Samuel V, Brendan G Hunt, and Sandra M Rehan (2018). "The effect of maternal care on gene expression and DNA methylation in a subsocial bee". In: *Nature communications* 9.(1), p. 3468.

Aschenbrenner, Ines Aline, Tomislav Cernava, Armin Erlacher, Gabriele Berg, and Martin Grube (2017). "Differential sharing and distinct co-occurrence networks among spatially close bacterial microbiota of bark, mosses and lichens". In: *Molecular Ecology* 26.(10), pp. 2826–2838.

Ascher, JS and J Pickering (2020). *Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila)*. `https://www.discoverlife.org/mp/20q?guide=Apoidea_species&flags=HAS:`.

Asgari, Sassan and Karyn N Johnson (2010). *Insect virology*. QR360 I591. Caister Academic.

Avarguès-Weber, Aurore, Daniele d'Amaro, Marita Metzler, J Garcia, and Adrian G Dyer (2017). "Recognition of human face images by the free flying wasp *Vespula vulgaris*". In: *Anim. Behav. Cogn* 4.(3), pp. 314–23.

Babalola, Olubukola Oluranti (2010). "Beneficial bacteria of agricultural importance". In: *Biotechnology letters* 32, pp. 1559–1570.

Balachandar, D, P Raja, K Kumar, and SP Sundaram (2007). "Non-rhizobial nodulation in legumes". In: *Biotechnol Mol Biol Rev* 2.(2), pp. 049–057.

Ballenghien, Marion, Nicolas Faivre, and Nicolas Galtier (2017). "Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions". In: *BMC biology* 15.(1), pp. 1–16.

Bangham, Jenny, Frank Jiggins, and Bruno Lemaitre (2006). "Insect immunity: the post-genomic era". In: *Immunity* 25.(1), pp. 1–5.

Bänsch, Svenja, Teja Tscharntke, Doreen Gabriel, and Catrin Westphal (2021). "Crop pollination services: Complementary resource use by social vs solitary bees facing crops with contrasting flower supply". In: *Journal of Applied Ecology* 58.(3), pp. 476–485.

Barribeau, Seth M et al. (2015). "A depauperate immune repertoire precedes evolution of sociality in bees". In: *Genome biology* 16.(1), pp. 1–21.

Basset, Yves et al. (2012). "Arthropod diversity in a tropical forest". In: *Science* 338.(6113), pp. 1481–1484.

Batra, Suzanne WT (1984). "Solitary bees". In: *Scientific American* 250.(2), pp. 120–127.

Beadle, Katherine et al. (2019). "Genomic insights into neonicotinoid sensitivity in the solitary bee *Osmia bicornis*". In: *PLoS genetics* 15.(2), e1007903.

Bedford, Mark T and Steven G Clarke (2009). "Protein arginine methylation in mammals: who, what, and why". In: *Molecular cell* 33.(1), pp. 1–13.

Belvin, Marcia P and Kathryn V Anderson (1996). "A conserved signaling pathway: the *Drosophila toll-dorsal* pathway". In: *Annual review of cell and developmental biology* 12.(1), pp. 393–416.

Ben-Shahar, Yehuda, Nichole L Dudek, and Gene E Robinson (2004). "Phenotypic deconstruction reveals involvement of manganese transporter *malvolio* in honey bee division of labor". In: *Journal of experimental biology* 207.(19), pp. 3281–3288.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal statistical society: series B (Methodological)* 57.(1), pp. 289–300.

Benoit, Joshua B et al. (2017). "Symbiont-induced odorant binding proteins mediate insect host hematopoiesis". In: *elife* 6, e19535.

Berens, Ali J, James H Hunt, and Amy L Toth (2015). "Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects". In: *Molecular biology and evolution* 32.(3), pp. 690–703.

Bettedi, Lucia, Mohamad F Aslam, Joanna Szular, Konstantinos Mandilaras, and Fanis Missirlis (2011). "Iron depletion in the intestines of Malvolio mutant flies does not occur in the absence of a multicopper oxidase". In: *Journal of Experimental Biology* 214.(6), pp. 971–978.

Beye, Martin et al. (2006). "Exceptionally high levels of recombination across the honey bee genome". In: *Genome research* 16.(11), pp. 1339–1344.

Bian, Guowu, Yao Xu, Peng Lu, Yan Xie, and Zhiyong Xi (2010). "The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*". In: *PLoS pathogens* 6.(4), e1000833.

Biesmeijer, Jacobus C and Thomas D Seeley (2005). "The use of waggle dance information by honey bees throughout their foraging careers". In: *Behavioral Ecology and Sociobiology* 59, pp. 133–142.

Biesmeijer, Jacobus C et al. (2006). "Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands". In: *Science* 313.(5785), pp. 351–354.

Bischoff, Vincent, Cécile Vignal, Ivo G Boneca, Tatiana Michel, Jules A Hoffmann, and Julien Royet (2004). "Function of the *Drosophila* pattern-recognition receptor PGRP-SD in the detection of Gram-positive bacteria". In: *Nature immunology* 5.(11), pp. 1175–1180.

Bonilla-Rosso, Germán and Philipp Engel (2018). "Functional roles and metabolic niches in the honey bee gut microbiota". In: *Current opinion in microbiology* 43, pp. 69–76.

Bonning, Bryony C (2019). "The insect virome: opportunities and challenges". In: *Current issues in molecular biology* 34.(1), pp. 1–12.

Boomsma, Jacobus J and Richard Gawne (2018). "Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation". In: *Biological Reviews* 93.(1), pp. 28–54.

Bossert, Silas, Elizabeth A Murray, Eduardo AB Almeida, Seán G Brady, Bonnie B Blaimer, and Bryan N Danforth (2019). "Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae". In: *Molecular phylogenetics and evolution* 130, pp. 121–131.

Bourtzis, Kostas, Androniki Nirgianaki, George Markakis, and Charalambos Savakis (1996). "*Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species". In: *Genetics* 144.(3), pp. 1063–1073.

Brady, Sean G, Sedonia Sipes, Adam Pearson, and Bryan N Danforth (2006). "Recent and simultaneous origins of eusociality in halictid bees". In: *Proceedings of the Royal Society B: Biological Sciences* 273.(1594), pp. 1643–1649.

Branstetter, Michael G et al. (2017). "Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees". In: *Current Biology* 27.(7), pp. 1019–1025.

Bray, Nicolas L, Harold Pimentel, Páll Melsted, and Lior Pachter (2016). "Near-optimal probabilistic RNA-seq quantification". In: *Nature biotechnology* 34.(5), pp. 525–527.

Brettell, Laura E, Markus Riegler, Corey O'Brien, and James M Cook (2020). "Occurrence of honey bee-associated pathogens in *Varroa*-free pollinator communities". In: *Journal of Invertebrate Pathology* 171, p. 107344.

Brettell, Laura E, Declan C Schroeder, and Stephen J Martin (2019). "RNAseq analysis reveals virus diversity within Hawaiian apiary insect communities". In: *Viruses* 11.(5), p. 397.

Broderick, Sarah, Xiaoxi Wang, Nicholas Simms, and Andrea Page-McCaw (Sept. 2012). "*Drosophila* Ninjurin A induces nonapoptotic cell death". In: *PLOS ONE* 7, pp. 1–13.

Brown, Michael F and Bridget P Austin (2021). "Bees and abstract concepts". In: *Current Opinion in Behavioral Sciences* 37, pp. 140–145.

Brutscher, Laura M, Katie F Daughenbaugh, and Michelle L Flenniken (2017). "Virus and dsRNA-triggered transcriptional responses reveal key components of honey bee antiviral defense". In: *Scientific Reports* 7.(1), pp. 1–15.

Brutscher, Laura M and Michelle L Flenniken (Dec. 2015). "RNAi and Antiviral Defense in the Honey Bee". In: *Journal of Immunology Research* 2015, p. 941897.

Butler, Èile, Magnus Alsterfjord, Tobias C Olofsson, Christofer Karlsson, Johan Malmström, and Alejandra Vásquez (2013). "Proteins of novel lactic acid bacteria from *Apis mellifera mellifera*: an insight into the production of known extracellular proteins during microbial stress". In: *BMC microbiology* 13, pp. 1–12.

Calderone, Nicholas W (2012). "Insect pollinated crops, insect pollinators and US agriculture: trend analysis of aggregate data for the period 1992–2009". In: *PloS one* 7.(5), e37235.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, et al. (2009). "BLAST+: architecture and applications". In: *BMC bioinformatics* 10.(1), pp. 1–9.

Cameron, Sydney A (2004). "Phylogeny and biology of neotropical orchid bees (Euglossini)". In: *Annual Reviews in Entomology* 49.(1), pp. 377–404.

Cameron, Sydney A and Ben M Sadd (2020). "Global trends in bumble bee health". In: *Annual review of entomology* 65, pp. 209–232.

Cantalapiedra, Carlos P, Ana Hernández-Plaza, Ivica Letunic, Peer Bork, and Jaime Huerta-Cepas (2021). "eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale". In: *Molecular biology and evolution* 38.(12), pp. 5825–5829.

Cardinal, Sophie and Bryan N Danforth (2011). "The antiquity and evolutionary history of social behavior in bees". In: *PLOS one* 6.(6), e21086.

Cariveau, Daniel P, J Elijah Powell, Hauke Koch, Rachael Winfree, and Nancy A Moran (2014). "Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (*Bombus*)". In: *The ISME journal* 8.(12), pp. 2369–2379.

Cerenius, Lage, Shun-ichiro Kawabata, Bok Luel Lee, Masaru Nonaka, and Kenneth Söderhäll (2010). "Proteolytic cascades and their involvement in invertebrate immunity". In: *Trends in biochemical sciences* 35.(10), pp. 575–583.

Chen, Quanmei et al. (2015). "Comparative proteomic analysis of silkworm fat body after knocking out fibroin heavy chain gene: a novel insight into cross-talk between tissues". In: *Functional & integrative genomics* 15.(5), pp. 611–637.

Chen, Wen-Ming, Lionel Moulin, Cyril Bontemps, Peter Vandamme, Gilles Béna, and Catherine Boivin-Masson (2003). "Legume Symbiotic Nitrogen Fixation byβ-Proteobacteria Is Widespread inNature". In: *Journal of bacteriology* 185.(24), pp. 7266–7272.

Cheng, Daifeng et al. (2019). "Symbiotic microbiota may reflect host adaptation by resident to invasive ant species". In: *PLoS pathogens* 15.(7), e1007942.

Chittka, Lars and Natacha Rossi (2022). "Social cognition in insects". In: *Trends in Cognitive Sciences* 26.(7), pp. 578–592.

Choe, Kwang-Min, Thomas Werner, Svenja Stoven, Dan Hultmark, and Kathryn V Anderson (2002). "Requirement for a peptidoglycan recognition protein (PGRP) in *Relish* activation and antibacterial immune responses in *Drosophila*". In: *Science* 296.(5566), pp. 359–362.

Cocom Pech, ME, W de J May-Itzá, LA Medina Medina, and JJG Quezada-Euán (2008). "Sociality in *Euglossa (Euglossa) viridissima* Friese (Hymenoptera, Apidae, Euglossini)". In: *Insectes Sociaux* 55, pp. 428–433.

Cogni, Rodrigo, Shuai Dominique Ding, André C Pimentel, Jonathan P Day, and Francis M Jiggins (2021). "*Wolbachia* reduces virus infection in a natural population of *Drosophila*". In: *Communications Biology* 4.(1), p. 1327.

Colla, Sheila R, Michael C Otterstatter, Robert J Gegear, and James D Thomson (2006). "Plight of the bumble bee: pathogen spillover from commercial to wild populations". In: *Biological conservation* 129.(4), pp. 461–467.

Colomer-Winter, Cristina et al. (2018). "Manganese acquisition is essential for virulence of *Enterococcus faecalis*". In: *PLoS pathogens* 14.(9), e1007102.

Cooper, W Rodney et al. (2017). "Bacterial endosymbionts of the psyllid *Cacopsylla pyricola* (Hemiptera: Psyllidae) in the Pacific northwestern United States". In: *Environmental Entomology* 46.(2), pp. 393–402.

Corbin, Brian D et al. (2008). "Metal chelation and inhibition of bacterial growth in tissue abscesses". In: *Science* 319.(5865), pp. 962–965.

Corby-Harris, Vanessa, Lucy A Snyder, Melissa R Schwan, Patrick Maes, Quinn S McFrederick, and Kirk E Anderson (2014). "Origin and effect of Alpha 2.2 Acetobacteraceae in honey bee larvae and description of *Parasaccharibacter apium* gen. nov., sp. nov". In: *Applied and environmental microbiology* 80.(24), pp. 7460–7472.

Cordes, Sara (2019). "Even bees know zero is less than one". In: *Learning & Behavior* 47, pp. 187–188.

Cornet, Luc et al. (2022). "Phylogenomic analyses of *Snodgrassella* isolates from honeybees and bumblebees reveal taxonomic and functional diversity". In: *Msystems* 7.(3), e01500–21.

Cotter, Sheena C, Stephen J Simpson, David Raubenheimer, and Kenneth Wilson (2011). "Macronutrient balance mediates trade-offs between immune function and life history traits". In: *Functional Ecology* 25.(1), pp. 186–198.

Cremer, Sylvia, Sophie AO Armitage, and Paul Schmid-Hempel (2007). "Social immunity". In: *Current biology* 17.(16), R693–R702.

Cremer, Sylvia, Christopher D Pull, and Matthias A Fürst (2018). "Social immunity: emergence and evolution of colony-level disease protection". In: *Annual Review of Entomology* 63, pp. 105–123.

Crozier, Ross H, Philip S Newey, Ellen A Schluens, Simon KA Robson, et al. (2010). "A masterpiece of evolution–*Oecophylla* weaver ants (Hymenoptera: Formicidae)". In: *Myrmecological News* 13.(5), pp. 57–71.

Cui, Peng et al. (2010). "A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing". In: *Genomics* 96.(5), pp. 259–265.

da Silva, Jack (2021). "Life history and the transitions to eusociality in the Hymenoptera". In: *Frontiers in Ecology and Evolution* 9.

Dar, Mudasir A, Kiran D Pawar, Jyoti P Jadhav, and Radhakrishna S Pandit (2015). "Isolation of cellulolytic bacteria from the gastro-intestinal tract of *Achatina fulica* (Gastropoda: Pulmonata) and their evaluation for cellulose biodegradation". In: *International Biodeterioration & Biodegradation* 98, pp. 73–80.

Darwin Tree of Life Project Consortium (2022). "Sequence locally, think globally: The Darwin tree of life project". In: *Proceedings of the National Academy of Sciences* 119.(4), e2115642118.

Daughenbaugh, Katie F et al. (2021). "Metatranscriptome analysis of sympatric bee species identifies bee virus variants and a new virus, *Andrena*-associated bee virus-1". In: *Viruses* 13.(2), p. 291.

Davis, James J et al. (2020). "The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities". In: *Nucleic acids research* 48.(D1), pp. D606–D612.

Davison, PJ and J Field (2016). "Social polymorphism in the sweat bee *Lasioglossum (Evylaeus) calceatum*". In: *Insectes Sociaux* 63, pp. 327–338.

De Bie, Tijl, Nello Cristianini, Jeffery P Demuth, and Matthew W Hahn (2006). "CAFE: a computational tool for the study of gene family evolution". In: *Bioinformatics* 22.(10), pp. 1269–1271.

Denison, Rachel and Valérie Raymond-Delpech (2008). "Insights into the molecular basis of social behaviour from studies on the honeybee, *Apis mellifera*". In: *Invertebrate Neuroscience* 8, pp. 1–9.

Dew, Rebecca M, Quinn S McFrederick, and Sandra M Rehan (2020). "Diverse diets with consistent core microbiome in wild bee pollen provisions". In: *Insects* 11.(8), p. 499.

Dietrich, Carsten, Tim Köhler, and Andreas Brune (2014). "The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events". In: *Applied and Environmental Microbiology* 80.(7), pp. 2261–2269.

Dixon, Philip (2003). "VEGAN, a package of R functions for community ecology". In: *Journal of vegetation science* 14.(6), pp. 927–930.

Dobin, Alexander and Thomas R Gingeras (2016). "Optimizing RNA-Seq mapping with STAR". In: *Data mining techniques for the life sciences*. Springer, pp. 245–262.

Dobin, Alexander et al. (2013). "STAR: ultrafast universal RNA-seq aligner". In: *Bioinformatics* 29.(1), pp. 15–21.

Dolezal, Adam G and Amy L Toth (2014). "Honey bee sociogenomics: a genome-scale perspective on bee social behavior and health". In: *Apidologie* 45, pp. 375–395.

— (2018). "Feedbacks between nutrition and disease in honey bee health". In: *Current opinion in insect science* 26, pp. 114–119.

Doublet, Vincent et al. (2017). "Unity in defence: honeybee workers exhibit conserved molecular responses to diverse pathogens". In: *BMC genomics* 18.(1), pp. 1–17.

Dressler, Robert L (1982). "Biology of the orchid bees (Euglossini)". In: *Annual review of ecology and systematics* 13.(1), pp. 373–394.

Drum, Nathan H and Walter C Rothenbuhler (1985). "Differences in non-stinging aggressive responses of worker honeybees to diseased and healthy bees in May and July". In: *Journal of Apicultural Research* 24.(3), pp. 184–187.

Dudzic, Jan Paul, Mark Austin Hanson, Igor Iatsenko, Shu Kondo, and Bruno Lemaitre (2019). "More than black or white: melanization and *toll* share regulatory serine proteases in *Drosophila*". In: *Cell reports* 27.(4), pp. 1050–1061.

Dunn, Olive Jean (1964). "Multiple comparisons using rank sums". In: *Technometrics* 6.(3), pp. 241–252.

Duplouy, Anne, Christelle Couchoux, Ilkka Hanski, and Saskya van Nouhuys (2015). "*Wolbachia* infection in a natural parasitoid wasp population". In: *PloS one* 10.(8), e0134843.

Duret, Laurent and Nicolas Galtier (2009). "Biased gene conversion and the evolution of mammalian genomic landscapes". In: *Annual review of genomics and human genetics* 10, pp. 285–311.

Dyrhage, Karl et al. (2022). "Genome Evolution of a Symbiont Population for Pathogen Defense in Honeybees". In: *Genome Biology and Evolution* 14.(11), evac153.

Edosa, Tariku Tesfaye et al. (2020). "Tm Spz6 Is Essential for Regulating the Immune Response to *Escherichia coli* and *Staphylococcus aureus* Infection in *Tenebrio molitor*". In: *Insects* 11.(2), p. 105.

Edwards, Edwards, Haag Haag, Collins Collins, Hutson Hutson, and Huang Huang (1998). "*Lactobacillus kunkeei* sp. nov.: a spoilage organism associated with grape juice fermentations". In: *Journal of Applied Microbiology* 84.(5), pp. 698–702.

Ellegaard, Kirsten M, Daniel Tamarit, Emelie Javelind, Tobias C Olofsson, Siv GE Andersson, and Alejandra Vásquez (2015). "Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut". In: *BMC genomics* 16.(1), pp. 1–22.

Elsik, Christine G et al. (2016). "Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine". In: *Nucleic acids research* 44.(D1), pp. D793–D800.

Emms, David M and Steven Kelly (2019). "OrthoFinder: phylogenetic orthology inference for comparative genomics". In: *Genome biology* 20.(1), pp. 1–14.

Endo, Akihito and Seppo Salminen (2013). "Honeybees and beehives are rich sources for fructophilic lactic acid bacteria". In: *Systematic and Applied Microbiology* 36.(6), pp. 444–448.

Endo, Akihito et al. (2012). "Characterization and emended description of *Lactobacillus kunkeei* as a fructophilic lactic acid bacterium". In: *International journal of systematic and evolutionary microbiology* 62.(Pt_3), pp. 500–504.

Engel, Michael S and Claus Rasmussen (2021). "Corbiculate bees". In: *Encyclopedia of Social Insects. Springer, Berlin*.

Engel, Philipp, Waldan K Kwong, and Nancy A Moran (2013). "*Frischella perrara* gen. nov., sp. nov., a gammaproteobacterium isolated from the gut of the honeybee, *Apis mellifera*". In: *International journal of systematic and evolutionary microbiology* 63.(Pt_10), pp. 3646–3651.

Engel, Philipp, Vincent G Martinson, and Nancy A Moran (2012). "Functional diversity within the simple gut microbiota of the honey bee". In: *Proceedings of the National Academy of Sciences* 109.(27), pp. 11002–11007.

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

Engel, Philipp, Ramunas Stepanauskas, and Nancy A Moran (2014). "Hidden diversity in honey bee gut symbionts detected by single-cell genomics". In: *PLoS genetics* 10.(9), e1004596.

Engel, Philipp et al. (2016). "The bee microbiome: impact on bee health and model for evolution and ecology of host-microbe interactions". In: *MBio* 7.(2), e02164–15.

Erwin, Douglas H (2000). "Macroevolution is more than repeated rounds of microevolution". In: *Evolution & development* 2.(2), pp. 78–84.

Evans, Jay D et al. (2006). "Immune pathways and defence mechanisms in honey bees *Apis mellifera*". In: *Insect molecular biology* 15.(5), pp. 645–656.

Fan, Yuanchan et al. (2022). "Comparative Transcriptome Investigation of *Nosema ceranae* Infecting Eastern Honey Bee Workers". In: *Insects* 13.(3), p. 241.

Federhen, Scott (2012). "The NCBI taxonomy database". In: *Nucleic acids research* 40.(D1), pp. D136–D143.

Ferreira, Álvaro Gil, Huw Naylor, Sara Santana Esteves, Inês Silva Pais, Nelson Eduardo Martins, and Luis Teixeira (2014). "The *Toll-dorsal* pathway is required for resistance to viral oral infection in Drosophila". In: *PLoS pathogens* 10.(12), e1004507.

Fischer, O and P Schmid-Hempel (2005). "Selection by parasites may increase host recombination frequency". In: *Biology Letters* 1.(2), pp. 193–195.

Forbes, Andrew A, Robin K Bagley, Marc A Beer, Alaine C Hippee, and Heather A Widmayer (2018). "Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order". In: *BMC ecology* 18.(1), pp. 1–11.

Forsgren, Eva, Tobias C Olofsson, Alejandra Vásquez, and Ingemar Fries (2010). "Novel lactic acid bacteria inhibiting *Paenibacillus larvae* in honey bee larvae". In: *Apidologie* 41.(1), pp. 99–108.

Fünfhaus, Anne, Julia Ebeling, and Elke Genersch (2018). "Bacterial pathogens of bees". In: *Current opinion in insect science* 26, pp. 89–96.

Gadagkar, Raghavendra (2011). "The birth of ant genomics". In: *Proceedings of the National Academy of Sciences* 108.(14), pp. 5477–5478.

Gadau, Jürgen et al. (2012). "The genomic impact of 100 million years of social evolution in seven ant species". In: *Trends in Genetics* 28.(1), pp. 14–21.

Galbraith, David A, Xingyu Yang, Elina Lastro Nino, Soojin Yi, and Christina Grozinger (2015). "Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*)". In: *PLoS pathogens* 11.(3), e1004713.

Gallai, Nicola, Jean-Michel Salles, Josef Settele, and Bernard E Vaissière (2009). "Economic valuation of the vulnerability of world agriculture confronted with pollinator decline". In: *Ecological economics* 68.(3), pp. 810–821.

Galtier, Nicolas and Laurent Duret (2007). "Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution". In: *TRENDS in Genetics* 23.(6), pp. 273–277.

Gan, Tian et al. (2021). "JNK Signaling in *Drosophila* aging and longevity". In: *International Journal of Molecular Sciences* 22.(17), p. 9649.

Garcia, JR et al. (2014). "Partner associations across sympatric broad-headed bug species and their environmentally acquired bacterial symbionts". In: *Molecular ecology* 23.(6), pp. 1333–1347.

Garibaldi, Lucas A et al. (2011). "Stability of pollination services decreases with isolation from natural areas despite honey bee visits". In: *Ecology letters* 14.(10), pp. 1062–1072.

Geer, Lewis Y et al. (2010). "The NCBI biosystems database". In: *Nucleic acids research* 38.(suppl_1), pp. D492–D496.

Gerth, Michael, Annemarie GEißler, and Christoph Bleidorn (2011). "Wolbachia infections in bees (Anthophila) and possible implications for DNA barcoding". In: *Systematics and Biodiversity* 9.(4), pp. 319–327.

Gerth, Michael, Abiya Saeed, Jennifer A White, and Christoph Bleidorn (2015). "Extensive screen for bacterial endosymbionts reveals taxon-specific distribution patterns among bees (Hymenoptera, Anthophila)". In: *FEMS Microbiology Ecology* 91.(6).

Gibbs, Jason, Seán G Brady, Kojun Kanda, and Bryan N Danforth (2012). "Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae)". In: *Molecular phylogenetics and evolution* 65.(3), pp. 926–939.

Goka, Koichi, Kimiko Okabe, Masahiro Yoneda, and Satomi Niwa (2001). "Bumblebee commercialization will cause worldwide migration of parasitic mites". In: *Molecular ecology* 10.(8), pp. 2095–2099.

Goulson, Dave, Gillian C Lye, and Ben Darvill (2008). "Decline and conservation of bumble bees". In: *Annual Review of Entomology* 53, pp. 191–208.

Goulson, Dave, Elizabeth Nicholls, Cristina Botıas, and Ellen L Rotheray (2015). "Bee declines driven by combined stress from parasites, pesticides, and lack of flowers". In: *Science* 347.(6229), p. 1255957.

Graur, Dan (1985). "Gene diversity in Hymenoptera". In: *Evolution* 39.(1), pp. 190–199.

Grozinger, Christina M and Michelle L Flenniken (2019). "Bee viruses: Ecology, pathogenicity, and impacts". In: *Annual review of entomology* 64.(1).

Gupta, Shishir K et al. (2015). "Scrutinizing the immune defence inventory of *Camponotus floridanus* applying total transcriptome sequencing". In: *BMC genomics* 16.(1), pp. 1–21.

Hammer, Tobin J, Eli Le, Alexia N Martin, and Nancy A Moran (2021). "The gut microbiota of bumblebees". In: *Insectes sociaux*, pp. 1–15.

Handy, Madeline Y, Dino L Sbardellati, Michael Yu, Nicholas W Saleh, Madeleine M Ostwald, and Rachel L Vannette (2022). "Incipiently social carpenter bees (*Xylocopa*) host distinctive gut bacterial communities and display geographical structure as revealed by full-length PacBio 16S rRNA sequencing". In: *Molecular Ecology*.

Harpur, Brock A and Sandra M Rehan (2021). "Connecting social polymorphism to single nucleotide polymorphism: population genomics of the small carpenter bee, *Ceratina australensis*". In: *Biological Journal of the Linnean Society* 132.(4), pp. 945–954.

Harpur, Brock A and Amro Zayed (2013). "Accelerated evolution of innate immunity proteins in social insects: adaptive evolution or relaxed constraint?" In: *Molecular biology and evolution* 30.(7), pp. 1665–1674.

Harpur, Brock A et al. (2019). "Integrative genomics reveals the genetics and evolution of the honey bee's social immune system". In: *Genome Biology and Evolution* 11.(3), pp. 937–948.

Harrison, Mark C et al. (2021). "Gene Coexpression Network Reveals Highly Conserved, Well-Regulated Anti-Ageing Mechanisms in Old Ant Queens". In: *Genome biology and evolution* 13.(6), evab093.

Hartfield, Matthew and Peter D Keightley (2012). "Current hypotheses for the evolution of sex and recombination". In: *Integrative zoology* 7.(2), pp. 192–209.

Harwood, Gyan, Gro Amdam, and Dalial Freitak (2019). "The role of Vitellogenin in the transfer of immune elicitors from gut to hypopharyngeal glands in honey bees (*Apis mellifera*)". In: *Journal of insect physiology* 112, pp. 90–100.

Havukainen, Heli, Daniel Münch, Anne Baumann, Shi Zhong, Øyvind Halskau, Michelle Krogsgaard, et al. (2013). "Vitellogenin recognizes cell damage through membrane binding and shields living cells from reactive oxygen species". In: *Journal of Biological Chemistry* 288.(39), pp. 28369–28381.

Hayward, Alexander, Tokiharu Takahashi, William G Bendena, Stephen S Tobe, and Jerome HL Hui (2010). "Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans". In: *FEBS letters* 584.(6), pp. 1273–1278.

He, Shulin et al. (2021). "Evidence for reduced immune gene diversity and activity during the evolution of termites". In: *Proceedings of the Royal Society B* 288.(1945), p. 20203168.

Hennig, W. (1981). *Insect Phylogeny*. Wiley-Blackwell.

Heo, Jun, Soo-Jin Kim, Jeong-Seon Kim, Seung-Beom Hong, and Soon-Wo Kwon (2020). "Comparative genomics of *Lactobacillus* species as bee symbionts and description of *Lactobacillus bombintestini* sp. nov., isolated from the gut of *Bombus ignitus*". In: *Journal of Microbiology* 58, pp. 445–455.

Herold, Heike M and Thomas Scheibel (2017). "Applicability of biotechnologically produced insect silks". In: *Zeitschrift für Naturforschung C* 72.(9-10), pp. 365–385.

Hess, George (1996). "Disease in metapopulation models: implications for conservation". In: *Ecology* 77.(5), pp. 1617–1632.

Hill, William G and Alan Robertson (1966). "The effect of linkage on limits to artificial selection". In: *Genetics Research* 8.(3), pp. 269–294.

Hillyer, Julián F (2016). "Insect immunology and hematopoiesis". In: *Developmental & Comparative Immunology* 58, pp. 102–118.

Hoggard, Stephen J, Peter D Wilson, Andrew J Beattie, and Adam J Stow (2011). "Social complexity and nesting habits are factors in the evolution of antimicrobial defences in wasps". In: *PLoS One* 6.(7), e21763.

Holm, Sture (1979). "A simple sequentially rejective multiple test procedure". In: *Scandinavian journal of statistics*, pp. 65–70.

Honeybee Genome Sequencing Consortium and others (2006). "Insights into social insects from the genome of the honeybee Apis mellifera". In: *Nature* 443.(7114), p. 931.

Hood, M Indriati and Eric P Skaar (2012). "Nutritional immunity: transition metals at the pathogen–host interface". In: *Nature Reviews Microbiology* 10.(8), pp. 525–537.

Hope, Adery CA (1968). "A simplified Monte Carlo significance test procedure". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 30.(3), pp. 582–598.

Horak, Richard D, Sean P Leonard, and Nancy A Moran (2020). "Symbionts shape host innate immunity in honeybees". In: *Proceedings of the Royal Society B* 287.(1933), p. 20201184.

Howard, Scarlett R, Aurore Avarguès-Weber, Jair E Garcia, Andrew D Greentree, and Adrian G Dyer (2018). "Numerical ordering of zero in honey bees". In: *Science* 360.(6393), pp. 1124–1126.

Hrdina, Alexandra and Igor Iatsenko (2021). "The roles of metals in insect-microbe interactions and immunity". In: *Current Opinion in Insect Science*.

Hu, Guowen, Lihua Zhang, Yueli Yun, and Yu Peng (2019). "Taking insight into the gut microbiota of three spider species: No characteristic symbiont was found

corresponding to the special feeding style of spiders". In: *Ecology and Evolution* 9.(14), pp. 8146–8156.

Huang, Shengwei, Ping Sheng, and Hongyu Zhang (2012). "Isolation and identification of cellulolytic bacteria from the gut of *Holotrichia parallela* larvae (Coleoptera: Scarabaeidae)". In: *International journal of molecular sciences* 13.(3), pp. 2563–2577.

Huerta-Cepas, Jaime et al. (2016). "eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences". In: *Nucleic acids research* 44.(D1), pp. D286–D293.

Huerta-Cepas, Jaime et al. (2019). "eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses". In: *Nucleic acids research* 47.(D1), pp. D309–D314.

Husemann, Martin, Marco Thomas Neiber, Jana Nickel, Carla VM Reinbold, Michael Kuhlmann, and Mathilde Cordellier (2021). "Mitogenomic phylogeny of bee families confirms the basal position and monophyly of Melittidae". In: *Zoologica Scripta* 50.(3), pp. 352–357.

Iatsenko, Igor, Alice Marra, Jean-Philippe Boquete, Jasquelin Peña, and Bruno Lemaitre (2020). "Iron sequestration by transferrin 1 mediates nutritional immunity in *Drosophila melanogaster*". In: *Proceedings of the National Academy of Sciences* 117.(13), pp. 7317–7325.

Iorizzo, Massimo et al. (2020). "Antagonistic activity against *Ascosphaera apis* and functional properties of *Lactobacillus kunkeei* strains". In: *Antibiotics* 9.(5), p. 262.

Ishizawa, Hidehiro, Masashi Kuroda, Kanako Inoue, Daisuke Inoue, Masaaki Morikawa, and Michihiko Ike (2019). "Colonization and competition dynamics of plant growth-promoting/inhibiting bacteria in the phytosphere of the duckweed *Lemna minor*". In: *Microbial ecology* 77, pp. 440–450.

Jha, Prabhat N et al. (2020). "Alterations in the endophyte-enriched root-associated microbiome of rice receiving growth-promoting treatments of urea fertilizer and Rhizobium biofertilizer". In: *Microbial ecology* 79, pp. 367–382.

Jo, A Reum and Jee-Young Imm (2017). "Effects of aronia extract on lifespan and age-related oxidative stress in *Drosophila melanogaster*". In: *Food science and biotechnology* 26.(5), pp. 1399–1406.

Jones, Beryl M, Callum J Kingwell, William T Wcislo, and Gene E Robinson (2017). "Caste-biased gene expression in a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of eusociality". In: *Proceedings of the Royal Society B: Biological Sciences* 284.(1846), p. 20162228.

Jones, Julia C, Andreas Wallberg, Matthew J Christmas, Karen M Kapheim, and Matthew T Webster (2019). "Extreme differences in recombination rate between the genomes of a solitary and a social bee". In: *Molecular Biology and Evolution* 36.(10), pp. 2277–2291.

Jones, Philip et al. (2014). "InterProScan 5: genome-scale protein function classification". In: *Bioinformatics* 30.(9), pp. 1236–1240.

Jørgensen, Frank Grønlund, Mikkel Heide Schierup, and Andrew G Clark (2007). "Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of *Apis mellifera*". In: *Molecular biology and evolution* 24.(2), pp. 611–619.

Kačániová, Miroslava, Margarita Terentjeva, Jana Žiarovská, and Przemysław Łukasz Kowalczewski (2020). "In vitro antagonistic effect of gut bacteriota isolated from indigenous honey bees and essential oils against *Paenibacillus larvae*". In: *International Journal of Molecular Sciences* 21.(18), p. 6736.

Kalantar, Katrina L et al. (2020). "IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring". In: *Gigascience* 9.(10), giaa111.

Kanamori, Yasushi et al. (2010). "The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involved in trehalose import into peripheral tissues". In: *Insect biochemistry and molecular biology* 40.(1), pp. 30–37.

Kapheim, Karen M, Makenna M Johnson, and Maggi Jolley (2021). "Composition and acquisition of the microbiome in solitary, ground-nesting alkali bees". In: *Scientific reports* 11.(1), p. 2993.

Kapheim, Karen M et al. (2015). "Genomic signatures of evolutionary transitions from solitary to group living". In: *Science* 348.(6239), pp. 1139–1143.

Kapheim, Karen M et al. (2019). "Draft genome assembly and population genetics of an agricultural pollinator, the solitary alkali bee (Halictidae: *Nomia melanderi*)". In: *G3: Genes, Genomes, Genetics* 9.(3), pp. 625–634.

Kapheim, Karen M et al. (2020). "Developmental plasticity shapes social traits and selection in a facultatively eusocial bee". In: *Proceedings of the National Academy of Sciences* 117.(24), pp. 13615–13625.

Katoh, Kazutaka, John Rozewicki, and Kazunori D Yamada (2019). "MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization". In: *Briefings in bioinformatics* 20.(4), pp. 1160–1166.

Katz, Kenneth, Oleg Shutov, Richard Lapoint, Michael Kimelman, J Rodney Brister, and Christopher O'Sullivan (2022). "The Sequence Read Archive: a decade more of explosive growth". In: *Nucleic acids research* 50.(D1), pp. D387–D390.

Kawakami, Takeshi et al. (2019). "Substantial heritable variation in recombination rate on multiple scales in honeybees and bumblebees". In: *Genetics* 212.(4), pp. 1101–1119.

Kehl-Fie, Thomas E et al. (2011). "Nutrient metal sequestration by calprotectin inhibits bacterial superoxide defense, enhancing neutrophil killing of *Staphylococcus aureus*". In: *Cell host & microbe* 10.(2), pp. 158–164.

Keightley, Peter D and Sarah P Otto (2006). "Interference among deleterious mutations favours sex and recombination in finite populations". In: *Nature* 443.(7107), pp. 89–92.

Keller, Alexander, Gudrun Grimmer, and Ingolf Steffan-Dewenter (2013). "Diverse microbiota identified in whole intact nest chambers of the red mason bee *Osmia bicornis* (Linnaeus 1758)". In: *PLoS One* 8.(10), e78296.

Keller, Alexander et al. (2018). "Wild bees and their nests host *Paenibacillus* bacteria with functional potential of avail". In: *Microbiome* 6.(1), pp. 1–10.

Keller, L (1998). "Queen lifespan and colony characteristics in ants and termites". In: *Insectes sociaux* 45, pp. 235–246.

Kent, Clement F, Amer Issa, Alexandra C Bunting, and Amro Zayed (2011). "Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*". In: *Molecular Ecology* 20.(24), pp. 5226–5235.

Kent, Clement F, Shermineh Minaei, Brock A Harpur, and Amro Zayed (2012). "Recombination is associated with the evolution of genome structure and worker behavior in honey bees". In: *Proceedings of the National Academy of Sciences* 109.(44), pp. 18012–18017.

Kent, Clement F and Amro Zayed (2013). "Evolution of recombination and genome structure in eusocial insects". In: *Communicative & Integrative Biology* 6.(2), pp. 18012–7.

Kent, Tyler V, Jasmina Uzunović, and Stephen I Wright (2017). "Coevolution between transposable elements and recombination". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.(1736), p. 20160458.

Kerstes, Niels AG, Camillo Bérénos, Paul Schmid-Hempel, and K Mathias Wegner (2012). "Antagonistic experimental coevolution with a parasite increases host recombination frequency". In: *BMC Evolutionary Biology* 12, pp. 1–9.

Kešnerová, Lucie, Ruben AT Mars, Kirsten M Ellegaard, Michaël Troilo, Uwe Sauer, and Philipp Engel (2017). "Disentangling metabolic functions of bacteria in the honey bee gut". In: *PLoS biology* 15.(12), e2003467.

Kešnerová, Lucie, Roxane Moritz, and Philipp Engel (2016). "Bartonella apis sp. nov., a honey bee gut symbiont of the class Alphaproteobacteria." In: *International journal of systematic and evolutionary microbiology* 66.(1), pp. 414–421.

Killer, J, Soňa Dubná, Ivo Sedláček, and Pavel Švec (2014a). "*Lactobacillus apis* sp. nov., from the stomach of honeybees (*Apis mellifera*), having an in vitro inhibitory effect on the causative agents of American and European foulbrood".

In: *International journal of systematic and evolutionary microbiology* 64.(Pt_1), pp. 152–157.

Killer, J, A Votavová, I Valterová, E Vlková, V Rada, and Z Hroncová (2014b). "*Lactobacillus bombi* sp. nov., from the digestive tract of laboratory-reared bumblebee queens (*Bombus terrestris*)". In: *International journal of systematic and evolutionary microbiology* 64.(Pt_8), pp. 2611–2617.

Killer, J et al. (2009). "*Bifidobacterium bombi* sp. nov., from the bumblebee digestive tract". In: *International journal of systematic and evolutionary microbiology* 59.(8), pp. 2020–2024.

Killer, J et al. (2011). "*Bifidobacterium actinocoloniiforme* sp. nov. and *Bifidobacterium bohemicum* sp. nov., from the bumblebee digestive tract". In: *International Journal of Systematic and Evolutionary Microbiology* 61.(6), pp. 1315–1321.

Kiran, Fadime, Sedat Sevin, and Ahmet Ceylan (2022). "Biocontrol potential of *Apilactobacillus kunkeei* EIR/BG-1 against infectious diseases in honey bees (Apis mellifera L.)" In: *Veterinary Research Communications*, pp. 1–13.

Klein, Alexandra-Maria et al. (2007). "Importance of pollinators in changing landscapes for world crops". In: *Proceedings of the royal society B: biological sciences* 274.(1608), pp. 303–313.

Koch, Hauke, Dharam P Abrol, Jilian Li, and Paul Schmid-Hempel (2013). "Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees". In: *Molecular ecology* 22.(7), pp. 2028–2044.

Koch, Hauke and Paul Schmid-Hempel (2011a). "Bacterial communities in central European bumblebees: low diversity and high specificity". In: *Microbial Ecology* 62.(1), pp. 121–133.

— (2011b). "Socially transmitted gut microbiota protect bumble bees against an intestinal parasite". In: *Proceedings of the National Academy of Sciences* 108.(48), pp. 19288–19292.

Koch, Hauke and Paul Schmid-Hempel (2012). "Gut microbiota instead of host genotype drive the specificity in the interaction of a natural host-parasite system". In: *Ecology letters* 15.(10), pp. 1095–1103.

Kocher, Sarah D and Robert J Paxton (2014). "Comparative methods offer powerful insights into social evolution in bees". In: *Apidologie* 45, pp. 289–305.

Kocher, Sarah D et al. (2013). "The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*". In: *Genome Biology* 14, pp. 1–14.

Kodama, Yuichi, Martin Shumway, and Rasko Leinonen (2012). "The Sequence Read Archive: explosive growth of sequencing data". In: *Nucleic acids research* 40.(D1), pp. D54–D56.

Kohno, Hiroki and Takeo Kubo (2019). "Genetics in the honey bee: Achievements and prospects toward the functional analysis of molecular and neural mechanisms underlying social behaviors". In: *Insects* 10.(10), p. 348.

Kosior, Andrzej et al. (2007). "The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe". In: *Oryx* 41.(1), pp. 79–88.

Kraaijeveld, AR, J Ferrari, and HCJ Godfray (2002). "Costs of resistance in insect-parasite and insect-parasitoid interactions". In: *Parasitology* 125.(7), S71–S82.

Krampis, Konstantinos and Claudia Wultsch (2015). "A review of cloud computing bioinformatics solutions for next-gen sequencing data analysis and research". In: *Methods in Next Generation Sequencing* 2.(1).

Krause, Christopher D, Zhi-Hong Yang, Young-Sun Kim, Jin-Hyung Lee, Jeffry R Cook, and Sidney Pestka (2007). "Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential". In: *Pharmacology & therapeutics* 113.(1), pp. 50–87.

Kumar, Sudhir et al. (2022). "TimeTree 5: An expanded resource for species divergence times". In: *Molecular Biology and Evolution* 39.(8), msac174.

Kwakman, Paulus HS, Anje A Te Velde, Leonie de Boer, Christina MJE Vandenbroucke-Grauls, and Sebastian AJ Zaat (2011). "Two major medicinal honeys have different mechanisms of bactericidal activity". In: *PloS one* 6.(3), e17709.

Kwong, Waldan K et al. (2017a). "Dynamic microbiome evolution in social bees". In: *Science Advances* 3.(3), e1600513.

Kwong, Waldan K, Amanda L Mancenido, and Nancy A Moran (2017b). "Immune system stimulation by the native gut microbiota of honey bees". In: *Royal Society open science* 4.(2), p. 170003.

— (2014). "Genome sequences of *Lactobacillus* sp. strains wkB8 and wkB10, members of the Firm-5 clade, from honey bee guts". In: *Genome announcements* 2.(6), e01176–14.

Kwong, Waldan K and Nancy A Moran (2013). "Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of *Snodgrassella alvi* gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and *Gilliamella apicola* gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order 'Enterobacteriales' of the Gammaproteobacteria". In: *International journal of systematic and evolutionary microbiology* 63.(Pt_6), pp. 2008–2018.

— (2015). "Evolution of host specialization in gut microbes: the bee gut as a model". In: *Gut microbes* 6.(3), pp. 214–220.

— (2016a). "*Apibacter adventoris* gen. nov., sp. nov., a member of the phylum Bacteroidetes isolated from honey bees". In: *International journal of systematic and evolutionary microbiology* 66.(3), p. 1323.

— (2016b). "Gut microbial communities of social bees". In: *Nature reviews microbiology* 14.(6), pp. 374–384.

Lang, Haoyu et al. (2022). "Specific strains of honeybee gut *lactobacillus* stimulate host immune system to protect against pathogenic *Hafnia alvei*". In: *Microbiology Spectrum* 10.(1), e01896–21.

Lapidge, Keryn L, Benjamin P Oldroyd, and Marla Spivak (2002). "Seven suggestive quantitative trait loci influence hygienic behavior of honey bees". In: *Naturwissenschaften* 89, pp. 565–568.

Larsen, Brendan B, Elizabeth C Miller, Matthew K Rhodes, and John J Wiens (2017). "Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life". In: *The Quarterly Review of Biology* 92.(3), pp. 229–265.

Lawley, Trevor D and Alan W Walker (2013). "Intestinal colonization resistance". In: *Immunology* 138.(1), pp. 1–11.

LeBuhn, Gretchen and Joshua Vargas Luna (2021). "Pollinator decline: what do we know about the drivers of solitary bee declines?" In: *Current opinion in insect science* 46, pp. 106–111.

Leinonen, Rasko, Hideaki Sugawara, Martin Shumway, and International Nucleotide Sequence Database Collaboration (2010). "The sequence read archive". In: *Nucleic acids research* 39.(suppl_1), pp. D19–D21.

Lester, Philip J et al. (2022). "Viral communities in the parasite Varroa destructor and in colonies of their honey bee host (Apis mellifera) in New Zealand". In: *Scientific Reports* 12.(1), pp. 1–13.

Letunic, Ivica and Peer Bork (2021). "Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation". In: *Nucleic acids research* 49.(W1), W293–W296.

Leulier, François et al. (2003). "The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition". In: *Nature immunology* 4.(5), pp. 478–484.

Levashina, Elena A, Luis F Moita, Stephanie Blandin, Gert Vriend, Marie Lagueux, and Fotis C Kafatos (2001). "Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*". In: *Cell* 104.(5), pp. 709–718.

Li, Leilei et al. (2015). "*Bombella intestini* gen. nov., sp. nov., an acetic acid bacterium isolated from bumble bee crop". In: *International journal of systematic and evolutionary microbiology* 65.(Pt_1), pp. 267–273.

Lim, Haw Chuan, Chia-Ching Chu, Manfredo J Seufferheld, and Sydney A Cameron (2015). "Deep sequencing and ecological characterization of gut microbial communities of diverse bumble bee species". In: *PloS one* 10.(3), e0118566.

Liu, Haoxuan et al. (2015). "Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee". In: *Genome biology* 16.(1), pp. 1–20.

Liu, Kun, Yuemei Dong, Yuzheng Huang, Jason L Rasgon, and Peter Agre (2013). "Impact of trehalose transporter knockdown on *Anopheles gambiae* stress adaptation and susceptibility to *Plasmodium falciparum* infection". In: *Proceedings of the National Academy of Sciences* 110.(43), pp. 17504–17509.

Liu, Yanjie et al. (2020). "*De novo* transcriptomic and metabolomic analyses reveal the ecological adaptation of high-altitude *Bombus pyrosoma*". In: *Insects* 11.(9), p. 631.

Lombardo, Michael P (2008). "Access to mutualistic endosymbiotic microbes: an underappreciated benefit of group living". In: *Behavioral Ecology and Sociobiology* 62, pp. 479–497.

López-Uribe, Margarita M, Warren B Sconiers, Steven D Frank, Robert R Dunn, and David R Tarpy (2016). "Reduced cellular immune response in social insect lineages". In: *Biology Letters* 12.(3), p. 20150984.

Lourenço, Anete P, Juliana R Martins, Márcia MG Bitondi, and Zilá LP Simões (2009). "Trade-off between immune stimulation and expression of storage protein genes". In: *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America* 71.(2), pp. 70–87.

Lourenço, Anete Pedro, Mônica Mazzei Florecki, Zilá Luz Paulino Simões, and Jay Daniel Evans (2018). "Silencing of *Apis mellifera dorsal* genes reveals their

role in expression of the antimicrobial peptide defensin-1". In: *Insect molecular biology* 27.(5), pp. 577–589.

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome biology* 15.(12), pp. 1–21.

Lu, Anrui et al. (2014). "Insect prophenoloxidase: the view beyond immunity". In: *Frontiers in physiology* 5, p. 252.

Lu, Dihong, Tiffany Baiocchi, and Adler R Dillman (2016). "Genomics of entomopathogenic nematodes and implications for pest control". In: *Trends in parasitology* 32.(8), pp. 588–598.

Lu, Huanhuan, Bo He, Youjin Hao, Zeyang Zhou, Chengyong Su, and Dunyuan Huang (2021). "Comparative Mitogenomic Analysis of Two Cuckoo Bees (Apoidea: Anthophila: Megachilidae) with Phylogenetic Implications". In: *Insects* 12.(1), p. 29.

Lucaci, Alexander G, Sadie R Wisotsky, Stephen D Shank, Steven Weaver, and Sergei L Kosakovsky Pond (2021). "Extra base hits: widespread empirical support for instantaneous multiple-nucleotide changes". In: *Plos one* 16.(3), e0248337.

Lucas, Eric R and Laurent Keller (2020). "The co-evolution of longevity and social life". In: *Functional Ecology* 34.(1), pp. 76–87.

Ludvigsen, Jane, Davide Porcellato, Gro V Amdam, and Knut Rudi (2018). "Addressing the diversity of the honeybee gut symbiont *Gilliamella*: description of *Gilliamella apis* sp. nov., isolated from the gut of honeybees (*Apis mellifera*)". In: *International Journal of Systematic and Evolutionary Microbiology* 68.(5), pp. 1762–1770.

Malik, Garima, Rahul Arora, Ritu Chaturvedi, and Manoj S Paul (2021). "Implementation of genetic engineering and novel omics approaches to enhance bioremediation: a focused review". In: *Bulletin of environmental contamination and toxicology*, pp. 1–8.

Mancini, MV et al. (2018). "Estimating bacteria diversity in different organs of nine species of mosquito by next generation sequencing". In: *BMC microbiology* 18.(1), pp. 1–10.

Manley, Robyn, Mike Boots, and Lena Wilfert (2015). "Emerging viral disease risk to pollinating insects: ecological, evolutionary and anthropogenic factors". In: *Journal of Applied Ecology* 52.(2), pp. 331–340.

Martin, Callum D, Michelle T Fountain, and Mark JF Brown (2021). "The potential for parasite spill-back from commercial bumblebee colonies: a neglected threat to wild bees?" In: *Journal of Insect Conservation* 25.(3), pp. 531–539.

Martin, Marcel (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *EMBnet. journal* 17.(1), pp. 10–12.

Martinson, Vincent G, Bryan N Danforth, Robert L Minckley, Olav Rueppell, Salim Tingek, and Nancy A Moran (2011). "A simple and distinctive microbiota associated with honey bees and bumble bees". In: *Molecular ecology* 20.(3), pp. 619–628.

Martinson, Vincent G, Jamie Moy, and Nancy A Moran (2012). "Establishment of characteristic gut bacteria during development of the honeybee worker". In: *Applied and environmental microbiology* 78.(8), pp. 2830–2840.

Mattarelli, Paola, Giovanna Felis, Charles MAP Franz, and Michael G Gaenzle (2021). "Proposal to reclassify four *Lactobacillus* species as *Apilactobacillus bombintestini*, *Companilactobacillus suantsaicola*, *Lactiplantibacillus garii* and *Levilactobacillus suantsaiihabitans*". In: *International Journal of Systematic and Evolutionary Microbiology* 71.(3), p. 004704.

McFrederick, Quinn S and Sandra M Rehan (2016). "Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee". In: *Molecular ecology* 25.(10), pp. 2302–2311.

— (2019). "Wild bee pollen usage and microbial communities co-vary across landscapes". In: *Microbial Ecology* 77, pp. 513–522.

McFrederick, Quinn S, William T Wcislo, Michael C Hout, and Ulrich G Mueller (2014). "Host species and developmental stage, but not host social structure, affects bacterial community structure in socially polymorphic bees". In: *FEMS Microbiology Ecology* 88.(2), pp. 398–406.

McFrederick, Quinn S, William T Wcislo, Douglas R Taylor, Heather D Ishak, Scot E Dowd, and Ulrich G Mueller (2012). "Environment or kin: whence do bees obtain acidophilic bacteria?" In: *Molecular Ecology* 21.(7), pp. 1754–1768.

McFrederick, Quinn S et al. (2017). "Flowers and wild megachilid bees share microbes". In: *Microbial ecology* 73, pp. 188–200.

McLaughlin, Gabrielle, Morgan Gueuning, David Genoud, Juerg E Frey, and Christophe Praz (2023). "Why are there so many species of mining bees (Hymenoptera, Andrenidae)? The possible roles of phenology and *Wolbachia* incompatibility in maintaining species boundaries in the *Andrena* proxima-complex". In: *Systematic Entomology* 48.(1), pp. 127–141.

McMenamin, Alexander J and Michelle L Flenniken (2018). "Recently identified bee viruses and their impact on bee pollinators". In: *Current opinion in insect science* 26, pp. 120–129.

Melicher, Dacotah, Elisabeth S Wilson, Julia H Bowsher, Steve S Peterson, George D Yocum, and Joseph P Rinehart (2019). "Long-distance transportation causes temperature stress in the honey bee, *Apis mellifera* (Hymenoptera: Apidae)". In: *Environmental entomology* 48.(3), pp. 691–701.

Mendes, Fábio K, Dan Vanderpool, Ben Fulton, and Matthew W Hahn (2021). "CAFE 5 models variation in evolutionary rates among gene families". In: *Bioinformatics* 36.(22-23), pp. 5516–5518.

Meusemann, Karen, Judith Korb, Maximilian Schughart, and Fabian Staubach (2020). "No evidence for single-copy immune-gene specific signals of selection in termites". In: *Frontiers in Ecology and Evolution* 8, p. 26.

Michel, Tatiana, Jean-Marc Reichhart, Jules A Hoffmann, and Julien Royet (2001). "*Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein". In: *Nature* 414.(6865), pp. 756–759.

Michener, Charles D (1969). "Comparative social behavior of bees". In: *Annual review of entomology* 14.(1), pp. 299–342.

Miller, Delaney L, Eric A Smith, and Irene LG Newton (2021). "A bacterial symbiont protects honey bees from fungal disease". In: *MBio* 12.(3), e00503–21.

Milutinović, Barbara, Robert Peuß, Kevin Ferro, and Joachim Kurtz (2016). "Immune priming in arthropods: an update focusing on the red flour beetle". In: *Zoology* 119.(4), pp. 254–261.

Mockler, Blair K, Waldan K Kwong, Nancy A Moran, and Hauke Koch (2018). "Microbiome structure influences infection by the parasite *Crithidia bombi* in bumble bees". In: *Applied and environmental microbiology* 84.(7), e02335–17.

Moran, Nancy A, Allison K Hansen, J Elijah Powell, and Zakee L Sabree (2012). "Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees". In: *PloS one* 7.(4), e36393.

Moret, Yannick and Jérôme Moreau (2012). "The immune role of the arthropod exoskeleton". In: *Invertebrate Survival Journal* 9.(2), pp. 200–206.

Morgan, Xochitl C, Nicola Segata, and Curtis Huttenhower (2013). "Biodiversity and functional genomics in the human microbiome". In: *Trends in genetics* 29.(1), pp. 51–58.

Moussian, Bernard (2010). "Recent advances in understanding mechanisms of insect cuticle differentiation". In: *Insect biochemistry and molecular biology* 40.(5), pp. 363–375.

Mueller, Ulrich G, Nicole M Gerardo, Duur K Aanen, Diana L Six, and Ted R Schultz (2005). "The evolution of agriculture in insects". In: *Annu. Rev. Ecol. Evol. Syst.* 36, pp. 563–595.

Munoz-Benavent, Maria, Ana Elena Perez-Cobas, Carlos Garcia-Ferris, Andres Moya, and Amparo Latorre (2021). "Insects' potential: understanding the functional

role of their gut microbiome". In: *Journal of Pharmaceutical and Biomedical Analysis* 194, p. 113787.

Murray, Elizabeth A, John Burand, Natalia Trikoz, Julia Schnabel, Heather Grab, and Bryan N Danforth (2019). "Viral transmission in honey bees and native bees, supported by a global black queen cell virus phylogeny". In: *Environmental microbiology* 21.(3), pp. 972–983.

Murray, Tomás E, Mary F Coffey, Eamonn Kehoe, and Finbarr G Horgan (2013). "Pathogen prevalence in commercially reared bumble bees and evidence of spillover in conspecific populations". In: *Biological Conservation* 159, pp. 269–276.

Näpflin, Kathrin and Paul Schmid-Hempel (2016). "Immune response and gut microbial community structure in bumblebees after microbiota transplants". In: *Proceedings of the Royal Society B: Biological Sciences* 283.(1831), p. 20160312.

Nappi, AJ and BM Christensen (2005). "Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity". In: *Insect biochemistry and molecular biology* 35.(5), pp. 443–459.

Naug, Dhruba and Scott Camazine (2002). "The role of colony organization on pathogen transmission in social insects". In: *Journal of theoretical Biology* 215.(4), pp. 427–439.

NCBI Resource Coordinators (2018). "Database resources of the national center for biotechnology information". In: *Nucleic acids research* 46.(D1), pp. D8–D13.

Negri, Pedro et al. (2019). "Towards precision nutrition: a novel concept linking phytochemicals, immune response and honey bee health". In: *Insects* 10.(11), p. 401.

Nelson, Annika S and Kailen A Mooney (2022). "The evolution and ecology of interactions between ants and honeydew-producing hemipteran insects". In: *Annual Review of Ecology, Evolution, and Systematics* 53, pp. 379–402.

Nelson, C Mindy, Kate E Ihle, M Kim Fondrk, Robert E Page Jr, and Gro V Amdam (2007). "The gene vitellogenin has multiple coordinating effects on social organization". In: *PLoS biology* 5.(3), e62.

Nouri, Shahideh, Emilyn E Matsumura, Yen-Wen Kuo, and Bryce W Falk (2018). "Insect-specific viruses: from discovery to potential translational applications". In: *Current opinion in virology* 33, pp. 33–41.

Nowbahari, Elise, Alexandra Scohier, Jean-Luc Durand, and Karen L Hollis (2009). "Ants, *Cataglyphis cursor*, use precisely directed rescue behavior to free entrapped relatives". In: *PloS one* 4.(8), e6573.

Obbard, Darren J, Karl HJ Gordon, Amy H Buck, and Francis M Jiggins (2009). "The evolution of RNAi as a defence against viruses and transposable elements". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.(1513), pp. 99–115.

Obbard, Darren J, Francis M Jiggins, Daniel L Halligan, and Tom J Little (2006). "Natural selection drives extremely rapid evolution in antiviral RNAi genes". In: *Current biology* 16.(6), pp. 580–585.

Oliphant, Scott A, Nathan S Watson-Haigh, Krista M Sumby, Jennifer Gardner, Scott Groom, and Vladimir Jiranek (2022). "*Apilactobacillus apisilvae* sp. nov., *Nicolia spurrieriana* gen. nov. sp. nov., *Bombilactobacillus folatiphilus* sp. nov. and *Bombilactobacillus thymidiniphilus* sp. nov., four new lactic acid bacterial isolates from stingless bees *Tetragonula carbonaria* and *Austroplebeia australis*". In: *International Journal of Systematic and Evolutionary Microbiology* 72.(9), p. 005588.

Olofsson, Tobias C, Magnus Alsterfjord, Bo Nilson, Èile Butler, and Alejandra Vásquez (2014). "*Lactobacillus apinorum* sp. nov., *Lactobacillus mellifer* sp. nov., *Lactobacillus mellis* sp. nov., *Lactobacillus melliventris* sp. nov., *Lactobacillus kimbladii* sp. nov., *Lactobacillus helsingborgensis* sp. nov. and *Lactobacillus kullabergensis* sp. nov., isolated from the honey stomach of the honeybee *Apis mellifera*". In: *International journal of systematic and evolutionary microbiology* 64.(Pt 9), p. 3109.

Ong, Sek Tong, Jason Zhe Shan Ho, Bow Ho, and Jeak Ling Ding (2006). "Iron-withholding strategy in innate immunity". In: *Immunobiology* 211.(4), pp. 295–314.

Ono, Masato, Takeshi Igarashi, Eishi Ohno, and Masami Sasaki (1995). "Unusual thermal defence by a honeybee against mass attack by hornets". In: *Nature* 377, pp. 334–336.

Orgad, Sara, Hannah Nelson, Daniel Segal, and Nathan Nelson (1998). "Metal ions suppress the abnormal taste behavior of the *Drosophila* mutant *malvolio*." In: *The Journal of experimental biology* 201.(1), pp. 115–120.

Osterman, Julia et al. (2021). "Global trends in the number and diversity of managed pollinator species". In: *Agriculture, Ecosystems & Environment* 322, p. 107653.

Otani, Saria, Nick Bos, and Sze H Yek (2016). "Transitional complexity of social insect immunity". In: *Frontiers in Ecology and Evolution* 4, p. 69.

Oxley, Peter R, Marla Spivak, and Benjamin P Oldroyd (2010). "Six quantitative trait loci influence task thresholds for hygienic behaviour in honeybees (*Apis mellifera*)". In: *Molecular Ecology* 19.(7), pp. 1452–1461.

Park, Doori et al. (2015). "Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing". In: *BMC genomics* 16.(1), pp. 1–16.

Park, Hee Geun et al. (2018). "Honeybee (*Apis cerana*) vitellogenin acts as an antimicrobial and antioxidant agent in the body and venom". In: *Developmental & Comparative Immunology* 85, pp. 51–60.

Park, Soyi et al. (2019). "TmToll-7 plays a crucial role in innate immune responses against Gram-negative bacteria by regulating 5 AMP genes in *Tenebrio molitor*". In: *Frontiers in immunology* 10, p. 310.

Park, Ji-Won et al. (2007). "Clustering of peptidoglycan recognition protein-SA is required for sensing lysine-type peptidoglycan in insects". In: *Proceedings of the National Academy of Sciences* 104.(16), pp. 6602–6607.

Pascall, David J, Matthew C Tinsley, Bethany L Clark, Darren J Obbard, and Lena Wilfert (2021). "Virus prevalence and genetic diversity across a wild bumblebee community". In: *Frontiers in microbiology* 12, p. 650747.

Patalano, Solenn et al. (2015). "Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies". In: *Proceedings of the National Academy of Sciences* 112.(45), pp. 13970–13975.

Pennacchio, Francesco and Michael R Strand (2006). "Evolution of developmental strategies in parasitic Hymenoptera". In: *Annu. Rev. Entomol.* 51, pp. 233–258.

Pérez-Sato, Juan-Antonio, Nicolas Châline, Stephen J Martin, WOH Hughes, and Francis LW Ratnieks (2009). "Multi-level selection for hygienic behaviour in honeybees". In: *Heredity* 102.(6), pp. 609–615.

Peters, Ralph S et al. (2017). "Evolutionary history of the Hymenoptera". In: *Current Biology* 27.(7), pp. 1013–1018.

Pie, Marcio R, Rebeca B Rosengaus, and James FA Traniello (2004). "Nest architecture, activity pattern, worker density and the dynamics of disease transmission in social insects". In: *Journal of Theoretical Biology* 226.(1), pp. 45–51.

Pimentel, Andre C, Cassia S Cesar, Marcos Martins, and Rodrigo Cogni (2021). "The antiviral effects of the symbiont bacteria *Wolbachia* in insects". In: *Frontiers in immunology* 11, p. 626329.

Porcheron, Gaëlle, Amélie Garénaux, Julie Proulx, Mourad Sabri, and Charles M Dozois (2013). "Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence". In: *Frontiers in cellular and infection microbiology* 3, p. 90.

Potts, Simon G, Jacobus C Biesmeijer, Claire Kremen, Peter Neumann, Oliver Schweiger, and William E Kunin (2010). "Global pollinator declines: trends, impacts and drivers". In: *Trends in ecology & evolution* 25.(6), pp. 345–353.

Powell, J Elijah, Sean P Leonard, Waldan K Kwong, Philipp Engel, and Nancy A Moran (2016). "Genome-wide screen identifies host colonization determinants in

a bacterial gut symbiont". In: *Proceedings of the National Academy of Sciences* 113.(48), pp. 13887–13892.

Praet, Jessy, Ivan Meeus, Margo Cnockaert, Maarten Aerts, Guy Smagghe, and Peter Vandamme (2015a). "*Bifidobacterium commune* sp. nov. isolated from the bumble bee gut". In: *Antonie Van Leeuwenhoek* 107, pp. 1307–1313.

Praet, Jessy, Ivan Meeus, Margo Cnockaert, Kurt Houf, Guy Smagghe, and Peter Vandamme (2015b). "Novel lactic acid bacteria isolated from the bumble bee gut: *Convivina intestini* gen. nov., sp. nov., *Lactobacillus bombicola* sp. nov., and *Weissella bombi* sp. nov." In: *Antonie Van Leeuwenhoek* 107, pp. 1337–1349.

Praet, Jessy, Maarten Aerts, Evie De Brandt, Ivan Meeus, Guy Smagghe, and Peter Vandamme (2016). "*Apibacter mensalis* sp. nov.: a rare member of the bumblebee gut microbiota". In: *International Journal of Systematic and Evolutionary Microbiology* 66.(4), pp. 1645–1651.

Praet, Jessy, Margo Cnockaert, Ivan Meeus, Guy Smagghe, and Peter Vandamme (2017). "*Gilliamella intestini* sp. nov., *Gilliamella bombicola* sp. nov., *Gilliamella bombi* sp. nov. and *Gilliamella mensalis* sp. nov.: Four novel *Gilliamella* species isolated from the bumblebee gut". In: *Systematic and Applied Microbiology* 40.(4), pp. 199–204.

Prigot-Maurice, Cybèle, Sophie Beltran-Bech, and Christine Braquart-Varnier (2022). "Why and how do protective symbionts impact immune priming with pathogens in invertebrates?" In: *Developmental & Comparative Immunology* 126, p. 104245.

Qiu, Bitao, Rasmus Stenbak Larsen, Ni-Chen Chang, John Wang, Jacobus J Boomsma, and Guojie Zhang (2018). "Towards reconstructing the ancestral brain gene-network regulating caste differentiation in ants". In: *Nature ecology & evolution* 2.(11), pp. 1782–1791.

Qiu, Ping, Pang Chu Pan, and Shubha Govind (1998). "A role for the *Drosophila* Toll/Cactus pathway in larval hematopoiesis". In: *Development* 125.(10), pp. 1909–1920.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rajkumar, Mani, Noriharu Ae, and Helena Freitas (2009). "Endophytic bacteria and their potential to enhance heavy metal phytoextraction". In: *Chemosphere* 77.(2), pp. 153–160.

Ramalho, Manuela de O, Zoe Kim, Steven Wang, and Corrie S Moreau (2021). "*Wolbachia* across social insects: patterns and Implications". In: *Annals of the Entomological Society of America* 114.(2), pp. 206–218.

Ramalho, Manuela de O, Alexsandro Santana Vieira, Mayara Cristina Pereira, Corrie S Moreau, and Odair Correa Bueno (2018). "Transovarian transmission of *Blochmannia* and *Wolbachia* endosymbionts in the neotropical weaver ant *Camponotus textor* (Hymenoptera, Formicidae)". In: *Current microbiology* 75, pp. 866–873.

Ramirez, Santiago R, David W Roubik, Charlotte Skov, and Naomi E Pierce (2010). "Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae)". In: *Biological Journal of the Linnean Society* 100.(3), pp. 552–572.

Ratzka, Carolin, Chunguang Liang, Thomas Dandekar, Roy Gross, and Heike Feldhaar (2011). "Immune response of the ant *Camponotus floridanus* against pathogens and its obligate mutualistic endosymbiont". In: *Insect biochemistry and molecular biology* 41.(8), pp. 529–536.

Raymann, Kasie and Nancy A Moran (2018). "The role of the gut microbiome in health and disease of adult honey bee workers". In: *Current opinion in insect science* 26, pp. 97–104.

Raymann, Kasie, Zack Shaffer, and Nancy A Moran (2017). "Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees". In: *PLoS biology* 15.(3), e2001861.

Rehan, Sandra M, Karl M Glastad, Sarah P Lawson, and Brendan G Hunt (2016). "The genome and methylome of a subsocial small carpenter bee, Ceratina calcarata". In: *Genome biology and evolution* 8.(5), pp. 1401–1410.

Rehan, Sandra M, Karl M Glastad, Michael A Steffen, Cameron R Fay, Brendan G Hunt, and Amy L Toth (2018). "Conserved genes underlie phenotypic plasticity in an incipiently social bee". In: *Genome biology and evolution* 10.(10), pp. 2749–2758.

Rehan, Sandra M and Amy L Toth (2015). "Climbing the social ladder: the molecular evolution of sociality". In: *Trends in ecology & evolution* 30.(7), pp. 426–433.

Reichhart, Jean Marc, David Gubb, and Vincent Leclerc (2011). "The *Drosophila* serpins: multiple functions in immunity and morphogenesis". In: *Methods in enzymology*. Vol. 499. Elsevier, pp. 205–225.

Remnant, Emily J et al. (2017). "A diverse range of novel RNA viruses in geographically distinct honey bee populations". In: *Journal of virology* 91.(16), e00158–17.

Richard, Freddie-Jeanne, Holly L Holt, and Christina M Grozinger (2012). "Effects of immunostimulation on social behavior, chemical communication and genome-wide gene expression in honey bee workers (*Apis mellifera*)". In: *Bmc Genomics* 13.(1), pp. 1–18.

Roberts, John MK, Denis L Anderson, and Peter A Durr (2017). "Absence of deformed wing virus and *Varroa destructor* in Australia provides unique perspectives on honeybee viral landscapes and colony losses". In: *Scientific reports* 7.(1), pp. 1–11.

Robinson, Gene E, Christina M Grozinger, and Charles W Whitfield (2005). "Sociogenomics: social life in molecular terms". In: *Nature Reviews Genetics* 6.(4), pp. 257–270.

Roditi, Isabel and Michael J Lehane (2008). "Interactions between trypanosomes and tsetse flies". In: *Current opinion in microbiology* 11.(4), pp. 345–351.

Rodríguez-García, Cristina et al. (2021). "Transferrin-mediated iron sequestration suggests a novel therapeutic strategy for controlling *Nosema* disease in the honey bee, *Apis mellifera*". In: *PLoS Pathogens* 17.(2), e1009270.

Romiguier, Jonathan et al. (2014). "Population genomics of eusocial insects: the costs of a vertebrate-like effective population size". In: *Journal of Evolutionary Biology* 27.(3), pp. 593–603.

Rose, Patrick P et al. (2011). "Natural resistance-associated macrophage protein is a cellular receptor for sindbis virus in both insect and mammalian hosts". In: *Cell host & microbe* 10.(2), pp. 97–104.

Rothenbuhler, Walter C (1964). "Behavior genetics of nest cleaning in honey bees. IV. Responses of F 1 and backcross generations to disease-killed brood". In: *American Zoologist* 4.(2), pp. 111–123.

Royet, Julien, Dipika Gupta, and Roman Dziarski (2011). "Peptidoglycan recognition proteins: modulators of the microbiome and inflammation". In: *Nature Reviews Immunology* 11.(12), pp. 837–851.

Rubin, Benjamin ER, Beryl M Jones, Brendan G Hunt, and Sarah D Kocher (2019). "Rate variation in the evolution of non-coding DNA associated with social evolution in bees". In: *Philosophical Transactions of the Royal Society B* 374.(1777), p. 20180247.

Russell, Jacob A (2012). "The ants (Hymenoptera: Formicidae) are unique and enigmatic hosts of prevalent *Wolbachia* (Alphaproteobacteria) symbionts". In: *Myrmecological News* 16.(January, 2011), pp. 7–23.

Sackton, Timothy B (2019). "Comparative genomics and transcriptomics of host–pathogen interactions in insects: evolutionary insights and future directions". In: *Current opinion in insect science* 31, pp. 106–113.

— (2020). "Studying natural selection in the era of ubiquitous genomes". In: *Trends in Genetics* 36.(10), pp. 792–803.

Sackton, Timothy B, Brian P Lazzaro, Todd A Schlenke, Jay D Evans, Dan Hultmark, and Andrew G Clark (2007). "Dynamic evolution of the innate immune system in *Drosophila*". In: *Nature genetics* 39.(12), pp. 1461–1468.

Sackton, Timothy B, John H Werren, and Andrew G Clark (2013). "Characterizing the infection-induced transcriptome of *Nasonia vitripennis* reveals a preponderance of taxonomically-restricted immune genes". In: *PloS one* 8.(12), e83984.

Sadd, Ben M and Seth M Barribeau (2013). "Heterogeneity in infection outcome: lessons from a bumblebee-trypanosome system". In: *Parasite Immunology* 35.(11), pp. 339–349.

Sadd, Ben M, Yvonne Kleinlogel, Regula Schmid-Hempel, and Paul Schmid-Hempel (2005). "Trans-generational immune priming in a social insect". In: *Biology letters* 1.(4), pp. 386–388.

Sadd, Ben M and Paul Schmid-Hempel (2006). "Insect immunity shows specificity in protection upon secondary pathogen exposure". In: *Current Biology* 16.(12), pp. 1206–1210.

Sadd, Ben M et al. (2015). "The genomes of two key bumblebee species with primitive eusocial organization". In: *Genome biology* 16.(1), pp. 1–32.

Saeed, Abiya and Jennifer A White (2015). "Surveys for maternally-inherited endosymbionts reveal novel and variable infections within solitary bee species". In: *Journal of invertebrate pathology* 132, pp. 111–114.

Saleh, Nicholas W, Jonas Henske, and Santiago R Ramırez (2022). "Experimental disruption of social structure reveals totipotency in the orchid bee, *Euglossa dilemma*". In: *Evolution* 76.(7), pp. 1529–1545.

Salmela, Heli, Gro V Amdam, and Dalial Freitak (2015). "Transfer of immunity from mother to offspring is mediated via egg-yolk protein vitellogenin". In: *PLoS pathogens* 11.(7), e1005015.

Salmela, Heli and Liselotte Sundström (2018). "Vitellogenin in inflammation and immunity in social insects". In: *Inflammation and Cell Signaling* 5.

Sanders, Jon G, Scott Powell, Daniel JC Kronauer, Heraldo L Vasconcelos, Megan E Frederickson, and Naomi E Pierce (2014). "Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes". In: *Molecular Ecology* 23.(6), pp. 1268–1283.

Sane, Sanjay P, Sree Subha Ramaswamy, and Sreekrishna Varma Raja (2020). "Insect architecture: structural diversity and behavioral principles". In: *Current Opinion in Insect Science* 42, pp. 39–46.

Sayers, Eric W et al. (2021). "Database resources of the national center for biotechnology information". In: *Nucleic acids research* 49.(D1), p. D10.

Sayers, Eric W et al. (2023). "Database resources of the National Center for Biotechnology Information in 2023". In: *Nucleic acids research* 51.(D1), pp. D29–D38.

Scardovi, V and LD Trovatelli (1969). "New species of bifid bacteria from *Apis mellifica* L. and *Apis indica* F. A contribution to the taxonomy and biochemistry of the genus *Bifidobacterium*." In: *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Zweite Naturwissenschaftliche Abt.: Allgemeine, Landwirtschaftliche und Technische Mikrobiologie* 123.(1), pp. 64–88.

Scherfer, Christoph, Huaping Tang, Zakaria Kambris, Nouara Lhocine, Carl Hashimoto, and Bruno Lemaitre (2008). "*Drosophila* Serpin-28D regulates hemolymph phenoloxidase activity and adult pigmentation". In: *Developmental biology* 323.(2), pp. 189–196.

Schmid-Hempel, P (1995). "Parasites and social insects". In: *Apidologie* 26.(3), pp. 255–271.

Schmid-Hempel, Paul (1998). *Parasites in social insects*. Princeton University Press.

— (2005). "Evolutionary ecology of insect immune defenses". In: *Annual review of entomology* 50, p. 529.

Schmid-Hempel, Paul and Regula Schmid-Hempel (1993). "Transmission of a pathogen in Bombus terrestris, with a note on division of labour in social insects". In: *Behavioral Ecology and Sociobiology* 33, pp. 319–327.

Schmid-Hempel, Regula et al. (2014). "The invasion of southern S outh A merica by imported bumblebees and associated parasites". In: *Journal of Animal Ecology* 83.(4), pp. 823–837.

Schoch, Conrad L et al. (2020). "NCBI Taxonomy: a comprehensive update on curation, resources and tools". In: *Database* 2020.

Schoonvaere, Karel, Lina De Smet, Guy Smagghe, Andy Vierstraete, Bart P Braeckman, and Dirk C de Graaf (2016). "Unbiased RNA shotgun metagenomics in social and solitary wild bees detects associations with eukaryote parasites and new viruses". In: *PLoS One* 11.(12), e0168456.

Schoonvaere, Karel, Guy Smagghe, Frédéric Francis, and Dirk C de Graaf (2018). "Study of the metatranscriptome of eight social and solitary wild bee species reveals novel viruses and bee parasites". In: *Frontiers in microbiology* 9, p. 177.

Schwarz, Michael P, Miriam H Richards, and Bryan N Danforth (2007). "Changing paradigms in insect social evolution: insights from halictine and allodapine bees". In: *Annu. Rev. Entomol.* 52, pp. 127–150.

Schwarz, Ryan S, Qiang Huang, and Jay D Evans (2015). "Hologenome theory and the honey bee pathosphere". In: *Current opinion in insect science* 10, pp. 1–7.

Seehuus, Siri-Christine, Kari Norberg, Ulrike Gimsa, Trygve Krekling, and Gro V Amdam (2006). "Reproductive protein protects functionally sterile honey bee workers from oxidative stress". In: *Proceedings of the National Academy of Sciences* 103.(4), pp. 962–967.

Séguret, Alice et al. (2021). "Transcriptomic signatures of ageing vary in solitary and social forms of an orchid bee". In: *Genome biology and evolution* 13.(6), evab075.

Shang, Yanfang, Peng Feng, and Chengshu Wang (2015). "Fungi that infect insects: altering host behavior and beyond". In: *PLoS pathogens* 11.(8), e1005037.

Sheehan, Gerard, Amy Garvey, Michael Croke, and Kevin Kavanagh (2018). "Innate humoral immune defences in mammals and insects: The same, with differences?" In: *Virulence* 9.(1), pp. 1625–1639.

Shell, Wyatt A and Sandra M Rehan (2022). "Comparative metagenomics reveals expanded insights into intra-and interspecific variation among wild bee microbiomes". In: *Communications biology* 5.(1), p. 603.

Shell, Wyatt A et al. (2021). "Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees". In: *Communications Biology* 4.(1), p. 253.

Shorter, JR and Olav Rueppell (2012). "A review on self-destructive defense behaviors in social insects". In: *Insectes sociaux* 59, pp. 1–10.

Shykoff, Jacqui A and Paul Schmid-Hempel (1991). "Parasites and the advantage of genetic variability within social insect colonies". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 243.(1306), pp. 55–58.

Silverman, Neal and Tom Maniatis (2001). "NF-κB signaling pathways in mammalian and insect innate immunity". In: *Genes & development* 15.(18), pp. 2321–2342.

Singh, Rohini and Timothy A Linksvayer (2020). "*Wolbachia*-infected ant colonies have increased reproductive investment and an accelerated life cycle". In: *Journal of Experimental Biology* 223.(9), jeb220079.

Siozios, Stefanos, Josephine Moran, Mary Chege, Gregory DD Hurst, and Juan C Paredes (2019). "Complete reference genome assembly for *Commensalibacter* sp. strain AMU001, an acetic acid bacterium Isolated from the gut of honey bees". In: *Microbiology resource announcements* 8.(1), e01459–18.

Sirviö, A et al. (2006). "High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant *Acromyrmex echinatior*". In: *Journal of evolutionary biology* 19.(5), pp. 1475–1485.

Sirviö, Anu, J Spencer Johnston, Tom Wenseleers, and Pekka Pamilo (2011). "A high recombination rate in eusocial Hymenoptera: evidence from the common wasp *Vespula vulgaris*". In: *BMC genetics* 12.(1), pp. 1–7.

Smith, Eric A, Sylvie A Martin-Eberhardt, Delaney L Miller, Audrey J Parish, and Irene LG Newton (2019). "Draft genome sequence of a *Bombella apis* strain

isolated from honey bees". In: *Microbiology Resource Announcements* 8.(47), e01329–19.

Smith, Eric A et al. (2021). "Reclassification of seven honey bee symbiont strains as *Bombella apis*". In: *International journal of systematic and evolutionary microbiology* 71.(9).

Sohn, Yu Jung et al. (2021). "Chemoautotroph *Cupriavidus necator* as a potential game-changer for global warming and plastic waste problem: A review". In: *Bioresource Technology* 340, p. 125693.

Soneson, Charlotte, Michael I Love, and Mark D Robinson (2015). "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences". In: *F1000Research* 4.

Soro, A, J Field, C Bridge, SC Cardinal, and RJ Paxton (2010). "Genetic differentiation across the social transition in a socially polymorphic sweat bee, *Halictus rubicundus*". In: *Molecular Ecology* 19.(16), pp. 3351–3363.

Spivak, Marla and Gary S Reuter (2001). "*Varroa destructor* infestation in untreated honey bee (Hymenoptera: Apidae) colonies selected for hygienic behavior". In: *Journal of Economic Entomology* 94.(2), pp. 326–331.

Standage, Daniel S, Ali J Berens, Karl M Glastad, Andrew J Severin, Volker P Brendel, and Amy L Toth (2016). "Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect". In: *Molecular ecology* 25.(8), pp. 1769–1784.

Steele, Margaret I, Waldan K Kwong, Marvin Whiteley, and Nancy A Moran (2017). "Diversification of type VI secretion system toxins reveals ancient antagonism among bee gut microbes". In: *MBio* 8.(6), e01630–17.

Stillson, Patrick T, David A Baltrus, and Alison Ravenscraft (2022). "Prevalence of an Insect-Associated Genomic Region in Environmentally Acquired Burkholderiaceae Symbionts". In: *Applied and Environmental Microbiology* 88.(9), e02502–21.

Stow, Adam et al. (2007). "Antimicrobial defences increase with sociality in bees". In: *Biology letters* 3.(4), pp. 422–424.

Stroeymeyt, Nathalie, Barbara Casillas-Pérez, and Sylvia Cremer (2014). "Organisational immunity in social insects". In: *Current Opinion in Insect Science* 5, pp. 1–15.

Suen, Garret et al. (2011). "The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle". In: *PLoS genetics* 7.(2), e1002007.

Sumner, Seirian, Emeline Favreau, Katie Geist, Amy Toth, and Sandra Rehan (2022). "Molecular Patterns and Processes in Evolving Sociality: Lessons from Insects". In: *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Sun, Cheng et al. (2021). "Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits". In: *Molecular biology and evolution* 38.(2), pp. 486–501.

Sun, Qian, Kenneth F Haynes, and Xuguo Zhou (2018). "Managing the risks and rewards of death in eusocial insects". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.(1754), p. 20170258.

Sutherland, Tara D, Sarah Weisman, Andrew A Walker, and Stephen T Mudie (2012). "The coiled coil silk of bees, ants, and hornets". In: *Biopolymers* 97.(6), pp. 446–454.

Sutherland, Tara D et al. (2006). "A highly divergent gene cluster in honey bees encodes a novel silk family". In: *Genome Research* 16.(11), pp. 1414–1421.

Suyama, Mikita, David Torrents, and Peer Bork (2006). "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments". In: *Nucleic acids research* 34.(suppl_2), W609–W612.

Szathmáry, Eörs and John Maynard Smith (1995). "The major evolutionary transitions". In: *Nature* 374.(6519), pp. 227–232.

Tamarit, Daniel, Kirsten M Ellegaard, Johan Wikander, Tobias Olofsson, Alejandra Vasquez, and Siv GE Andersson (2015). "Functionally structured genomes in *Lactobacillus kunkeei* colonizing the honey crop and food products of honeybees and stingless bees". In: *Genome Biology and Evolution* 7.(6), pp. 1455–1473.

Teets, Nicholas M and David L Denlinger (2014). "Surviving in a frozen desert: environmental stress physiology of terrestrial Antarctic arthropods". In: *Journal of Experimental Biology* 217.(1), pp. 84–93.

Thamm, Markus and Ricarda Scheiner (2014). "PKG in honey bees: spatial expression, *Amfor* gene expression, sucrose responsiveness, and division of labor". In: *Journal of Comparative Neurology* 522.(8), pp. 1786–1799.

Tibbetts, Elizabeth A (2002). "Visual signals of individual identity in the wasp *Polistes fuscatus*". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269.(1499), pp. 1423–1428.

Toth, Amy L and Sandra M Rehan (2017). "Molecular evolution of insect sociality: an eco-evo-devo perspective". In: *Annual Review of Entomology* 62, pp. 419–442.

Toth, Amy L and Gene E Robinson (2007). "Evo-devo and the evolution of social behavior". In: *Trends in Genetics* 23.(7), pp. 334–341.

Towb, Par, Andreas Bergmann, and Steven A Wasserman (Dec. 2001). "The protein kinase *Pelle* mediates feedback regulation in the *Drosophila* Toll signaling pathway". In: *Development* 128.(23), pp. 4729–4736.

Traniello, James FA, Rebeca B Rosengaus, and Keely Savoie (2002). "The development of immunity in a social insect: evidence for the group facilitation of disease resistance". In: *Proceedings of the National Academy of Sciences* 99.(10), pp. 6838–6842.

Trapp, Judith, Alison McAfee, and Leonard J Foster (2017). "Genomics, transcriptomics and proteomics: enabling insights into social evolution and disease challenges for managed and wild bees". In: *Molecular ecology* 26.(3), pp. 718–739.

Troha, Katia, Joo Hyun Im, Jonathan Revah, Brian P Lazzaro, and Nicolas Buchon (2018). "Comparative transcriptomics reveals *CrebA* as a novel regulator of infection tolerance in *D. melanogaster*". In: *PLoS Pathogens* 14.(2), e1006847.

Tsegaye, Bahiru, Chandrajit Balomajumder, and Partha Roy (2019). "Isolation and characterization of novel lignolytic, cellulolytic, and hemicellulolytic bacteria from

wood-feeding termite *Cryptotermes brevis*". In: *International Microbiology* 22, pp. 29–39.

Tsvetkov, Nadejda, Victoria J MacPhail, Sheila R Colla, and Amro Zayed (2021). "Conservation genomics reveals pesticide and pathogen exposure in the declining bumble bee *Bombus terricola*". In: *Molecular ecology* 30.(17), pp. 4220–4230.

Turnbull, Christine et al. (2011). "Antimicrobial strength increases with group size: implications for social evolution". In: *Biology Letters* 7.(2), pp. 249–252.

Turnbull, Christine et al. (2012). "Primordial enemies: fungal pathogens in thrips societies". In: *PLoS One* 7.(11), e49737.

Valanne, Susanna, Jing-Huan Wang, and Mika Rämet (2011). "The *Drosophila toll* signaling pathway". In: *The Journal of Immunology* 186.(2), pp. 649–656.

Vallet-Gely, Isabelle, Bruno Lemaitre, and Frédéric Boccard (2008). "Bacterial strategies to overcome insect defences". In: *Nature Reviews Microbiology* 6.(4), pp. 302–313.

Van Den Hurk, Andrew F et al. (2012). "Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*". In: *PLoS neglected tropical diseases* 6.(11), e1892.

Vásquez, Alejandra et al. (2012). "Symbionts as major modulators of insect health: lactic acid bacteria and honeybees". In: *PloS one* 7.(3), e33188.

Vásquez-Procopio, Johana, Subhash Rajpurohit, and Fanis Missirlis (2020). "Cuticle darkening correlates with increased body copper content in *Drosophila melanogaster*". In: *Biometals* 33.(6), pp. 293–303.

Venkat, Aarti, Matthew W Hahn, and Joseph W Thornton (2018). "Multinucleotide mutations cause false inferences of lineage-specific positive selection". In: *Nature ecology & evolution* 2.(8), pp. 1280–1288.

Vetvicka, V and P Sima (2017). "Various roles of β-glucan in invertebrates". In: *Invertebrate Survival Journal* 14.(1), pp. 488–493.

Viljakainen, Lumi, Jaana Jurvansuu, Ida Holmberg, Tobias Pamminger, Silvio Erler, and Sylvia Cremer (2018). "Social environment affects the transcriptomic response to bacteria in ant queens". In: *Ecology and evolution* 8.(22), pp. 11031–11070.

von Frisch, K (1946). "Die tänze der Bienen". In: *Österr. Zool. Z* 1.(1).

Voulgari-Kokota, Anna, Quinn S McFrederick, Ingolf Steffan-Dewenter, and Alexander Keller (2019). "Drivers, diversity, and functions of the solitary-bee microbiota". In: *Trends in Microbiology* 27.(12), pp. 1034–1044.

Waiker, Prashant, Fabiano Carlos Pinto de Abreu, Danielle Luna-Lucena, Flávia Cristina Paula Freitas, Zilá Luz Paulino Simões, and Olav Rueppell (2021). "Recombination mapping of the Brazilian stingless bee *Frieseomelitta varia* confirms high recombination rates in social hymenoptera". In: *BMC genomics* 22.(1), pp. 1–12.

Wallberg, Andreas, Sylvain Glémin, and Matthew T Webster (2015). "Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*". In: *PLoS genetics* 11.(4), e1005189.

Wallberg, Andreas et al. (2019). "A hybrid *de novo* genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds". In: *BMC genomics* 20, pp. 1–19.

Wang, Baohong, Mingfei Yao, Longxian Lv, Zongxin Ling, and Lanjuan Li (2017). "The human microbiota in health and disease". In: *Engineering* 3.(1), pp. 71–82.

Wang, Chengshu and Sibao Wang (2017). "Insect pathogenic fungi: genomics, molecular interactions, and genetic improvements". In: *Annual Review of Entomology* 62.(1), pp. 73–90.

Wang, Cong et al. (2018). "*Lactobacillus panisapium* sp. nov., from honeybee *Apis cerana* bee bread". In: *International Journal of Systematic and Evolutionary Microbiology* 68.(3), p. 703.

Wang, John et al. (2013). "A Y-like social chromosome causes alternative colony organization in fire ants". In: *Nature* 493.(7434), pp. 664–668.

Wang, Meng C, Dirk Bohmann, and Heinrich Jasper (2005). "JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling". In: *Cell* 121.(1), pp. 115–125.

Wang, Xiaofei et al. (2021). "High-fat diets with differential fatty acids induce obesity and perturb gut microbiota in honey bee". In: *International journal of molecular sciences* 22.(2), p. 834.

Wang, Zhiqiang, Kai Feng, Fang Tang, and Meng Xu (2021). "Activation of the Host Immune Response in *Hyphantria cunea* (Drury)(Lepidoptera: Noctuidae) Induced by *Serratia marcescens Bizio*". In: *Insects* 12.(11), p. 983.

Waterhouse, Robert M, Brian P Lazzaro, and Timothy B Sackton (2020). "Characterization of insect immune systems from genomic data". In: *Immunity in Insects*. Springer, pp. 3–34.

Wcislo, William T, Laura Arneson, Kari Roesch, Victor Gonzalez, Adam Smith, and Hermógenes Fernández (2004). "The evolution of nocturnal behaviour in sweat bees, *Megalopta genalis* and *M. ecuadoria* (Hymenoptera: Halictidae): an escape from competitors and enemies?" In: *Biological Journal of the Linnean Society* 83.(3), pp. 377–387.

Weber, Jacob J, Michael R Kanost, and Maureen J Gorman (2020). "Iron binding and release properties of transferrin-1 from *Drosophila melanogaster* and *Manduca sexta*: Implications for insect iron homeostasis". In: *Insect biochemistry and molecular biology* 125, p. 103438.

Webster, Matthew T and Laurence D Hurst (2012). "Direct and indirect consequences of meiotic recombination: implications for genome evolution". In: *Trends in genetics* 28.(3), pp. 101–109.

Weinert, Lucy A, Eli V Araujo-Jnr, Muhammad Z Ahmed, and John J Welch (2015). "The incidence of bacterial endosymbionts in terrestrial arthropods". In: *Proceedings of the Royal Society B: Biological Sciences* 282.(1807), p. 20150249.

Wenzel, M, I Schönig, M Berchtold, P Kämpfer, and H König (2002). "Aerobic and facultatively anaerobic cellulolytic bacteria from the gut of the termite *Zootermopsis angusticollis*". In: *Journal of applied microbiology* 92.(1), pp. 32–40.

Werren, John H, Laura Baldo, and Michael E Clark (2008). "*Wolbachia*: master manipulators of invertebrate biology". In: *Nature Reviews Microbiology* 6.(10), pp. 741–751.

Wheeler, William Morton (1911). "The ant-colony as an organism". In: *Journal of Morphology* 22.(2), pp. 307–325.

Wilfert, L, Juergen Gadau, and P Schmid-Hempel (2007). "Variation in genomic recombination rates among animal taxa and the case of social insects". In: *Heredity* 98.(4), pp. 189–197.

Wilson, Edward O and Bert Hölldobler (2005). "Eusociality: origin and consequences". In: *Proceedings of the National Academy of Sciences* 102.(38), pp. 13367–13371.

Wilson-Rich, Noah, Marla Spivak, Nina H Fefferman, and Philip T Starks (2009). "Genetic, individual, and group facilitation of disease resistance in insect societies". In: *Annual review of entomology* 54, pp. 405–423.

Wolter, Laura A, Shota Suenami, and Ryo Miyazaki (2021). "*Frischella japonica* sp. nov., an anaerobic member of the Orbales in the Gammaproteobacteria, isolated from the gut of the eastern honey bee, *Apis cerana japonica Fabricius*". In: *International Journal of Systematic and Evolutionary Microbiology* 71.(3).

Wu, Shitao, Sai Yin, and Bing Zhou (2022). "Molecular physiology of iron trafficking in Drosophila melanogaster". In: *Current Opinion in Insect Science*, p. 100888.

Wu, Yan-Yan et al. (2017). "Sublethal effects of imidacloprid on targeting muscle and ribosomal protein related genes in the honey bee Apis mellifera L." In: *Scientific reports* 7.(1), pp. 1–8.

Wurm, Yannick and Laurent Keller (2010). "Parasitoid wasps: from natural history to genomic studies". In: *Current Biology* 20.(5), R242–R244.

Xie, Wan-Ying, Jian-Qiang Su, and Yong-Guan Zhu (2015). "Phyllosphere bacterial community of floating macrophytes in paddy soil environments as revealed by Illu-

mina high-throughput sequencing". In: *Applied and Environmental Microbiology* 81.(2), pp. 522–532.

Yang, Ziheng et al. (1997). "PAML: a program package for phylogenetic analysis by maximum likelihood". In: *Computer applications in the biosciences* 13.(5), pp. 555–556.

Yang, Ziheng (2007). "PAML 4: phylogenetic analysis by maximum likelihood". In: *Molecular biology and evolution* 24.(8), pp. 1586–1591.

Yang, Ziheng and Rasmus Nielsen (2002). "Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages". In: *Mol. Biol. Evol.* 19.(6), pp. 908–917.

Yek, Sze H, Jacobus J Boomsma, and Morten Schiøtt (2013). "Differential gene expression in *Acromyrmex* leaf-cutting ants after challenges with two fungal pathogens". In: *Molecular ecology* 22.(8), pp. 2173–2187.

Yu, Bin, Qi Sang, Guoqing Pan, Chunfeng Li, and Zeyang Zhou (2020). "A *Toll-Spätzle* pathway in the immune response of *Bombyx mori*". In: *Insects* 11.(9), p. 586.

Yu, Yang et al. (2010). "Diversity of innate immune recognition mechanism for bacterial polymeric meso-diaminopimelic acid-type peptidoglycan in insects". In: *Journal of Biological Chemistry* 285.(43), pp. 32937–32945.

Yun, Ji-Hyun, June-Young Lee, Dong-Wook Hyun, Mi-Ja Jung, and Jin-Woo Bae (2017). "*Bombella apis* sp. nov., an acetic acid bacterium isolated from the midgut of a honey bee". In: *International journal of systematic and evolutionary microbiology* 67.(7), pp. 2184–2188.

Zdobnov, Evgeny M et al. (2017). "OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs". In: *Nucleic acids research* 45.(D1), pp. D744–D749.

Zendo, Takeshi et al. (2020). "Kunkecin A, a new nisin variant bacteriocin produced by the fructophilic lactic acid bacterium, *Apilactobacillus kunkeei* FF30-6 isolated from honey bees". In: *Frontiers in Microbiology* 11, p. 571903.

Zhang, Guangmei, Mazhar Hussain, and Sassan Asgari (2014). "Regulation of arginine methyltransferase 3 by a *Wolbachia*-induced microRNA in *Aedes aegypti* and its effect on *Wolbachia* and dengue virus replication". In: *Insect biochemistry and molecular biology* 53, pp. 81–88.

Zhang, Jianzhi, Rasmus Nielsen, and Ziheng Yang (2005). "Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level". In: *Molecular biology and evolution* 22.(12), pp. 2472–2479.

Zhang, Zi-Jing and Hao Zheng (2022). "Bumblebees with the socially transmitted microbiome: A novel model organism for gut microbiota research". In: *Insect Science* 29.(4), pp. 958–976.

Zhang, Shicui, Shaohui Wang, Hongyan Li, and Li Li (2011). "Vitellogenin, a multivalent sensor and an antimicrobial effector". In: *The international journal of biochemistry & cell biology* 43.(3), pp. 303–305.

Zhang, Shouke, Zikun Li, Jinping Shu, Huaijun Xue, Kai Guo, and Xudong Zhou (2022). "Soil-derived bacteria endow *Camellia* weevil with more ability to resist plant chemical defense". In: *Microbiome* 10.(1), p. 97.

Zhang, Shuning, Gina M Dailey, Elaine Kwan, Bernadette M Glasheen, Gyna E Sroga, and Andrea Page-McCaw (2006). "An MMP liberates the Ninjurin A ectodomain to signal a loss of cell adhesion". In: *Genes & development* 20.(14), pp. 1899–1910.

Zhang, Wei, Gianluca Tettamanti, Taha Bassal, Christa Heryanto, Ioannis Eleftherianos, and Amr Mohamed (2021). "Regulators and signalling in insect antimicrobial innate immunity: functional molecules and cellular pathways". In: *Cellular Signalling* 83, p. 110003.

Zheng, Hao, J Elijah Powell, Margaret I Steele, Carsten Dietrich, and Nancy A Moran (2017). "Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling". In: *Proceedings of the National Academy of Sciences* 114.(18), pp. 4775–4780.

Zhou, Cong-Zhao et al. (2000). “Fine organization of *Bombyx mori* fibroin heavy chain gene”. In: *Nucleic acids research* 28.(12), pp. 2413–2419.

Zhu, Yu-Xi, Zhang-Rong Song, Yue-Ling Song, Dian-Shu Zhao, and Xiao-Yue Hong (2020). “The microbiota in spider mite feces potentially reflects intestinal bacterial communities in the host”. In: *Insect science* 27.(5), pp. 859–868.

Zilber-Rosenberg, Ilana and Eugene Rosenberg (2008). “Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution”. In: *FEMS microbiology reviews* 32.(5), pp. 723–735.

Chapter 6

Supplementary Materials

6.1 Overview

Most of the tables for each of these chapters are available from the Supplementary Materials file uploaded alongside this thesis. These are large tables (i.e. results of selection analyses or differential expression analyses across many thousand genes). As an example, these tables are referred to as "**S2.1**", with the first number denoting the chapter of the thesis it refers to, and the second number referring to the order of the tables. Smaller tables and any supplementary figures are included in the following Appendices corresponding to the three data chapters.

6.1.1 Data Availability

The complete bioinformatic and analysis pipeline with accompanying scripts, trees and directions for each data chapter analysis are available as public github repositories.

Chapter 2: github.com/LMee17/Proj0_Analysis

Chapter 3: github.com/LMee17/Comp_Hym

Chapter 4: github.com/LMee17/AnthoMicroComp

6.2 Appendix A: Chapter 2 Supplementary Material

6.2.1 Tables

Table 6.1: Results from Kruskal-Wallis rank sum test and post hoc Dunn's test comparing gene functions and their mean dN/dS ratios. P-values adjusted using Bonferonni correction. $< 0.05 = *; < 0.01 = **; < 0.001 = ***; < 0.0001 = ****$

Kruskal-Wallis				
	χ^2	df	P	
	982.07	4	$< 2.2e-16$	
Dunn's Post-hoc Test				
Comparison	Z	P	P**.adj**	**Significance**
Background-Effector	0.4823365	6.30E-01	1.0000	
Background-Non-Canon	-4.8830629	1.04E-06	0.0000	****
Background-Receptor	-2.9747286	2.93E-03	0.0293	*
Background-Signalling	3.4614192	5.37E-04	0.0054	**
Effector-Non-Canon	-2.5268645	1.15E-02	0.1151	
Effector-Receptor	-2.5118179	1.20E-02	0.1201	
Effector-Signalling	1.0800784	2.80E-01	1.0000	
Non-Canon-Receptor	-0.7671345	4.43E-01	1.0000	
Non-Canon-Signalling	5.9545586	2.61E-09	0.0000	****
Receptor-Signalling	4.151095	3.31E-05	0.0003	***

Table 6.2: Results from Kruskal-Wallis rank sum test and post hoc Dunn's test comparing gene functions and their mean GC content per branch test categories. P-values adjusted using Bonferonni correction. $< 0.05 = *; < 0.01 = **; < 0.001 = ***; < 0.0001 = ****$

SOLITARY				
Kruskal-Wallis				
	χ^2	df	P	
	331.39	4	$< 2.2e-16$	
Dunn's Post-hoc Test				
Comparison	Z	P	P**.adj**	**Significance**
Background-Effector	-2.911004	3.60E-03	0.0360	*
Background-Non-Canon	-16.87972	6.34E-64	0.0000	****
Background-Receptor	-5.758384	8.49E-09	0.0000	****
Background-Signalling	2.35948	1.83E-02	0.1830	
Effector-Non-Canon	-4.57435	4.78E-06	0.0000	****
Effector-Receptor	-2.235623	2.54E-02	0.2538	
Effector-Signalling	3.664296	2.48E-04	0.0025	**
Non-Canon-Receptor	1.508227	1.31E-01	1.0000	
Non-Canon-Signalling	13.611666	3.41E-42	0.0000	****
Receptor-Signalling	6.243044	4.29E-10	0.0000	****
POST ORIGIN				
Kruskal-Wallis				
	χ^2	df	P	
	434.79	4	$< 2.2e-16$	
Dunn's Post-hoc Test				
Comparison	Z	P	P**.adj**	**Significance**
Background-Effector	-3.801089	1.44E-04	0.0014	**
Background-Non-Canon	-19.082652	3.52E-81	0.0000	****
Background-Receptor	-6.858491	6.96E-12	0.0000	****
Background-Signalling	3.148349	1.64E-03	0.0164	*
Effector-Non-Canon	-4.707724	2.50E-06	0.0000	****
Effector-Receptor	-2.434571	1.49E-02	0.1491	
Effector-Signalling	4.81424	1.48E-06	0.0000	****
Non-Canon-Receptor	1.384557	1.66E-01	1.0000	
Non-Canon-Signalling	15.737158	8.41E-56	0.0000	****
Receptor-Signalling	7.575353	3.58E-14	0.0000	****
POST ELABORATION				
Kruskal-Wallis				
	χ^2	df	P	
	260.79	4	$< 2.2e-16$	
Dunn's Post-hoc Test				
Comparison	Z	P	P**.adj**	**Significance**
Background-Effector	-2.528808	1.14E-02	0.1145	
Background-Non-Canon	-14.951046	1.53E-50	0.0000	****
Background-Receptor	-5.399999	6.66E-08	0.0000	****

Continued on next page

Table 6.2: continued

Background-Signalling	1.650724	9.88E-02	0.9879	
Effector-Non-Canon	-4.096745	4.19E-05	0.0004	***
Effector-Receptor	-2.233823	2.55E-02	0.2549	
Effector-Signalling	3.008459	2.63E-03	0.0263	*
Non-Canon-Receptor	1.060468	2.89E-01	1.0000	
Non-Canon-Signalling	11.737654	8.17E-32	0.0000	****
Receptor-Signalling	5.622469	1.88E-08	0.0000	****

Continued on next page

Table 6.3: Results of chi-squared tests of significant difference between numbers of positively-selected genes from the canonical immune gene class versus positively-selected genes from the background gene class per lineage. *P*-values adjusted using Benjamini-Hochburg procedure.

Branch(es) Tested	**Proportion under Positive Selection**		**X-Squared**	**p.value**	**adj.p.value**
	Canon	**Background**	**(df = NA)**		**(BH)**
All Post Elaboration	0	0.008	1.496	0.4	1
All Post Origin	0	0.003	0.595	0.666	1
All Solitary	0.011	0.001	10.561	0.035	0.385
Apis	0.005	0.015	1.194	0.371	1
Ceratina	0	0.009	1.596	0.272	1
Corbiculates	0	0.002	0.366	1	1
Dufourea	0.022	0.018	0.146	0.774	1
Habropoda	0.038	0.031	0.23	0.663	1
Lasioglossum	0.054	0.064	0.307	0.647	1
Megachile	0.011	0.009	0.058	1	1
Melipona	0.134	0.127	0.09	0.822	1

APPENDIX

Table 6.4: Results of chi-squared tests of significant difference between numbers of positively-selected genes from the non-canonical immune gene class versus positively-selected genes from the background gene class per lineage. *P*-values adjusted using Benjamini-Hochburg procedure.

Branch(es) Tested	**Proportion under Positive Selection**		**X-Squared (df = NA)**	**p.value**	**adj.p.value (BH)**
	Non-Canon	**Background**			
All Post Elaboration	0	0.008	1.11	0.432	1
All Post Origin	0	0.003	0.442	1	1
All Solitary	0	0.001	0.171	1	1
Apis	0.022	0.015	0.373	0.727	1
Ceratina	0.007	0.009	0.026	1	1
Corbiculates	0	0.002	0.269	1	1
Dufourea	0.007	0.018	0.862	0.521	1
Habropoda	0.007	0.031	2.628	0.126	1
Lasioglossum	0.08	0.064	0.565	0.483	1
Megachile	0	0.009	1.259	0.414	1
Melipona	0.13	0.127	0.015	1	1

APPENDIX

Table 6.5: Results of Chi-squared Tests for (1) equality of proportions without continuity correction and (2) trend in proportions between the three socialities for the three different gene classes.

	Canon	**Non-Canon**	**Background**
		Proportion PSG	
Solitary	0.05913978	0.01449275	0.05531915
Origin	0.05376344	0.07971014	0.07340426
Elaboration	0.13978495	0.15217391	0.14237589
χ **(df =2)**[1]	11.199	19.391	289.62
p**-value**	0.003701	1.69E-04	<2.2e-16
χ **(df = 1)**[2]	7.8413	17.351	260
p**-value**	0.005107	3.11E-05	<2.2e-16

6.3 Appendix B: Chapter 3 Supplementary Material

6.3.1 Figures

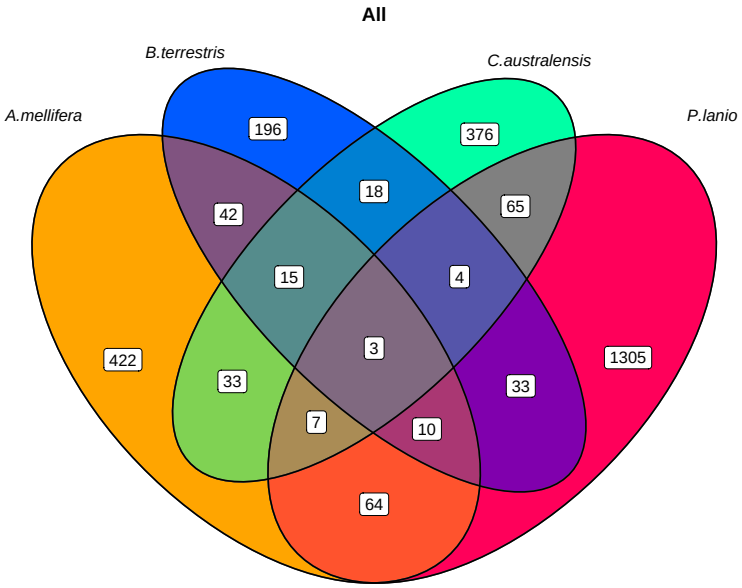

Figure 6.1: Venn diagram illustrating the overlap between significant differentially expressed genes across the four species. While some genes are shared, the majority of differentially expressed genes are species-specific.

6.3.2 Tables

Table 6.6: Data quality summary of the 48 libraries sequenced by Novogene. Raw reads: total amount of reads of raw data, each four lines taken as one unit. For paired-end sequencing, it equals the amount of read1 and read2, otherwise it equals the amount of read1 for single-end sequencing. Raw data: (Raw reads) * (sequence length), calculating in G. For paired-end sequencing like PE150, sequencing length equals 150, otherwise it equals 50 for sequencing like SE50. Effective: (Clean reads/Raw reads)*100%. Error: base error rate. Q20, Q30: (Base count of Phred value > 20 or 30) / (Total base count). GC: (G C base count) / (Total base count).

Sample	**Species**	**Treatment**	**Raw reads**	**Raw data (G)**	**Effective (%)**	**Error(%)**	**Q20(%)**	**Q30(%)**	**GC(%)**
AM_N_056	*A. mellifera*	Naïve	54482796	8.2	98.91	0.03	97.49	93.39	39.22
AM_N_068	*A. mellifera*	Naïve	47023956	7.1	98.94	0.02	98.12	94.59	38.85
AM_N_O72	*A. mellifera*	Naïve	58727372	8.8	98.68	0.02	98.08	94.46	38.76
AM_P_053	*A. mellifera*	Wound	41436842	6.2	97.71	0.03	97.44	93.49	38.7
AM_P_057	*A. mellifera*	Wound	40824400	6.1	97.52	0.02	97.91	94.31	38.43
AM_P_061	*A. mellifera*	Wound	49320690	7.4	97.05	0.03	96.99	92.64	38.74
AM_SL_050	*A. mellifera*	Gram +ve	44827680	6.7	97.49	0.02	97.91	94.31	38.23
AM_SL_054	*A. mellifera*	Gram +ve	44278362	6.6	98.04	0.03	97.37	93.39	39.29
AM_SL_058	*A. mellifera*	Gram +ve	49211636	7.4	97.69	0.03	97.7	93.86	37.89
AM_SM_051	*A. mellifera*	Gram -ve	45103132	6.8	97.79	0.03	97.66	93.94	38.29
AM_SM_059	*A. mellifera*	Gram -ve	47302312	7.1	97.58	0.03	97.46	93.34	37.95
AM_SM_075	*A. mellifera*	Gram -ve	61256040	9.2	97.95	0.03	97.51	93.61	38.46
BT_A_N_001	*B. terrestris*	Naïve	59387534	8.9	98.02	0.02	98.22	94.89	40.11
BT_A_N_025	*B. terrestris*	Naïve	71269776	10.7	97.19	0.02	98.02	94.37	40.87
BT_A_N_029	*B. terrestris*	Naïve	58462570	8.8	97.33	0.03	97.96	94.23	40.14
BT_A_P_014	*B. terrestris*	Wound	56547648	8.5	98.32	0.03	97.9	94.14	40.59
BT_A_P_026	*B. terrestris*	Wound	51543334	7.7	98.33	0.02	98.03	94.37	40.48
BT_A_P_030	*B. terrestris*	Wound	56037980	8.4	98.32	0.02	97.97	94.23	40.41

Continued on next page

APPENDIX

Table 6.6: continued

Sample	Species	Treatment	Raw reads	Raw data(G)	Effective(%)	Error(%)	Q20(%)	Q30(%)	GC(%)
BT_A_SL_003	*B. terrestris*	Gram +ve	63645146	9.5	98.65	0.02	98.1	94.49	40.06
BT_A_SL_027	*B. terrestris*	Gram +ve	57190288	8.6	98.18	0.03	97.8	93.91	40.31
BT_A_SL_031	*B. terrestris*	Gram +ve	56537792	8.5	98.24	0.03	97.79	93.88	40.78
BT_A_SM_004	*B. terrestris*	Gram -ve	49977678	7.5	98.48	0.02	98.08	94.47	40.63
BT_A_SM_028	*B. terrestris*	Gram -ve	44286492	6.6	98.31	0.02	98.04	94.43	40.84
BT_A_SM_032	*B. terrestris*	Gram -ve	48284592	7.2	98.12	0.02	98.2	94.83	40.39
CA_N_007	*C. australensis*	Naïve	44939930	6.7	99.23	0.03	97.93	94.06	42.23
CA_N_012	*C. australensis*	Naïve	42572374	6.4	99.04	0.03	97.89	94.03	43.75
CA_N_030	*C. australensis*	Naïve	44500540	6.7	98.65	0.03	97.9	94.02	42.59
CA_P_013	*C. australensis*	Wound	51977978	7.8	99.13	0.03	97.88	94.09	43.62
CA_P_041	*C. australensis*	Wound	39764440	6	99.14	0.03	97.86	94	43.36
CA_P_055	*C. australensis*	Wound	48569954	7.3	99.27	0.03	97.64	93.4	43.37
CA_SL_019	*C. australensis*	Gram +ve	42260116	6.3	99.21	0.03	97.96	94.15	42.51
CA_SL_048	*C. australensis*	Gram +ve	53289224	8	98.72	0.03	97.97	94.18	43.27
CA_SL_056_1	*C. australensis*	Gram +ve	51943162	7.8	98.98	0.03	97.71	93.66	44.74
CA_SM_029	*C. australensis*	Gram -ve	41554372	6.2	98.73	0.02	97.99	94.29	44.02

Continued on next page

Table 6.6: continued

Sample	**Species**	**Treatment**	**Raw reads**	**Raw data(G)**	**Effective(%)**	**Error(%)**	**Q20(%)**	**Q30(%)**	**GC(%)**
CA_SM_039	*C. australensis*	Gram -ve	48953682	7.3	99.31	0.02	97.97	94.23	44.36
CA_SM_049	*C. australensis*	Gram -ve	47715162	7.2	99.29	0.03	97.96	94.16	42.15
PL_N_1906	*P. lanio*	Naïve	45000402	6.8	97.88	0.03	97.89	94.1	36.51
PL_N_1908	*P. lanio*	Naïve	58095240	8.7	98.59	0.02	98.15	94.55	35.89
PL_N_1909	*P. lanio*	Naïve	40590682	6.1	98.33	0.02	98.04	94.42	36.69
PL_P_1912	*P. lanio*	Wound	46513058	7	98.44	0.02	97.96	94.26	36.33
PL_P_1915	*P. lanio*	Wound	40675238	6.1	98.23	0.02	98.1	94.53	36.87
PL_P_1920	*P. lanio*	Wound	47068644	7.1	98.34	0.02	98.15	94.52	36.11
PL_SL_1926	*P. lanio*	Gram +ve	46177060	6.9	98.68	0.02	98.04	94.44	36.58
PL_SL_1927	*P. lanio*	Gram +ve	47556674	7.1	98.77	0.02	98.11	94.48	36.39
PL_SL_1930	*P. lanio*	Gram +ve	49905478	7.5	98.8	0.02	98.15	94.56	36.05
PL_SM_1936	*P. lanio*	Gram -ve	39110038	5.9	98.49	0.03	97.85	93.98	36.48
PL_SM_1939	*P. lanio*	Gram -ve	42193284	6.3	98.52	0.02	98.05	94.41	36.58
PL_SM_1940	*P. lanio*	Gram -ve	48538260	7.3	98.35	0.02	97.98	94.28	36.49

APPENDIX

Table 6.7: Pseudoalignment mapping stats from `Kallisto` logs

Species	**Sample**	**Mapping Build**	**Reads Processed**	**Reads Pseudoaligned**	**Reads Pseudoaligned (%)**
A. mellifera	AM_N_056	Amel_HAv3.1 (GCA_003254395.2)	27237028	23745681	87.18
A. mellifera	AM_N_068	Amel_HAv3.1 (GCA_003254395.2)	23508991	11009520	46.83
A. mellifera	AM_N_O72	Amel_HAv3.1 (GCA_003254395.2)	29360104	10474427	35.68
A. mellifera	AM_P_053	Amel_HAv3.1 (GCA_003254395.2)	20712635	17668715	85.3
A. mellifera	AM_P_057	Amel_HAv3.1 (GCA_003254395.2)	20405246	8761874	42.94
A. mellifera	AM_P_061	Amel_HAv3.1 (GCA_003254395.2)	24651402	21272778	86.29
A. mellifera	AM_SL_050	Amel_HAv3.1 (GCA_003254395.2)	22406740	11690960	52.18
A. mellifera	AM_SL_054	Amel_HAv3.1 (GCA_003254395.2)	22132296	19075749	86.19
A. mellifera	AM_SL_058	Amel_HAv3.1 (GCA_003254395.2)	24598864	13092867	53.23
A. mellifera	AM_SM_051	Amel_HAv3.1 (GCA_003254395.2)	22545198	19853375	88.06
A. mellifera	AM_SM_059	Amel_HAv3.1 (GCA_003254395.2)	23643235	15771681	66.71
A. mellifera	AM_SM_075	Amel_HAv3.1 (GCA_003254395.2)	30619653	26625996	86.96

Continued on next page

APPENDIX

Table 6.7: continued

Species	**Sample**	**Mapping Build**	**Reads Processed**	**Reads Pseudoaligned**	**Reads Pseudoaligned (%)**
B. terrestris	BT_A_N_001	iyBomTerr1.2 (GCA_910591885.2)	29688012	26869169	90.51
B. terrestris	BT_A_N_025	iyBomTerr1.2 (GCA_910591885.2)	35625431	32749886	91.93
B. terrestris	BT_A_N_029	iyBomTerr1.2 (GCA_910591885.2)	29222637	26692382	91.34
B. terrestris	BT_A_P_014	iyBomTerr1.2 (GCA_910591885.2)	28263874	25783586	91.22
B. terrestris	BT_A_P_026	iyBomTerr1.2 (GCA_910591885.2)	25764849	23575192	91.5
B. terrestris	BT_A_P_030	iyBomTerr1.2 (GCA_910591885.2)	28011633	25677745	91.67
B. terrestris	BT_A_SL_003	iyBomTerr1.2 (GCA_910591885.2)	31816484	29039288	91.27
B. terrestris	BT_A_SL_027	iyBomTerr1.2 (GCA_910591885.2)	28586708	26098702	91.3
B. terrestris	BT_A_SL_031	iyBomTerr1.2 (GCA_910591885.2)	28261816	25646289	90.75
B. terrestris	BT_A_SM_004	iyBomTerr1.2 (GCA_910591885.2)	24985516	22531400	90.18
B. terrestris	BT_A_SM_028	iyBomTerr1.2 (GCA_910591885.2)	22137709	20166038	91.09
B. terrestris	BT_A_SM_032	iyBomTerr1.2 (GCA_910591885.2)	24135446	22102785	91.58

Continued on next page

APPENDIX

Table 6.7: continued

Species	**Sample**	**Mapping Build**	**Reads Processed**	**Reads Pseudoaligned**	**Reads Pseudoaligned (%)**
C. australensis	CA_N_007	ASM430768v.1 (GCA_004307685.1)	22466256	13168565	58.61
C. australensis	CA_N_012	ASM430768v.1 (GCA_004307685.1)	21282603	13568826	63.76
C. australensis	CA_N_030	ASM430768v.1 (GCA_004307685.1)	22246700	13036516	58.6
C. australensis	CA_P_013	ASM430768v.1 (GCA_004307685.1)	25982762	16327993	62.84
C. australensis	CA_P_041	ASM430768v.1 (GCA_004307685.1)	19877916	12343889	62.1
C. australensis	CA_P_055	ASM430768v.1 (GCA_004307685.1)	24279801	13416947	55.26
C. australensis	CA_SL_019	ASM430768v.1 (GCA_004307685.1)	21126035	12307818	58.26
C. australensis	CA_SL_048	ASM430768v.1 (GCA_004307685.1)	26641030	16155702	60.64
C. australensis	CA_SL_056	ASM430768v.1 (GCA_004307685.1)	25966418	14624755	56.32
C. australensis	CA_SM_029	ASM430768v.1 (GCA_004307685.1)	20773939	12585291	60.58
C. australensis	CA_SM_039	ASM430768v.1 (GCA_004307685.1)	24473060	13776525	56.29
C. australensis	CA_SM_049	ASM430768v1 (GCA_004307685.1)	23853859	15211421	63.77

Continued on next page

Table 6.7: continued

Species	**Sample**	**Mapping Build**	**Reads Processed**	**Reads Pseudoaligned**	**Reads Pseudoaligned (%)**
P. lanio	PL_N_1906	ASM131383v1 (GCA_001313835.1)	22497361	17059717	75.83
P. lanio	PL_N_1908	ASM131383v1 (GCA_001313835.1)	29043615	22711910	78.2
P. lanio	PL_N_1909	ASM131383v1 (GCA_001313835.1)	20292440	15155137	74.68
P. lanio	PL_P_1912	ASM131383v1 (GCA_001313835.1)	23254094	17670872	75.99
P. lanio	PL_P_1915	ASM131383v1 (GCA_001313835.1)	20335494	15324139	75.36
P. lanio	PL_P_1920	ASM131383v1 (GCA_001313835.1)	23531406	17572336	74.68
P. lanio	PL_SL_1926	ASM131383v1 (GCA_001313835.1)	23085782	17778454	77.01
P. lanio	PL_SL_1927	ASM131383v1 (GCA_001313835.1)	23776036	18492772	77.78
P. lanio	PL_SL_1930	ASM131383v1 (GCA_001313835.1)	24949811	18898378	75.75
P. lanio	PL_SM_1936	ASM131383v1 (GCA_001313835.1)	19552850	14865578	76.03
P. lanio	PL_SM_1939	ASM131383v1 (GCA_001313835.1)	21093368	16418993	77.84
P. lanio	PL_SM_1940	ASM131383v1 (GCA_001313835.1)	24266444	18512281	76.29

APPENDIX

Table 6.8: Number of differentially expressed genes ($FDR < 0.1$) within species according to treatment. The direction of regulation is indicated by the direction of the arrows.

Treatment	***A. mellifera***	***B. terrestris***	***C. australensis***	***P. lanio***
Wound	115 ↑, 60 ↓	29 ↑, 23 ↓	202 ↑, 56 ↓	657 ↑, 308 ↓
Gram Positive	139 ↑, 73 ↓	84 ↑, 27 ↓	94 ↑, 205 ↓	980 ↑, 543 ↓
Gram Negative	284 ↑, 203 ↓	101 ↑, 165 ↓	88 ↑, 28 ↓	646 ↑, 480 ↓

Table 6.9: Overrepresented GO terms likely to be involved in immune responses across species and the direction of expression in the condition over-reppresented. There were considerably more immune-associated GO terms identified in *A. mellifera* and *B. terrestris* than the two field caught species.

	GO:ID	**GO:term**	***A.mellifera***	***B. terrestris***	***C. australensis***	***P.lanio***
Apoptosis and	GO:0010941	regulation of cell death	↑Pos			
Autophagy	GO:0070513	death domain binding	↑Pos			
	GO:0019778	Atg12 activating enzyme activity			↑Wound,↑Pos,↑Neg	
	GO:0019779	Atg8 activating enzyme activity			↑Wound,↑Pos,↑Neg	
Defensive	GO:0006967	positive regulation of antifungal peptides	↑Wound,↑Pos,↑Neg	↓Wound,↑Pos		
Response	GO:0050830	defense response to Gram-positive bacteria	↑Wound,↑Pos,↓Neg	↑Pos,↑Neg		
	GO:0009617	response to bacterium	↑Wound,↑Pos			
	GO:0009253	peptidoglycan catabolic process	↑Neg			
	GO:0009620	response to fungus	↑Wound,↑Pos,↑Neg	↓Wound,↑Pos,↓Neg		
	GO:0045429	positive regulation of nitric oxide biosynthesis	↑Neg			
	GO:0002229	defense to oomycetes	↑Wound,↑Pos,↑Neg	↑Pos,↑Neg		↑Pos
	GO:0019732	antifungal humoral response	↑Wound			
	GO:0051607	defense response to virus		↑Pos		

Continued on next page

APPENDIX

Table 6.9: continued

	GO:ID	GO:term	*A.mellifera*	*B. terrestris*	*C. australensis*	*P.lanio*
	GO:0002807	positive regulation of antimicrobial peptide biosynthetic process				↑Pos
General	GO:0004867	serine-type endopeptidase inhibitor activity	↑Wound,↑Pos,↑Neg	↑Wound,↑Pos,↑Neg		
	GO:0045087	innate immune response	↑Wound			
	GO:0097677	STAT family protein binding		↑Pos		
	GO:1990782	protein tyrosine kinase binding		↑Pos		
Melanisation,	GO:0030097	haemopoiesis	↑Wound,↑Neg			
Encapsulation,	GO:0045610	regulation of haemocyte differentiation	↑Neg			
and Phagocytosis	GO:0035172	haemocyte proliferation	↑Neg			
	GO:1903707	negative regulation of haemopoiesis		↓Neg		
	GO:0035007	regulation of melanisation defense response	↑Wound,↑Pos,↑Neg	↑Pos,↑Neg		
	GO:0035009	negative regulation of melanisation defense response	↑Wound,↑Pos	↑Wound,↑Pos,↓Neg		
	GO:0006585	dopamine biosynthetic process from tyrosine	↑Wound,↑Pos,↑Neg	↑Neg		
	GO:0042438	melanin biosynthetic process	↑Neg	↑Wound,↑Pos,↑Neg		

Continued on next page

Table 6.9: continued

	GO:ID	GO:term	*A.mellifera*	*B. terrestris*	*C. australensis*	*P.lanio*
	GO:0048067	cuticle pigmentation	↑Neg	↑Wound,↑Pos,↑Neg		
	GO:0004058	aromatic-L-amino-acid decarboxylase activity	↑Wound,↑Pos,↑Neg	↑Neg		
	GO:0007564	regulation of chitin-based cuticle tanning		↑Neg		
	GO:0006583	melanin biosynthesis process from tyrosine		↑Neg		↑Wound,↓Pos,↑Neg
	GO:0048022	negative regulation of melanin biosynthetic process		↑Neg		
	GO:0090383	phagosome acidification				↑Pos
Recognition	GO:0042834	peptidoglycan binding	↑Neg			
	GO:0001872	(1− >3)-β-D-glucan binding	↑Pos	↑Pos,↓Neg	↑Wound,↑Pos,↑Neg	
	GO:0032491	detection of molecule of fungal origin	↑Pos	↑Pos,↑Neg	↑Wound,↑Pos,↑Neg	
	GO:0002752	cell surface pattern recognition receptor signalling pathway		↑Pos	↑Wound,↑Pos,↑Neg	
	GO:0038187	pattern recognition receptor activity		↑Pos	↑Wound,↑Pos,↑Neg	
	GO:0061060	negative regulation of peptidoglycan recognition protein				

Continued on next page

APPENDIX

Table 6.9: continued

	GO:ID	GO:term	*A.mellifera*	*B. terrestris*	*C. australensis*	*P.lanio*
		signalling pathway			↑Wound	
Toll Signalling	GO:0008063	Toll signalling pathway	↑Wound,↑Pos,↑Neg	↓Neg		
Pathway	GO:0045751	negative regulation of Toll signalling pathway	↑Wound,↑Pos	↑Wound,↑Pos,↑Neg		
	GO:0005121	Toll binding		↑Wound		
	GO:0070976	TIR domain binding	↑Wound,↑Pos,↑Neg	↑Pos,↑Neg		↑Pos
Wounding	GO:0009611	response to wounding	↑Wound	↓Neg		
	GO:0042060	wound healing	↓Wound			
	GO:0045752	positive regulation of Toll signalling pathway		↓Wound		
	GO:0006723	cuticle hydrocarbon biosynthetic process			↑Wound	
	GO:0040003	chitin-based cuticle development				↑Pos,↑Neg

APPENDIX

6.4 Appendix C: Chapter 4 Supplementary Material

6.4.1 Figures

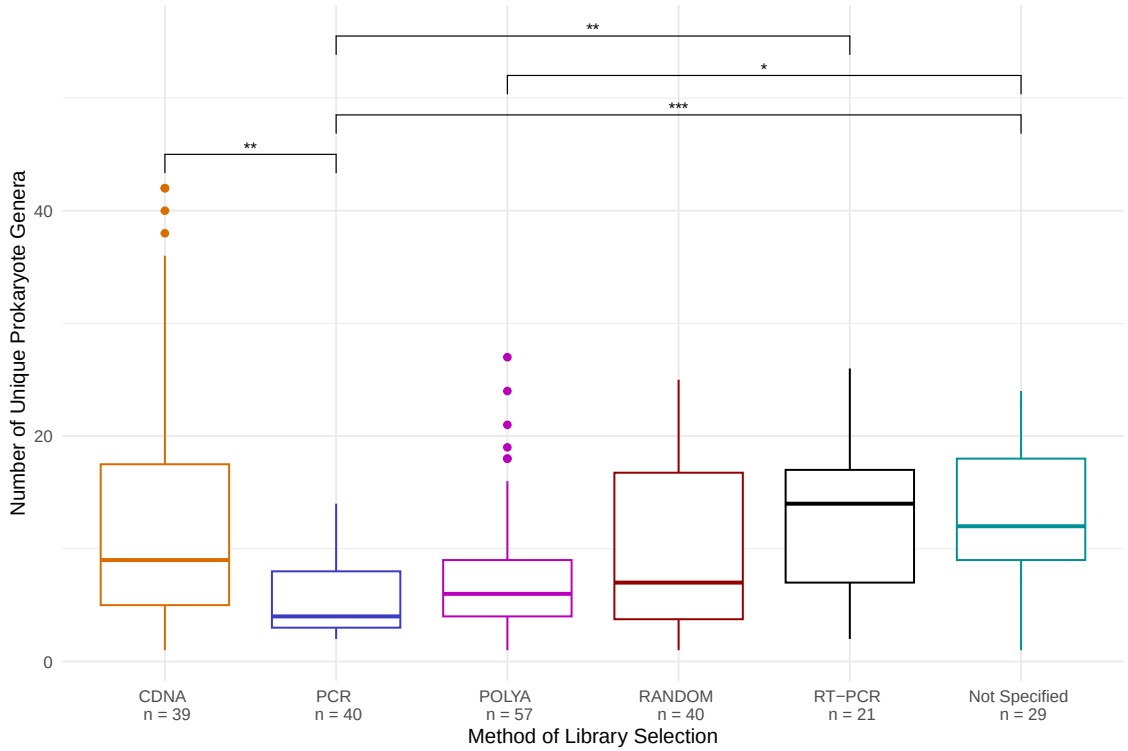

Figure 6.2: Number of unique prokaryotic genera detected in a sample versus the method of sample library preparation. Capitalised preparations are those taken directly from SRA metadata as was input when sample data was uploaded. Instances where no answer was given in this field are categorised as "Not Specified". The majority of samples were from libraries prepared with poly(A) enrichment, which only differed significantly from libraries with unspecified preparations ($P.adj = 0.0149$, $Z = 3.291$, Dunn's test). There were otherwise other significant differences between certain library preparations and number of detected bacterial genera (see **Appendix** 6.4: Table 6.12). Significance: $P < 0.05 = *$, $P < 0.01 = **$, $P < 0.001 = ***$, $P < 0.0001 = ****$.

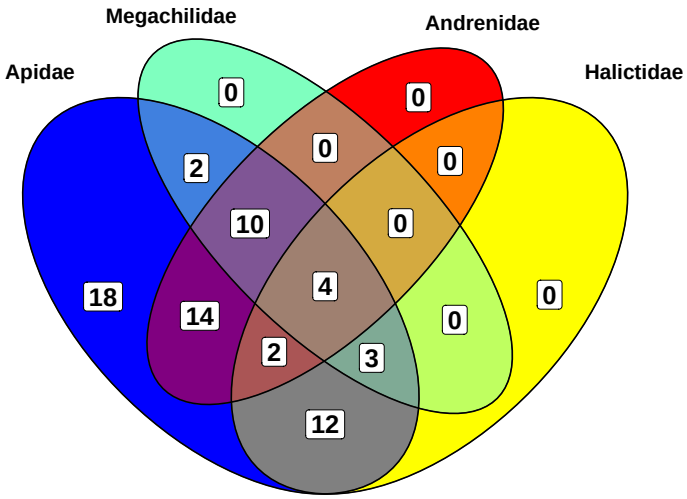

Figure 6.3: Overlap of bacterial taxa detected in different host families. The only unique taxa are found in Apidae.

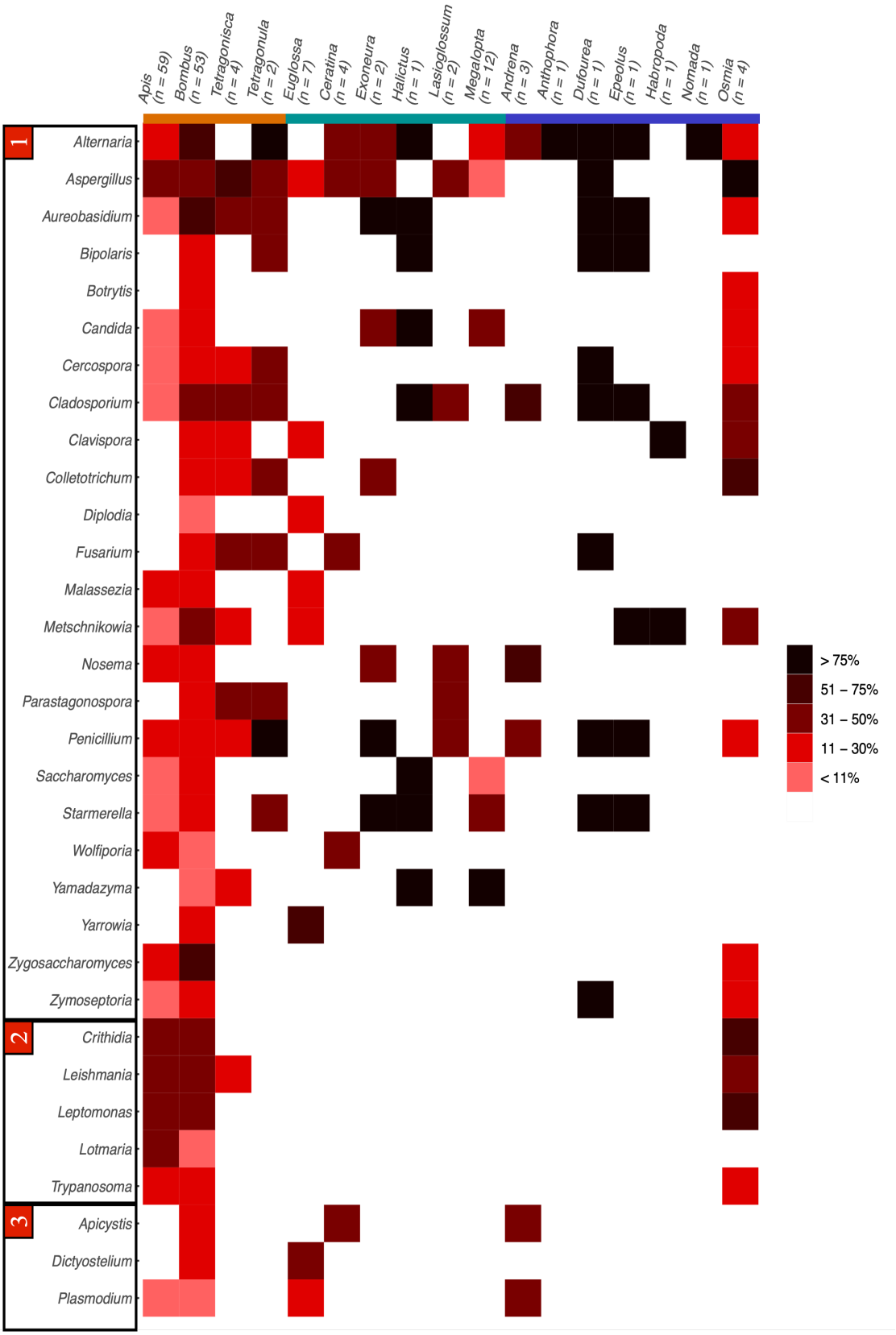

Figure 6.4: Heatmap of all detected eukaryote taxa and their prevalence in each genus of host samples tested after filtering. Eukaryotic taxa are ordered into 1) fungi, 2) trypanosomatids and 3) other.

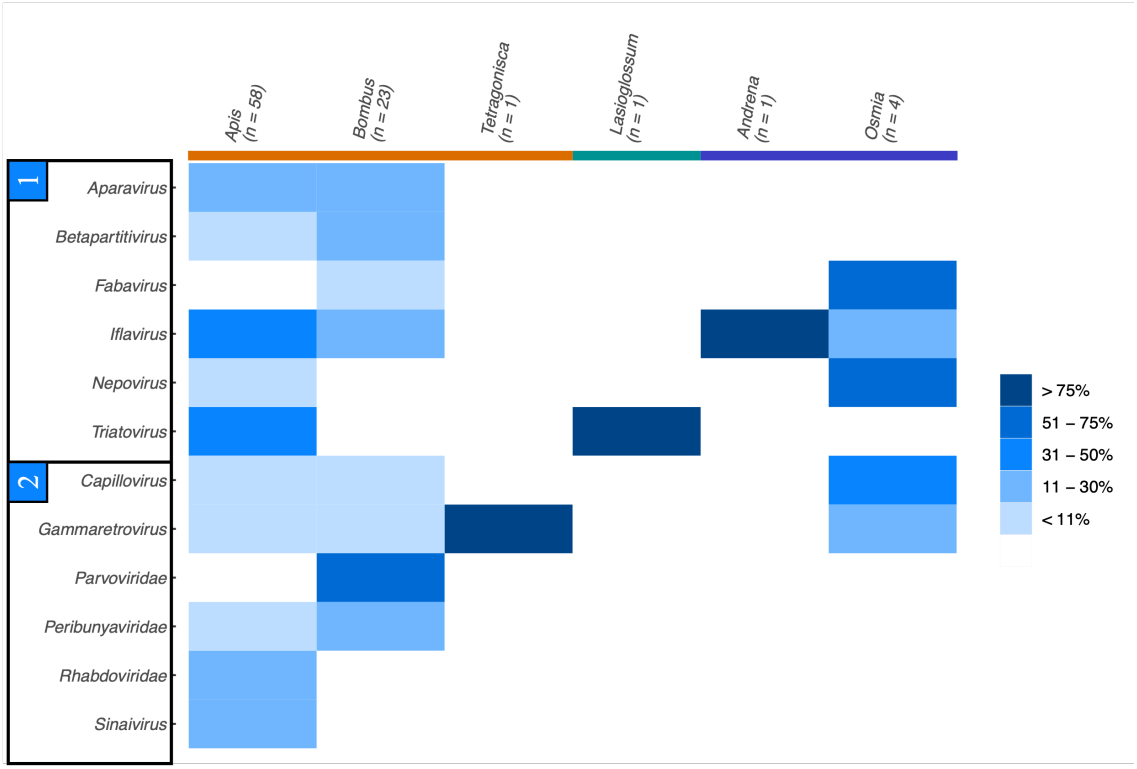

Figure 6.5: Detected viral prevalence in each host genus that passed data filtering. Host genera are coloured according to social lifestyle category: orange = obligately eusocial, facultatively eusocial = green, blue = solitary.

6.4.2 Tables

Table 6.10: List of non-*Apis mellifera* species included in the analysis with genome mapping information

Species	n	Genome Species	Assembly	GenBank	Reference
Andrena spp.	4	*Andrena dorsata*	iyAndDors1.1	GCA_929108735.1	Darwin Tree of Life Project Consortium 2022
Andrena camellia	4	*Andrena haemorrhoa*	iyAndHaem1.1	GCA_910592295.1	Darwin Tree of Life Project Consortium 2022
Andrena cineraria	1	*Andrena hattorfiana*	iyAndHatt1.1	GCA_944738655.1	Darwin Tree of Life Project Consortium 2022
Andrena fulva	1	*Andrena haemorrhoa*	iyAndHaem1.1	GCA_910592295.1	Darwin Tree of Life Project Consortium 2022
Andrena haemorrhoa	2	*Andrena haemorrhoa*	iyAndHaem1.1	GCA_910592295.1	Darwin Tree of Life Project Consortium 2022
Andrena vaga	1	*Andrena hattorfiana*	iyAndHatt1.1	GCA_944738655.1	Darwin Tree of Life Project Consortium 2022
Anthophora plumipes	1	*Habropoda laboriosa*	ASM126327v1	GCA_001263275.1	Kapheim et al. 2015
Apis cerana	5	*Apis cerana*	ACSNU-2.0	GCA_001442555.1	Park et al. 2015
Apis mellifera	87	*Apis mellifera*	# undergone in CZID pipeline "Bee" host genome		
Bombus spp.	1	*Bombus terrestris*	iyBomTerr1.2	GCA_910591885.2	Darwin Tree of Life Project Consortium 2022
Bombus breviceps	1	*Bombus breviceps*	ASM1482592v1	GCA_014825925.1	Sun et al. 2021
Bombus confusus	1	*Bombus confusus*	ASM1473747v1	GCA_014737475.1	Sun et al. 2021
Bombus consobrinus	1	*Bombus confusus*	ASM1473745v1	GCA_014737475.1	Sun et al. 2021
Bombus difficillimus	1	*Bombus difficillimus*	ASM1473752v1	GCA_014737525.1	Sun et al. 2021
Bombus haemorrhoidalis	1	*Bombus haemorrhoidalis*	ASM1482597v1	GCA_014825975.1	Sun et al. 2021

Continued on next page

APPENDIX

Table 6.10: continued

Species	n	Genome Species	Assembly	GenBank	Reference
Bombus ignitus	1	*Bombus ignitus*	ASM1482587v1	GCA_014825875.1	Sun et al. 2021
Bombus lucorum	2	*Bombus terrestris*	iyBomTerr1.2	GCA_910591885.2	Darwin Tree of Life Project Consortium 2022
Bombus opulentus	1	*Bombus opulentis*	ASM1473740v1	GCA_014737405.1	Sun et al. 2021
Bombus pascuorum	9	*Bombus pascuorum*	iyBomPasc1.1	GCA_905332965.1	Darwin Tree of Life Project Consortium 2022
Bombus picipes	1	*Bombus picipes*	ASM1473748v1	GCA_014737485.1	Sun et al. 2021
Bombus pyrosoma	7	*Bombus pyrosoma*	ASM1482585v1	GCA_014825855.1	Sun et al. 2021
Bombus rupestris	1	*Bombus skorikovi*	ASM1473735v1	GCA_014737355.1	Sun et al. 2021
Bombus sibiricus	1	*Bombus sibiricus*	ASM1473750v1	GCA_014737505.1	Sun et al. 2021
Bombus soroeensis	1	*Bombus soreensis*	ASM1473736v1	GCA_014737365.1	Sun et al. 2021
Bombus superbus	1	*Bombus superbus*	ASM1473738v1	GCA_014737385.1	Sun et al. 2021
Bombus terrestris	20	*Bombus terrestris*	iyBomTerr1.2	GCA_910591885.2	Darwin Tree of Life Project Consortium 2022
Bombus terricola	12	*Bombus terrestris*	iyBomTerr1.2	GCA_910591885.2	Darwin Tree of Life Project Consortium 2022
Bombus turneri	1	*Bombus turneri*	ASM1482582v1	GCA_014825825.1	Sun et al. 2021
Bombus waltoni	1	*Bombus waltoni*	ASM1473739v1	GCA_014737395.1	Sun et al. 2021
Ceratina australensis	5	*Ceratina australensis*	Caustv1	GCA_004307685.1	Rehan et al. 2018
Dufourea novaeangliae	1	*Dufourea novaeangliae*	ASM127255v1	GCA_001272555.1	Kapheim et al. 2015
Epeolus variegatus	1	*Nomada fabriciana*	iyNomFabr1.1	GCA_907165295.1	Darwin Tree of Life Project Consortium 2022
Eufriesea mexicana	1	*Eufriesea mexicana*	ASM148370v1	GCA_001483705.1	Kapheim et al. 2015
Euglossa dilemma	7	*Eufriesea mexicana*	ASM148370v1	GCA_001483705.1	Kapheim et al. 2015
Euglossa viridissima	20	*Eufriesea mexicana*	ASM148370v1	GCA_001483705.1	Kapheim et al. 2015
Exoneura spp.	1	*Exoneura robusta*	ASM1945341v1	GCA_019453415.1	Shell et al. 2021

Continued on next page

Table 6.10: continued

Species	**n**	**Genome Species**	**Assembly**	**GenBank**	**Reference**
Habropoda laboriosa	1	*Habropoda laboriosa*	ASM126327v1	GCA_001263275.1	Kapheim et al. 2015
Halictus sexcinctus	1	*Lasioglossum laticentre*	iyLasLatv2.1	GCA_916610255.1	Darwin Tree of Life Project Consortium 2022
Lasioglossum spp.	2	*Lasioglossum laticentre*	iyLasLatv2.1	GCA_916610255.1	Darwin Tree of Life Project Consortium 2022
Megalopta genalis	22	*Megalopta genalis*	USU_MGEN_1.2	GCA_011865705.1	Kapheim et al. 2020
Nomada lathburiana	1	*Nomada fabriciana*	iyNomFabr1.1	GCA_907165295.1	Darwin Tree of Life Project Consortium 2022
Osmia bicornis	8	*Osmia bicornis*	iOsmBic2.1	GCA_907164935.1	Darwin Tree of Life Project Consortium 2022
Osmia cornuta	4	*Osmia bicornis*	iOsmBic2.1	GCA_907164935.1	Darwin Tree of Life Project Consortium 2022
Tetragonisca angustula	6	*Tetragonula mellipes*	Tetragonula_mellipes1_1.1	GCA_011634685.1	Unpublished
Tetragonula carbonaria	2	*Tetragonula mellipes*	Tetragonula_mellipes1_1.1	GCA_011634685.1	Unpublished

APPENDIX

Table 6.11: Pairwise PERMANOVA assessing differences between factor levels per sociality / host family / continent in bacterial community analysis. Adjusted p-value computed using Benjamini-Hochburg (FDR) correction. * : < 0.05, ** : < 0.01, *** : < 0.001.

SOCIALITY				
Factor Level 1	**Factor Level 2**	P	P**.adj**	**Significance**
O. Eusocial	F. Eusocial	0.0063	0.0189	*
O. Eusocial	Solitary	0.5898	0.5898	
F. Eusocial	Solitary	0.3425	0.51375	
FAMILY				
Factor Level 1	**Factor Level 2**	P	P**.adj**	**Significance**
Apidae	Megachilidae	0.0095	0.01425	*
Apidae	Andrenidae	0.2973	0.35676	
Apidae	Halictidae	1.00E-04	4.00E-04	***
Megachilidae	Andrenidae	0.6234	0.6234	
Megachilidae	Halictidae	2.00E-04	4.00E-04	***
Andrenidae	Halictidae	2.00E-04	4.00E-04	***
CONTINTENT				
Factor Level 1	**Factor Level 2**	P	P**.adj**	**Significance**
Europe	Oceania	0.0011	0.0165	*
Europe	Asia	0.2094	0.3141	
Europe	North America	0.1456	0.3141	
Europe	Africa	0.7528	0.7528	
Europe	South America	0.3049	0.4158	
Oceania	Asia	0.1639	0.3141	
Oceania	North America	0.1374	0.3141	
Oceania	Africa	0.2008	0.3141	
Oceania	South America	0.3705	0.4275	
Asia	North America	0.0326	0.2445	
Asia	Africa	0.0937	0.3141	
Asia	South America	0.5599	0.5999	
North America	Africa	0.1943	0.3141	
North America	South America	0.1593	0.3141	
Africa	South America	0.3333	0.4167	

Table 6.12: Results from Kruskal-Wallis rank sum test and post hoc Dunn's test comparing methods of sample preparation and number of unique detected bacterial taxa. P-values adjusted using Bonferonni correction. $< 0.05 = *; < 0.01 = **; < 0.001 = * * *; < 0.0001 = * * **$

Kruskal-Wallis				
	χ^2	df	P	
	29.685	5	$1.701e-05$	
Dunn's Post-hoc Test				
Comparison	Z	P	P**.adj**	**Significance**
CDNA - Not Specified	-0.93580822	3.49E-01	1.0000	
CDNA - PCR	3.56666066	3.62E-04	0.0054	**
Not Specified - PCR	4.23175839	2.32E-05	0.0003	***
CDNA - POLYA	2.5088042	1.21E-02	0.1817	
Not Specified - POLYA	3.29170152	9.96E-04	0.0149	*
PCR - POLYA	-1.3636648	1.73E-01	1.0000	
CDNA - RANDOM	1.27205774	2.03E-01	1.0000	
Not Specified - RANDOM	2.11455179	3.45E-02	0.5170	
PCR - RANDOM	-2.30926507	2.09E-02	0.3139	
POLYA - RANDOM	-1.1397934	2.54E-01	1.0000	
CDNA - RT-PCR	-0.74222234	4.58E-01	1.0000	
Not Specified - RT-PCR	0.09969933	9.21E-01	1.0000	
PCR - RT-PCR	-3.72392354	1.96E-04	0.0029	**
POLYA - RT-PCR	-2.82934746	4.66E-03	0.0700	
POLYA - RANDOM	-1.80775584	7.06E-02	1.0000	

Table 6.13: Prevalence of corbiculate core and other bee associated bacterial microbes in the three corbiculate tribes and other bees (ordered by sociality).

Microbial Taxa	**Sample Category**	**Prevalence**	**Prevalence Factor**
Gilliamella	Apini	0.791	61 - 80%
Gilliamella	Bombini	0.825	81 - 100%
Gilliamella	F.Eusocial Non-Corbiculates	0	
Gilliamella	Meliponini	0.429	41 - 60%
Gilliamella	Euglossini	0	
Gilliamella	Solitary Non-Corbiculates	0.042	< 5%
Snodgrassella	Apini	0.779	61 - 80%
Snodgrassella	Bombini	0.719	61 - 80%
Snodgrassella	F.Eusocial Non-Corbiculates	0	
Snodgrassella	Meliponini	0.714	61 - 80%
Snodgrassella	Euglossini	0	
Snodgrassella	Solitary Non-Corbiculates	0.042	< 5%
Bombilactobacillus	Apini	0.663	61 - 80%
Bombilactobacillus	Bombini	0.316	21 - 40%
Bombilactobacillus	F.Eusocial Non-Corbiculates	0	
Bombilactobacillus	Meliponini	0.143	5 - 20%
Bombilactobacillus	Euglossini	0	
Bombilactobacillus	Solitary Non-Corbiculates	0.042	< 5%
Lactobacillus: Firm-5	Apini	0.826	81 - 100%
Lactobacillus: Firm-5	Bombini	0.526	41 - 60%
Lactobacillus: Firm-5	F.Eusocial Non-Corbiculates	0	
Lactobacillus: Firm-5	Meliponini	0.857	81 - 100%
Lactobacillus: Firm-5	Euglossini	0	
Lactobacillus: Firm-5	Solitary Non-Corbiculates	0	
Bifidobacterium	Apini	0.570	41 - 60%
Bifidobacterium	Bombini	0.474	41 - 60%
Bifidobacterium	F.Eusocial Non-Corbiculates	0	
Bifidobacterium	Meliponini	1	81 - 100%
Bifidobacterium	Euglossini	0	
Bifidobacterium	Solitary Non-Corbiculates	0.125	5 - 20%
Frischella	Apini	0.616	61 - 80%
Frischella	Bombini	0.456	41 - 60%
Frischella	F.Eusocial Non-Corbiculates	0	
Frischella	Meliponini	0.286	21 - 40%
Frischella	Euglossini	0	
Frischella	Solitary Non-Corbiculates	0.042	< 5%
Parasaccharibacter	Apini	0.105	5 - 20%
Parasaccharibacter	Bombini	0.070	5 - 20%
Parasaccharibacter	F.Eusocial Non-Corbiculates	0.222	21 - 40%
Parasaccharibacter	Meliponini	0.143	5 - 20%
Parasaccharibacter	Euglossini	0.077	5 - 20%
Parasaccharibacter	Solitary Non-Corbiculates	0	
Bombella	Apini	0.198	5 - 20%

Continued on next page

Table 6.13: continued

Microbial Taxa	**Sample Category**	**Prevalence**	**Prevalence Factor**
Bombella	Bombini	0.070	5 - 20%
Bombella	F.Eusocial Non-Corbiculates	0.444	41 - 60%
Bombella	Meliponini	0.429	41 - 60%
Bombella	Euglossini	0.038	$< 5\%$
Bombella	Solitary Non-Corbiculates	0	
Apibacter	Apini	0.140	5 - 20%
Apibacter	Bombini	0.351	21 - 40%
Apibacter	F.Eusocial Non-Corbiculates	0	
Apibacter	Meliponini	0	
Apibacter	Euglossini	0	
Apibacter	Solitary Non-Corbiculates	0	
Apilactobacillus	Apini	0.337	21 - 40%
Apilactobacillus	Bombini	0.053	5 - 20%
Apilactobacillus	F.Eusocial Non-Corbiculates	0.556	41 - 60%
Apilactobacillus	Meliponini	0.143	5 - 20%
Apilactobacillus	Euglossini	0.038	$< 5\%$
Apilactobacillus	Solitary Non-Corbiculates	0	
Commensalibacter	Apini	0.523	41 - 60%
Commensalibacter	Bombini	0.088	5 - 20%
Commensalibacter	F.Eusocial Non-Corbiculates	0	
Commensalibacter	Meliponini	0	
Commensalibacter	Euglossini	0	
Commensalibacter	Solitary Non-Corbiculates	0.083	5 - 20%
Bartonella	Apini	0.558	41 - 60%
Bartonella	Bombini	0.018	$< 5\%$
Bartonella	F.Eusocial Non-Corbiculates	0	
Bartonella	*Meliponini*	0	
Bartonella	*Euglossini*	0	
Bartonella	Solitary Non-Corbiculates	0	

Continued on next page

Table 6.14: Average relative abundances (RA) of *Wolbachia* in tribes were it was detected (median and mean). Detected levels were substantially lower in the obligately eusocial tribes (Apini and Bombini) than in the solitary tribes (Andrenini, Osmiini).

Tribe	**Median**	**Mean**	**SD**
Apini	0	0.004	0.029
Bombini	0	0.002	0.013
Andrenini	0.571	0.570	0.431
Osmiini	< 0.001	0.012	0.020