

Hunting for healthy microbiomes: determining the core microbiomes of *Ceratina*, *Megalopta*, and *Apis* bees and how they associate with microbes in bee collected pollen

Peter Graystock, ^{1,*}

Email peter@graystock.info

Sandra M. Rehan, ²

Quinn S. McFrederick, ¹

¹ Department of Entomology, University of California Riverside, 900 University Avenue, Riverside, CA, 92521 USA

² Department of Biological Sciences, University of New Hampshire, 46 College Road, Durham, NH, 03824 USA

Abstract

Social corbiculate bees such as honey bees and bumble bees maintain a specific beneficial core microbiome which is absent in wild bees. It has been suggested that maintaining this microbiome can prevent disease and keep bees healthy. The main aim of our study was to identify if there are any core bacterial groups in the non-corbiculate bees *Ceratina* and *Megalopta* that have been previously overlooked. We additionally test for associations between the core bee microbes and pollen provisions to look for potential transmission between the two. We identify three enterotypes in *Ceratina* samples, with thirteen core bacterial phylotypes in *Ceratina* females: *Rosenbergiella*, *Pseudomonas*, *Gilliamella*, *Lactobacillus*, *Caulobacter*, *Snodgrassella*, *Acinetobacter*, *Corynebacterium*, *Sphingomonas*, *Commenralibacter*, *Methylobacterium*, *Massilia*, and *Stenotrophomonas*, plus 19 in pollen (6 of which are shared by bees). Unlike *Apis* bees, whose gut microbial community differs compared to their pollen, *Ceratina* adults and pollen largely share a similar microbial composition and enterotype

difference was largely explained by pollen age. *Megalopta* displays a highly diverse composition of microbes throughout all adults, yet *Lactobacillus* and *Saccharibacter* was prevalent in 90% of adults as core bacteria. Only *Lactobacillus* was both a core bee and pollen provision microbe in all three species. The consequences of such diversity in core microbiota between bee genera and their associations with pollen are discussed in relation to identifying potentially beneficial microbial taxa in wild bees to aid the conservation of wild, understudied, non-model bee species.

AQ1

Keywords

Core microbiome
Hymenoptera
Pollen diet
16S
Foraging ecology
Bacterial diversity
Enterotype
Pollen provision

Electronic supplementary material

The online version of this article (doi: 10.1007/s10592-017-0937-7) contains supplementary material, which is available to authorized users.

Introduction

Microbial associates, ranging from mutualists to parasites, are a ubiquitous feature of life on earth. Diverse and dynamic communities of bacteria exist within the guts of insects. While parasites often provide the most dramatic and observable effect on the host, the roles of mutualists within the gut are increasingly found to have important and often profound effects on host health and fitness. In crickets (*Acheta domesticus*) gut bacteria have been shown to increase digestive efficiency and allow hosts to grow steadily when faced with diet change (Kaufman and Klug 1991). The evolution of herbivorous ants was likely facilitated by the symbiosis between ants and a clade of nitrogen-fixing bacteria, Rhizobiales. These symbionts are specific to ant microbiomes and are highly abundant in herbivorous ants (Russell et al. 2009). Profoundly, in *D. melanogaster* the gut microbe *Lactobacillus plantarum* influences the host's sex pheromone production and thus has a key role in sexual selection

(Sharon et al. 2010). With such a clear and often instrumental role in host fitness, it is surprising that the promotion and maintenance of healthy microbiota is still an often-overlooked concept in conservation biology.

Identifying what may be considered a ‘healthy microbiome’ in non-model organisms is a great challenge in microbial ecology. By understanding the composition of a particular species ‘healthy microbiome’, dysbiosis can be quickly diagnosed and appropriate action taken to reestablish a healthy microbiota in its host (Cho and Blaser 2012; Kong et al. 2012). Identifying the core microbiome, however, can be a challenge, due to the large diversity of gut microbes within an individual. Additionally, the countless interactions within this microbial community make understanding the function of each microbe elusive. Our understanding of microbiomes in non-model organisms is particularly hampered by a dearth of samples (rare species) or a low-priority research effort. Microbes that are more likely to have important functional roles can be identified in a population by determining the ‘core microbiota’ (Shade and Handelsman 2012). A core microbiota lists the shared microbes present in the majority of healthy replicate samples (Turnbaugh et al. 2007). This may be consistent throughout a species or it may subdivide into several clusters/enterotypes (Hamady and Knight 2009).

Insect microbiota tend to be dominated by two bacterial phyla, Proteobacteria and Firmicutes, but there is no insect-wide core microbiota, as variation exists between and often within insect groups (Colman et al. 2012; Yun et al. 2014). Variation in bacterial communities across insect hosts may be driven by a variety of factors including host phylogeny, geography, and diet. Across the world there are approximately 17,000 named species of bees that exhibit a variety of diets and either social or solitary lifestyles (Michener 2007). The honey bee (*Apis mellifera*) is considered the de facto model bee species, with bumble bee (*Bombus*) species also frequently used as a model in bee research (Woodard et al. 2015). Both *Apis* and *Bombus* are corbiculate bees, which account for only 5% of all bee species (Gallai et al. 2009; Martins et al. 2014), but due to their model status and commercial importance have the best studied microbiome of bees. The honey bee, *A. mellifera*, has a core set of bacterial species that are present in all workers (Martinson et al. 2011). These core bacterial species are similar but distinct from that of bumble bees (Martinson et al. 2011; Koch and Schmid-Hempel 2011a). The split between honey bee and bumble bee lineages occurred around 90 Mya (Cardinal et al. 2010). Within the *Bombus* genus, most speciation events occurred in the last 10 My and subtle yet consistent microbiota differences can be detected between bumble bee species within Central Europe (Koch and Schmid-Hempel

2011a). Unlike the consistent, core microbial community found in healthy honey bees, bumble bees exhibit two enterotypes, one dominated by the core bacteria and a second dominated by possible pathogens (Li et al. 2015).

The same core bacteria found in these model bees appear to be absent in non-corbiculate bees (Martinson et al. 2011; Koch et al. 2013). The stability and host-specificity of such an highly derived ‘core’ microbiota in bumble bees and honey bees is believed to be, in part, due to social transmission (Koch et al. 2013; Powell et al. 2014). Solitary and primitively social bees besides the bumble bees, on the other hand, appear to rely more on horizontal transmission of microbes from the environment (McFrederick et al. 2012, 2014). Environmental transmission is likely to occur at shared flowers where a diversity of microbes including bacteria, protists, fungi and yeasts can transmit during pollen and nectar foraging (Brysch-Herzberg 2004; McFrederick et al. 2014, 2016; Graystock et al. 2015; McFrederick and Rehan 2016). Reliance on environmental transmission does not preclude healthy solitary bees from having a core microbial community however, though it may make the acquisition more sporadic. Here, we seek to identify candidate microbial taxa that may be beneficial to bees by looking for core bacteria. We identify the core microbiota of *Ceratina calcarata*, a solitary non-corbiculate bee that is a close relative of *A. mellifera*. Secondly, we determine if a relationship exists between core microbes in adult bee guts and those found in pollen provisions, the food on which larvae are reared. Finally, to place our results in context, we conduct the same analyses using publicly available data from another apid bee, *A. mellifera*, and their corbicular pollen (Corby-Harris et al. 2014a), plus the facultatively eusocial halictid bees, *Megalopta* spp. and their pollen provisions (McFrederick et al. 2014).

Methods

To analyze microbiome composition across adult bees and their associated pollen samples, we analyzed 96 previously published pollen provision samples from *C. calcarata* (McFrederick and Rehan 2016) along with the 16 adult samples collected for this study. *Ceratina calcarata* is a subsocial member of the Apidae (Rehan and Richards 2010), and we collected new samples for this study at the same site (Durham, New Hampshire) and date (July 2014) as in McFrederick and Rehan (2016).

To look for similar associations in other bee-pollen samples we analyzed complete datasets of *Apis* bees and their corbicular pollen, plus *Megalopta* bees and their pollen provisions from previously published work (Corby-Harris et al. 2014a; McFrederick et al. 2014). Details for sample collection are included in the pertinent

publications. Details regarding sample identity, collection date and site, and sample type are included in supplementary table 1.

Molecular methods

We extracted DNA of samples and control blanks by following the protocol in the DNeasy Blood and tissue kit (Qiagen, Valencia, CA), but with the addition of thorough tissue lysing with steel and glass beads to ensure extraction of recalcitrant gram positive bacteria (Engel et al. 2013). Briefly, samples were lysed in 180 μ l of ATL buffer for 3 min at 30hz before being incubated with 20 μ l of Proteinase K at 57 °C overnight. The standard Qiagen protocol was thereafter employed. We sequenced DNA using the Illumina MiSeq platform with bacterial 16S primers 799F–1115R for all *C. calcarata* samples. Our control blanks did not visibly amplify during library prep PCR and whilst this does not preclude errant DNA, it does suggest there was not a significant level of contamination and the control blanks were subsequently not sequenced; however, doing so would enhance the identification of low-read contaminants.

Bioinformatics

We used MacQIIME 1.9.1 (Caporaso et al. 2010b) to run initial quality controls and for demultiplexing fastq files. We then used Sumacust for OTU clustering at 97% sequence identify and vSearch for chimera checking (Mercier et al. 2013; Rognes et al. 2016). We used the Ribosomal Database Project (RDP) Naïve Bayesian Classifier in QIIME to assign taxonomy to OTUs (Wang et al. 2007). We also ran additional BLAST searches using the 16S microbial database from NCBI (accessed 28 October, 2016). We then compared RDP results against the top BLAST hit. Any discrepancies were then resolved based on RDP confidence and BLAST percentage identity. At the same time any non-bacterial hits and contaminants such as the intercellular bacteria *Wolbachia* and *Sodalis* were removed from the dataset. Raw *Ceratina* adults data are available on the NCBI Sequence database and have been assigned accession numbers SAMN06141735 to SAMN06141750.

To analyze *C. calcarata* pollen and adult samples, we aligned the quality-filtered dataset using the PyNAST aligner and the greengenes database (DeSantis et al. 2006; Caporaso et al. 2010a). Next, we filtered the resulting sequence alignment and checked the results manually using Mesquite v.3.04 (Mesquite Project Team 2014). We reconstructed the phylogeny of the bacterial OTUs using FastTree (Price et al. 2009). To determine subsampling depth for subsequent analyses, we plotted rarefaction curves. Based on the saturation point of these curves, we determined

that any samples with fewer than 668 reads should be removed. In addition to our *Ceratina* dataset, we also used previously published data on *Megalopta* gut microbes and their pollen provisions, plus *Apis* gut (crop and full gut) samples and their corbicular pollen (Corby-Harris et al. 2014a; McFrederick et al. 2014). For all three OTU tables, OTUs were binned by genus and the core microbial genera were identified using a QIIME script, whereby microbial taxa must be present in over 50% of samples and have an average relative abundance $> 1\%$ within the sample type. To identify if any clustering/enterotypes were present within each dataset, we employed the same method as both Arumugam et al. (2011) and Li et al. (2015). Briefly, from the summarized OTU table, a Jensen-Shannon distance matrix was generated and optimal cluster number was evaluated via Calinski–Harabasz (CH) index with validation of cluster strength by the silhouette method (Caliński and Harabasz 1974; Rousseeuw 1987). Using this cluster k number, the Jensen-Shannon distance matrix was used to generate the clusters with the Partitioning Around Mediod (PAM) method (Reynolds et al. 2006). Enterotypes/clusters were visualized in a principal coordinate analysis (PCoA) plot. When the clusters contained three or more samples, we compared the relative abundance of core bacterial taxa between clusters using a Mann–Whitney U test, or if there were more than two clusters, Kruskal–Wallis tests. Finally, p -values were adjusted using the Benjamini–Hochberg FDR method to account for potential false discovery (Benjamini and Hochberg 1995).

Results

After filtering, the *Ceratina* dataset contained 15 females and 94 pollen samples, the *Megalopta* dataset contained 20 females and 23 pollen samples, and the *Apis* dataset contained 41 bee and 13 pollen samples.

Core bee microbiota

Overall thirteen microbe taxa were described as core bacteria-present in female *Ceratina* samples (Fig. 1). Two were identified in *Megalopta* and four in *Apis* (Fig. 2). Only one of these taxa (*Lactobacillus*) was identified as a core microbial group in all three of the bee genera.

Fig. 1

Relative abundance of microbial taxa identified as being either core microbes ($>50\%$ prevalence, $>1\%$ relative abundance) in *Ceratina* females (*orange panel*), *Ceratina* pollen provisions (*light blue panel*) or in both (*purple panel*). The average relative

abundances are shown for both female *Ceratina* (red bars) plus their pollen provisions (blue bars) in all core taxa. Error bars represent standard error

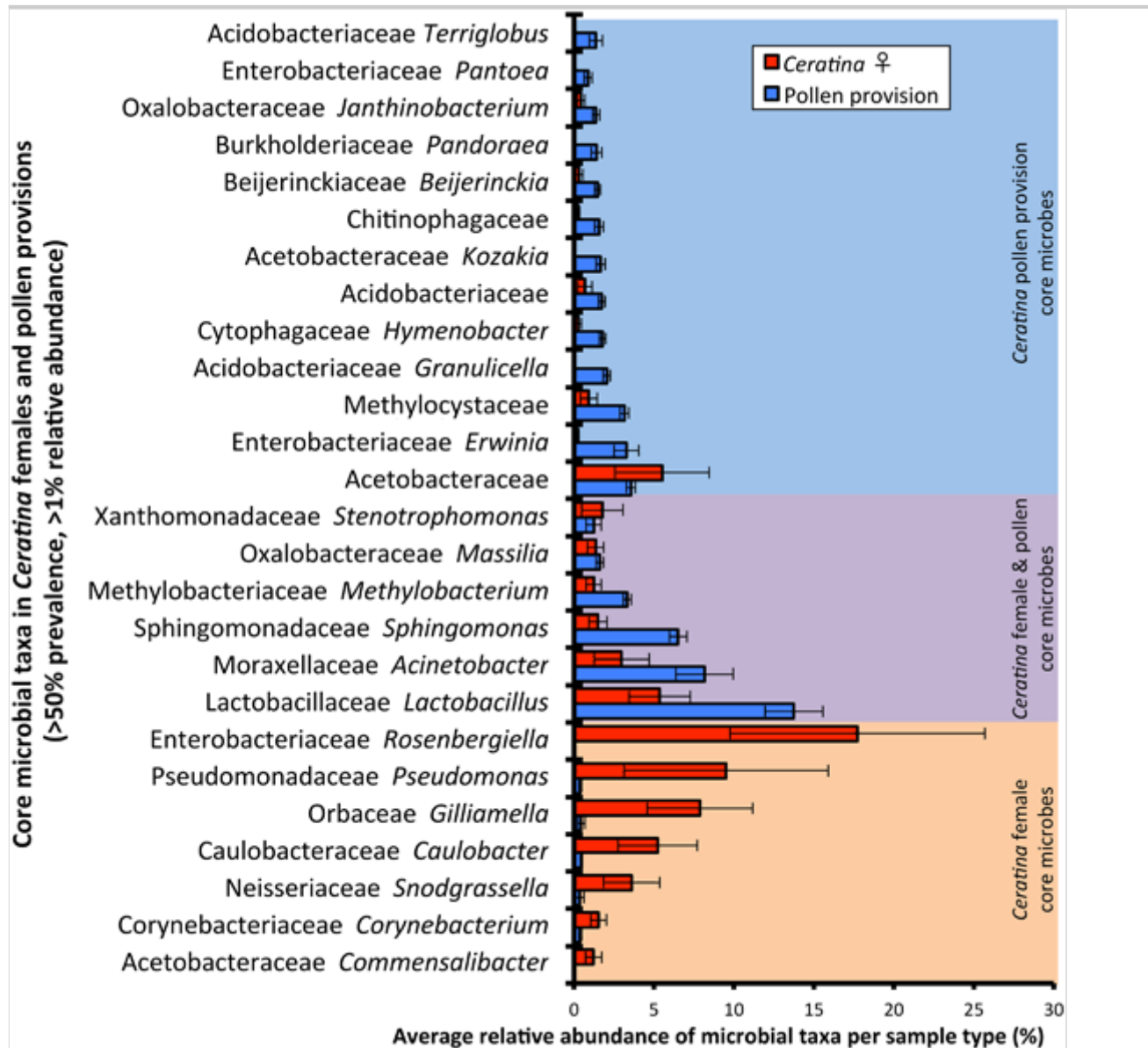
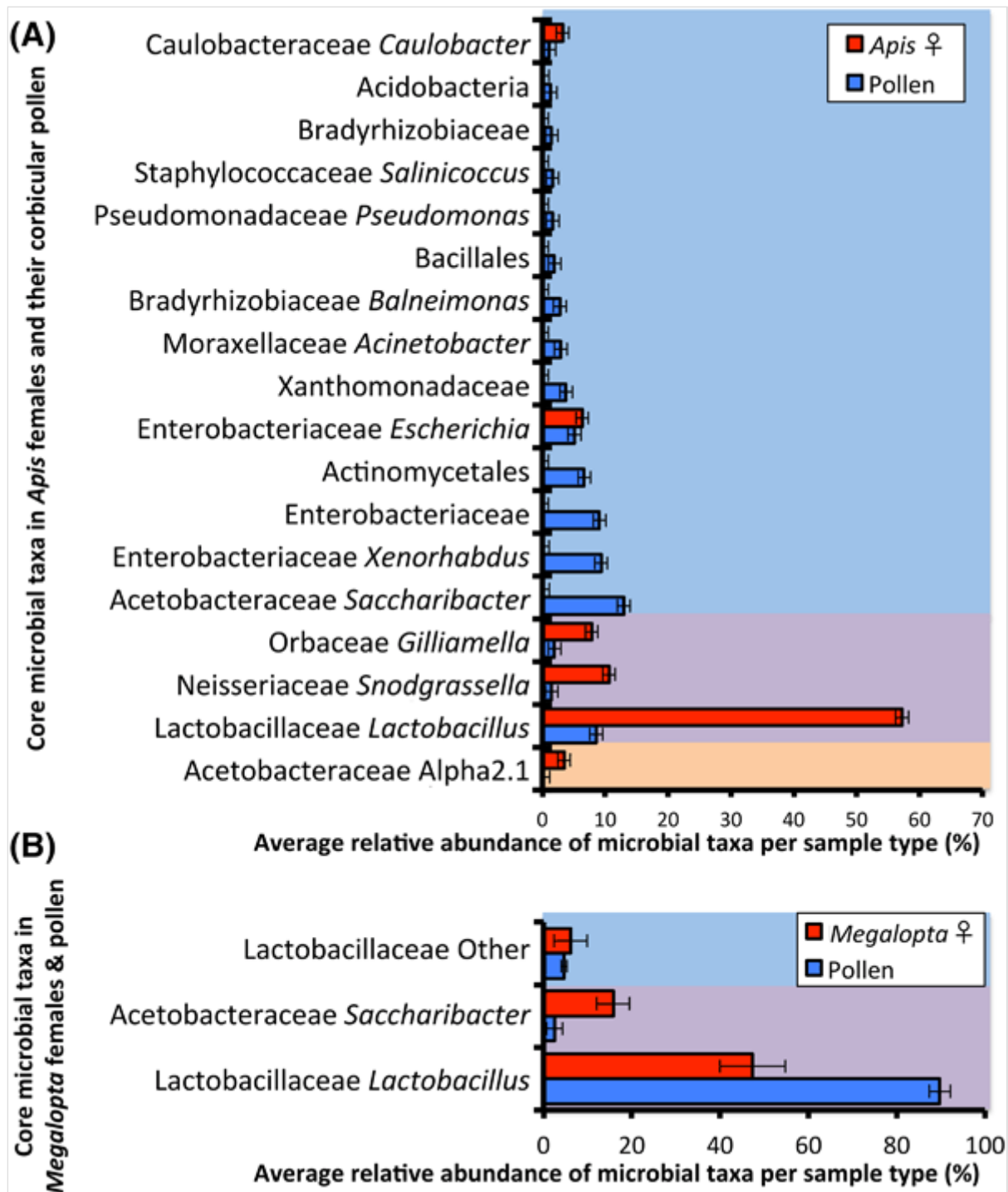


Fig. 2

Relative abundance of microbial taxa identified within samples of female *Megalopta* (a), and *Apis* bees (b), including their pollen provisions (corbicular pollen in *Apis*) as being either core microbes (>50% prevalence, >1% relative abundance) in *adult* females (orange panel), pollen (light blue panel) or in both (purple panel). The average relative abundances are shown for both female (red bars) plus their pollen provisions (blue bars) in all core taxa. Error bars represent standard error



Core pollen microbes

Overall 19 taxa were described as core bacteria in the *Ceratina* pollen samples (Fig. 1). *Megalopta* pollen and *Apis* corbicular pollen contained 3 and 17 core microbial groups respectively (Fig. 2). In addition to being core to all bees, *Lactobacillus* was also a core pollen microbial group found in the 3 pollen data sets. In both of the solitary bee datasets, *Lactobacillus* was relatively more abundant in the pollen, whereas in honey bees, it was most abundant in the bee gut (Figs. 1, 2).

Associations between adult *Ceratina* and pollen provisions

the microbial communities present in *Ceratina* females and their pollen provisions form three distinct clusters with a silhouette value of 0.08, suggesting overlap (Fig. 3, S1). *Ceratina* enterotype 1 contains most (66.6%) of the females and 42.5% of pollen samples. *Ceratina* enterotype 2 is a mixed group, representing only a small percentage of both female and pollen samples (6.6% and 15% respectively). *Ceratina* enterotype 3 is also a mixed cluster, containing the remaining 26.6% females and 42.5% of the pollen samples. All *Ceratina* female core and pollen core microbes were tested for relative abundance differences per enterotype. The relative abundance of four microbial taxa (*Methylobacterium*, *Sphingomonas*, *Lactobacillus*, and *Acinetobacter*) that are core groups to both bees and pollen differed significantly between *Ceratina* enterotypes 1–3 (KW = 14, d.f. = 2, $p < 0.001$; KW = 20, d.f. = 2, $p < 0.001$; KW = 65, d.f. = 2, $p < 0.001$; KW = 38, d.f. = 2, $p < 0.001$ respectively. Fig. 4; Table S1). In addition, the relative abundance of 1 core pollen microbe (Methylocystaceae) also differed significantly between the three enterotypes (KW = 15 d.f. = 2, $p = 0.005$; Table S1; Fig. 4). The remaining 21 microbial groups that had been identified as core microbes within *Ceratina* females and/or their pollen provisions did not differ between the enterotypes. Overall *Ceratina* enterotype 1 is characterized by relatively high *Sphingomonas*, Methylocystaceae, and *Methylobacterium*; *Ceratina* enterotype 2 has relatively high *Acinetobacter*; *Ceratina* enterotype 3 has relatively high *Lactobacillus*. The various *Ceratina* enterotypes were not specific to pollen provisions from different nests, but did tend to be in particular areas within nests ($\chi^2 = 30.9$, d.f. 22, $p = 0.09$; KW = 10.5, d.f. = 2, $p = 0.005$ respectively; Fig. 5). By comparing the position of pollen provision within the nest, we can determine its relative age within a nest (Fig. 5 a). Generally pollen provisions with a microbial community profile of *Ceratina* enterotype 1 were found in the freshest provisions (median cell position 4th), the oldest provisions were dominated by *Ceratina* enterotype 2 (median cell position 9.5), whilst provisions displaying a *Ceratina* enterotype 3 community had an intermediate cell position (median 6th) within the nest (Fig. 5 b). It should however be noted that most nests had 7 or fewer cells, therefore trends in the older cells may be biased by few nest replicates (Fig. 5 c).

Fig. 3

Clustering between microbial communities within datasets of female *Ceratina* (triangles) and their pollen provisions (circles). Enterotype 1 (green), Enterotype 2 (red) and Enterotype 3 (blue), were determined via Calinski–Harabasz and generated

using the Partitioning Around Mediod (PAM) method on Jensen-Shannon distance matrices. Silhouette = 0.08

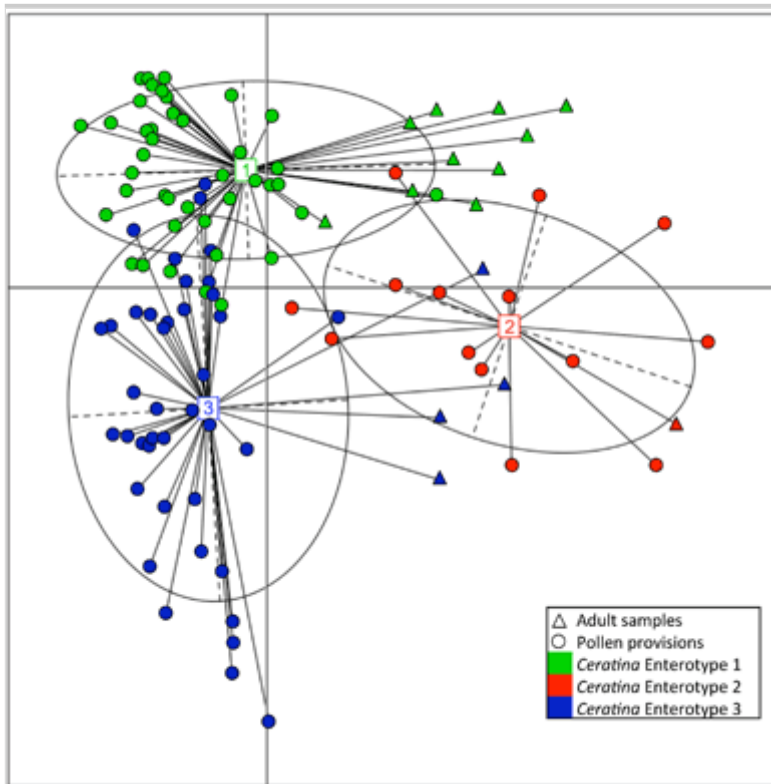


Fig. 4

Heatmap showing the abundance of the 26 core microbial phylotypes (rows) in *Ceratina* female and pollen samples. The presence of each phylotype is indicated by the presence of white to black shading where darker shading represents higher abundance as indicated by the scale bar and annotated in red text. The phylotypes are ordered by hierarchical clustering based on community dissimilarity. Cluster 1 contains most (66.6%) of the females and 42.5% of pollen samples. Cluster 2 is a mixed group, representing only a small percentage of both female and pollen samples (6.6 and 15% respectively). Cluster 3 is a mixed cluster, containing the remaining 26.6% of female and 42.5% of pollen samples. Significant differences (FDR corrected Kruskal–Wallis) in microbial abundances between the three community clusters are shown

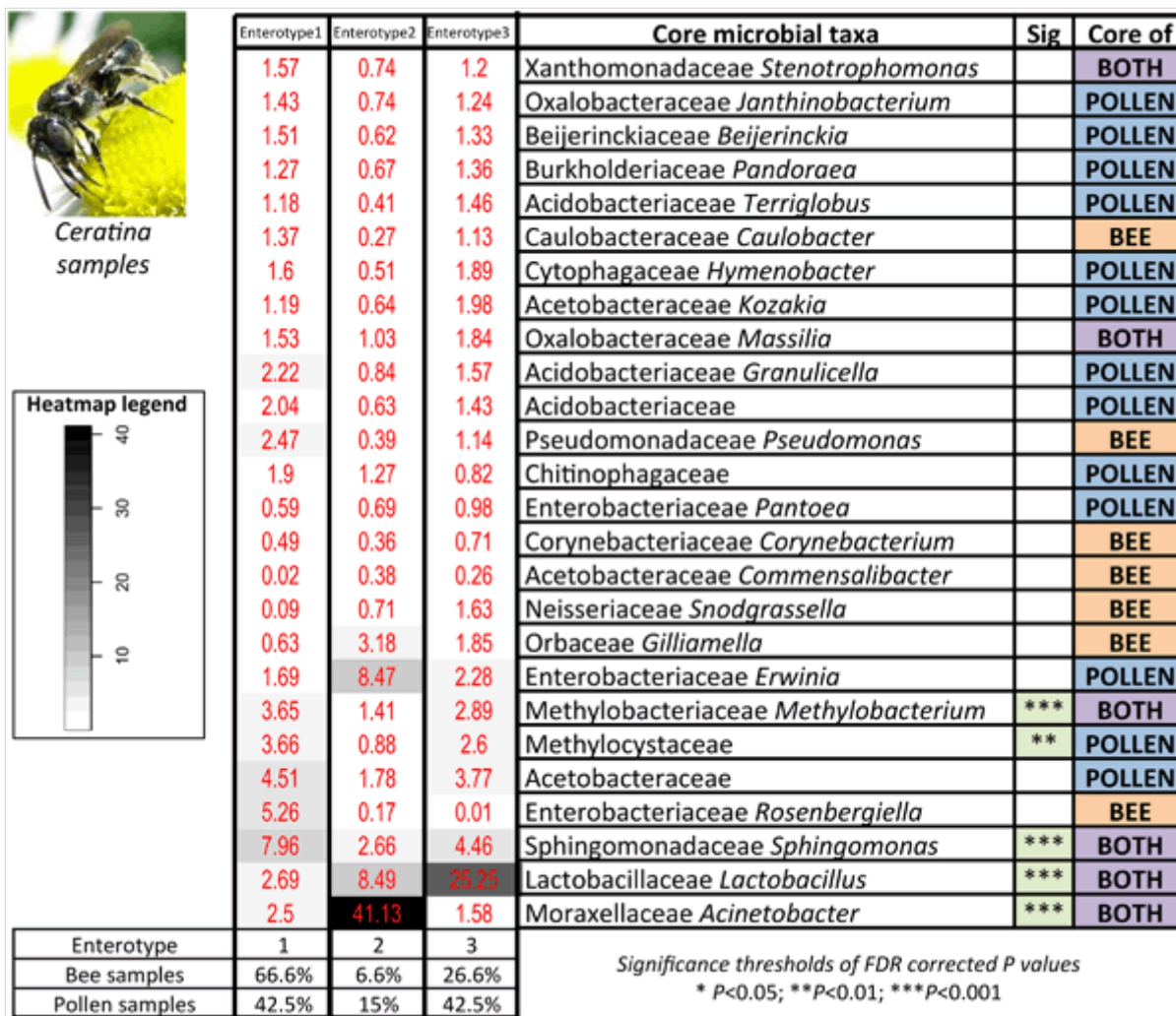
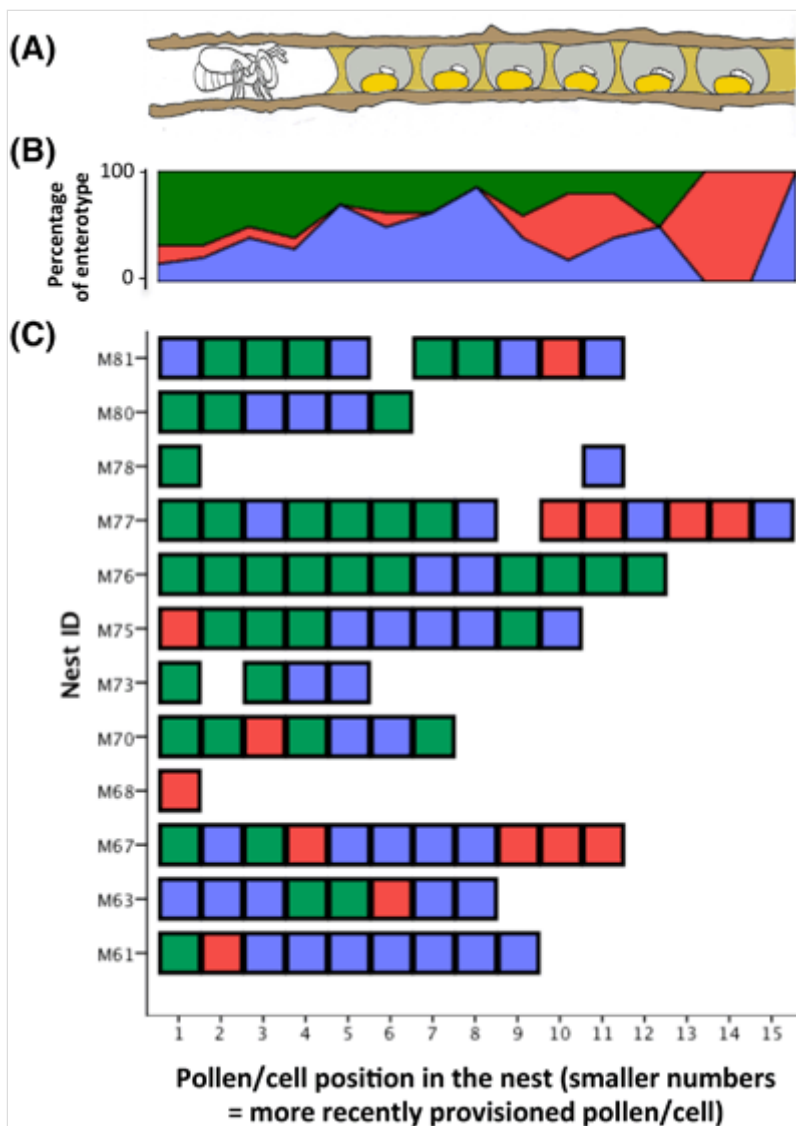


Fig. 5

Depiction of the pollen provision sequence of *Ceratina* females in their tubular nests (a), whereby the pollen provisions nearest the entrance are the freshest and those furthest away have been enclosed in a cell the longest. The proportion of pollen provisions exhibiting different microbial enterotypes differs over time (b); with *Ceratina* enterotype 1 (green) being immediately dominant before succeeding to *Ceratina* enterotype 3 (blue), then 2 (red). Nests have a mix of enterotypes, but few nests had more than 8 cells (c)

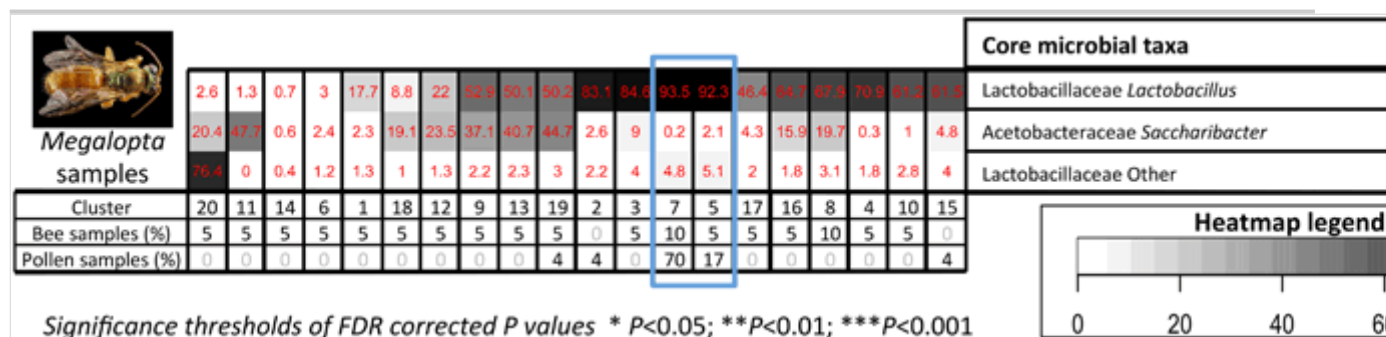


Associations between adult Megalopta and pollen provisions: although there are differences in the relative abundance of 3 core microbial groups, the microbial communities present in *Megalopta* females and their pollen provisions do not form distinctive enterotypes (sample $n = 43$, $K = 39$, Silhouette = 0.03; Fig. S2). As this K value is close to the sample number, we force the communities to assign within a K of 20 to see if any discreet clusters can be discerned. With 20 imposed *Megalopta* clusters, a silhouette value of 0.13 suggests overlap or (as in this case) low representation of each cluster (Figs. S3, S4). The microbial communities in 20 female *Megalopta*, split into 18 of the 20 different clusters. The 23 pollen samples group into five of the 20 clusters (three of these clusters also include a similar bee microbiome), with 70% of the pollen all in *Megalopta* cluster 7, suggesting pollen provisions tend to have a more similar microbial community than those of the *Megalopta* females (Fig. 6). Overall, *Lactobacillus* is core to both bees and pollen and is dominant throughout most of the samples. *Lactobacillaceae*_Other has a low dominance in pollen whilst *Saccharibacter*, despite being core to both pollen and

females, is more abundant in females. Only two of the clusters (7 and 5) contained more than three samples and so Mann–Whitney U tests on core microbes were only run between these two clusters. The abundance of *Saccharibacter* was significantly higher in cluster 5 than cluster 7 ($U = 0$, $p < 0.001$). The abundances of Lactobacillaceae *Lactobacillus* and Lactobacillaceae_Other were similar between Cluster 5 and 7 ($U = 61.5$, $p = 0.23$; $U = 30$, $p = 0.29$; Fig. 6; Table S2).

Fig. 6

Heatmap showing the abundance of the various core microbes in *Megalopta* females and pollen provisions within the 20 imposed distinct clusters of microbial communities. The presence of each phylotype is indicated by the presence of *white to black shading* where *darker shading* represents higher abundance as indicated by the *scale bar* and annotated in *red text*. The phylotypes are ordered by hierarchical clustering based on community dissimilarity. The microbial communities in the 20 female *Megalopta* split into 18 different ‘clusters’. The 23 pollen samples group into five clusters (three of these clusters also include a similar bee microbiome), with 70% of the pollen all in Cluster 7. Mann–Whitney U tests are only between Clusters 7 and 5 (outlined in *blue*) since these are the only ones with >3 samples



Associations between adult Apis bees and their corbicular pollen: the microbial communities present in *Apis* females and their corbicular pollen form two distinct clusters with a silhouette value of 0.37, suggesting minimal overlap between clusters (Fig. 7, S5). The sample composition in *Apis* cluster 1 comprises 90% of the female *Apis* samples along with 8% of the corbicular pollen samples and is therefore deemed to be a bee dominant cluster. The remaining 10% of bee samples group with 92% of the pollen samples in *Apis* cluster 2 (pollen dominant cluster). The *Apis* bee samples are a split of crop (foregut) extractions and gut samples—all gut samples are in Cluster 1 along with 71% of the crop samples. The bee samples present in Cluster 2 (pollen dominant cluster) are all crop samples. Each of the

eighteen core microbial taxa (17 core pollen, 4 core bee including 3 that are core to both) have significantly different abundances in each cluster (Fig. 8, Table S3).

Fig. 7

Clustering between microbial communities within *Apis* female guts (*circles*), *Apis* female crops (*triangles*) and their corbicular pollen (*diamonds*). Cluster 1 (*green*) and Cluster 2 (*red*) were determined via Calinski–Harabasz and generated using the Partitioning Around Mediod (PAM) method on Jensen–Shannon distance matrices; silhouette = 0.37

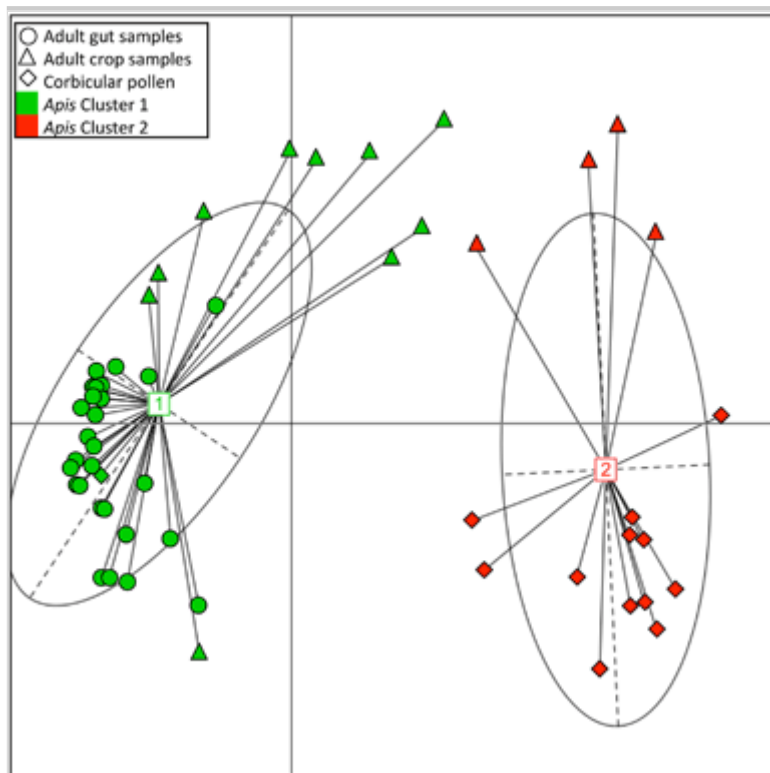
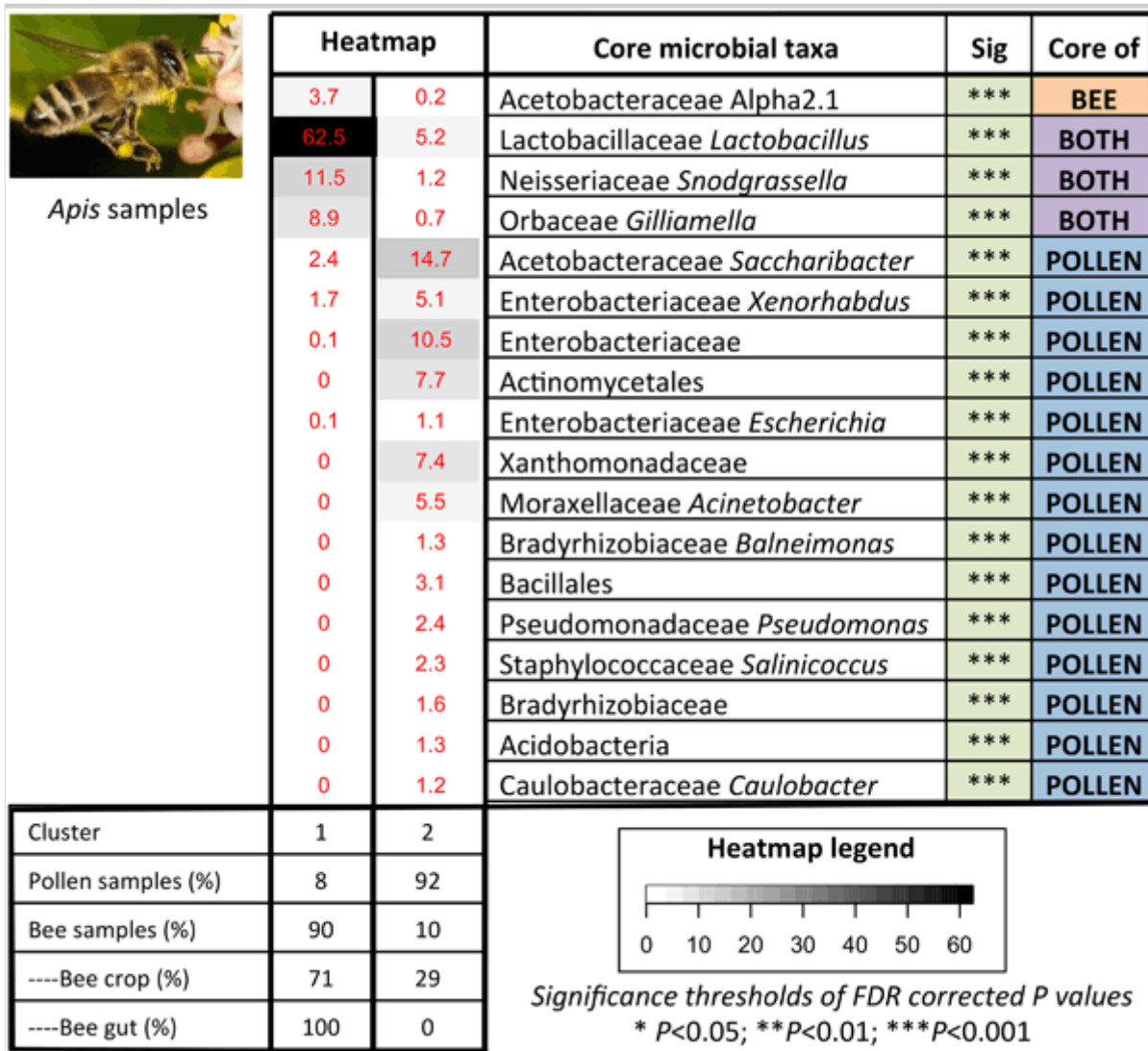


Fig. 8

Heatmap showing the abundance of the various core microbes in *Apis* females and their corbicular pollen within the two distinct clusters of microbial communities. The presence of each phylotype is indicated by the presence of *white to black shading* where *darker shading* represents higher abundance as indicated by the scale bar and annotated in *red text*. The phylotypes are ordered by hierarchical clustering based on community dissimilarity. The microbial communities in 90% of female *Apis* samples along with 8% of pollen samples separate into Cluster 1 (bee dominant cluster). The remaining 10% of bee samples group with 92% of the pollen samples in Cluster 2 (pollen dominant cluster). The *Apis* bee samples are a split of crop (foregut) extractions and gut samples—all gut samples are in Cluster 1 along with 71% of the

crop samples. The bee samples present in Cluster 2 (pollen dominant cluster) are all crop samples. Significant differences (Mann–Whitney U test) in microbial abundances between the two community clusters are shown



Discussion

The wild bees *Ceratina* and *Megalopta* both have core microbiota that are distinct from each other, and from model corbiculate bees such as *Apis* and *Bombus*. These core microbes may have the potential to play a role in the health of these wild bees. Certain microbial genera, however, are shared across host genera. For example, *Lactobacillus* was identified as core bacteria in *Ceratina*, *Megalopta* and *Apis* bees and was pervasive in pollen as well. In total *Ceratina* bees have thirteen core bacterial groups, three of which (*Lactobacillus*, *Snodgrassella* and *Gilliamella*) have previously been described as core bacteria in *Apis* bees. Two of these, *Gilliamella* and *Snodgrassella*, have been found to reduced infectivity of the common trypanosome parasite *Crithidia bombi* in bumble bees (Koch and Schmid-

Hempel 2011b). *Ceratina calcarata* bees ingest pollen before regurgitating it as pollen provisions (Michener 2007), so it is unsurprising that six of the adult *Ceratina* core microbial groups (*Lactobacillus*, *Acinetobacter*, *Sphingomonas*, *Methylobacterium*, *Massilia* and *Stenotrophomonas*) were also found to be core microbes of the pollen provisions. Despite overlap, three distinct enterotypes are identifiable in the female and pollen samples.

AQ2

Most of the *Ceratina* female samples and fresh pollen provisions display a *Ceratina* enterotype 1 community. This result is not surprising as it is expected that at the point of regurgitation, the female *Ceratina* bees' gut and the pollen provision would have a similar composition. As a different environment type, the static, high sugar environment of a pollen provision may promote the growth of different microbes than those that flourish inside the bee gut. We find some evidence of that here, with *Ceratina* enterotype 1 being largely found in adults and in freshly deposited pollen provisions. Older pollen provisions are then dominated by *Ceratina* enterotype 3 (which is characterized by high *Lactobacillus*), whereas the oldest pollen provisions tend to exhibit *Ceratina* enterotype 2 (dominated by *Acinetobacter*). Alternatively, or in addition to the microbial community in pollen provisions being driven by time spent in this static, brood cell environment, the differences may reflect a different cocktail of microbes being environmentally transmitted on different provisioning/foraging days. *Ceratina* pollen provisions had an additional 13 different core microbial groups, many of which have known associations with plants such as *Erwinia* and *Beijerinckia* (Starr and Chatterjee 1972; Becking 2006).

Megalopta bees have a core microbiome of two microbial groups, *Lactobaccillus* and *Saccharibacter*. Community clustering is not immediately apparent, yet *Lactobaccillus* and *Saccharibacter* are present in 90% of adults tested. In addition, both of these microbial groups are also core microbes found in their pollen provisions along with an undetermined genus of Lactobacillaceae, suggesting transmission between bee and pollen. Non-corbiculate bees appear to rely more on horizontal transmission from shared food and foraging surfaces with other bees (McFrederick et al. 2012, 2014).

Here we identify four core bacteria in *Apis* foragers. Previous research has also identified additional taxa including *Fischella perrara* as core *Apis* microbes, however this is rarely found in foraging bees like those sampled here (Corby-Harris et al. 2014a; Kwong and Moran 2016). Other microbes described previously as core honey bee microbes such as *Bartonella* and *Bifidobacterium* had a low relative

abundance (both 0.4%) and so were excluded by our core identification method (which excluded taxa <1% abundance) despite a high prevalence. This is likely an artifact of including crop samples as ‘adult bee’ samples, since these had a low abundance of several microbes, resulting in an overall lower average abundance. The microbial communities found in corbicular pollen of *Apis* clearly differentiated from those in honey bee gut with the exception of some crop samples, as would be expected for hosts that obtain their microbiota from their kin instead of food.

Overall the majority of *Ceratina* bees and pollen provisions share a similar microbial composition. This is in contrast to *Apis* and *Megalopta* that have relatively distinct adult core microbiomes compared to their pollen. In *Ceratina*, this large overlap between adults and pollen is likely due to the practice of ingesting and regurgitating their pollen. Despite this, both pollen and *Ceratina* adults have some distinct core microbes, suggesting the bees are maintaining a bee-specific microbiota beyond the crop. Whilst some of the core bee bacteria such as *Rosenbergiella*, *Lactobacillus*, and *Pseudomonas* have been found to inhabit flower environments, it is not clear if all core bee microbes are acquired from flowers or via a different route (McFrederick and Rehan 2016; McFrederick et al. 2016). *Megalopta* adults have a diverse range of microbial communities with only two core bacterial groups. The core microbial composition in *Apis* bees share similarities with that of *Ceratina* bees, yet despite clear community separation between adults and pollen, *Apis* adults share three core microbial groups (*Lactobacillus*, *Snodgrassella* and *Gilliamella*) with corbicular pollen. It should be noted, however, that we binned bacteria by genus, which does not allow us to differentiate between pollen- and bee-gut specialist species in the same genus, which are known to occur in honey bees (Corby-Harris et al. 2014b).

Honey bees are believed to maintain their core microbiome via interactions within the social colony environment (Powell et al. 2014). In the absence of this environment, *Ceratina* seem to be able to maintain a core microbiome even when the core microbes are not pervasive in the (flower) environment. By determining 13 microbes that are frequently associated with adult *Ceratina* bees, seven of which do not associate with their pollen provisions, we have identified good candidates to explore their potential functional importance to the health of these wild bees. The further identification of one of these microbial groups (*Lactobacillus*) in *Apis* and *Megalopta* strongly suggests that this may be a key microbial taxon in a diverse range bee species. *Lactobacillus* comprises species that appear to be specialists to the honey bee and bumble bee gut (Martinson et al. 2011; Olofsson et al. 2014) as

well as *L. micheneri* (McFrederick et al. 2016), which is found on both flowers and in association with a variety of wild bee species.

The microbiome of bees is likely to be critical to their health yet, for most wild bees, research of the honey bee microbiome can only provide limited insight (Engel et al. 2016). The use of a more diverse range of bees in research of bee health is therefore required in order to better understand the health and risks faced by declining bee species. As we learn more about the effects of these bacteria on the health of wild bees, an understanding of the composition of a healthy, core microbiome will become crucial for leveraging microbes to improve bee health. For example, pollinator habitat restoration efforts may need to consider flower plantings that increase the presence of core bacteria that are found in flowers, adults, and pollen provisions, such as *Lactobacillus* and *Saccharibacter*. Future work is needed to determine the role of these core bacteria in restoration of healthy pollinator communities.

Acknowledgements

We thank Sean Lombard and Nicholas Pizzi for assistance with nest collections, Krista Ciaccio and Wyatt Shell for nest processing, and Jason Rothman for DNA extractions and library preparation. Funding from the University of California Riverside to QSM, the New Hampshire Agricultural Experiment Station, Tuttle Research Foundation, and the University of New Hampshire to SMR supported this work. Media acknowledgements; *Ceratina calcarata* (in Fig. 4) photo by J.C. Lucier (CC BY-NC 2.0); *Megalopta genalis* (in Fig. 6) photo by Sam Droege (CC BY 2.0); *Apis mellifera* (in Fig. 8) photo by Gustavo Fotoopa (CC BY-NC-ND); *Ceratina* nest diagram (Fig. 5a) by Wyatt Shell.

Electronic supplementary material

Below is the link to the electronic supplementary material.

Supplementary material 1 (PDF 440 KB)

Supplementary material 2 (XLSX 63 KB)

References

Arumugam M, Raes J, Pelletier E et al (2011) Enterotypes of the human gut microbiome. *Nature* 473:174–180. doi: 10.1038/nature09944

Becking JH (2006) The genus *Beijerinckia*. In: *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, Springer, New York, pp 151–162

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300

Brysch-Herzberg M (2004) Ecology of yeasts in plant-bumblebee mutualism in Central Europe. *FEMS Microbiol Ecol* 50:87–100. doi: 10.1016/j.femsec.2004.06.003

Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat. Methods* 3:1–27. doi: 10.1080/03610917408548446

Caporaso JG, Bittinger K, Bushman FD et al (2010a) PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267. doi: 10.1093/bioinformatics/btp636

Caporaso JG, Kuczynski J, Stombaugh J et al (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. doi: 10.1038/nmeth.f.303

Cardinal S, Straka J, Danforth BN (2010) Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proc Natl Acad Sci U S A* 107:16207–16211. doi: 10.1073/pnas.1006299107

Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. *Nat Rev Genet* 13:260–270. doi: 10.1038/nrg3182

Colman DR, Toolson EC, Takacs-Vesbach CD (2012) Do diet and taxonomy influence insect gut bacterial communities? *Mol Ecol* 21:5124–5137. doi: 10.1111/j.1365-294X.2012.05752.x

Corby-Harris V, Maes P, Anderson KE (2014a) The bacterial communities associated with honey bee (*Apis mellifera*) foragers. *PLoS ONE*. doi: 10.1371/journal.pone.0095056

Corby-Harris V, Snyder L a, Schwan MR et al (2014b) Origin and effect of Acetobacteraceae Alpha 2.2 in honey bee larvae and description of *Parasaccharibacter apium*, gen. nov., sp. nov. *Appl Environ Microbiol*. doi: 10.1128/AEM.02043-14

DeSantis TZ, Hugenholtz P, Larsen N et al (2006) Greengenes, a chimera-checked 16 S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072. doi: 10.1128/AEM.03006-05

Engel P, James RR, Koga R et al (2013) Standard methods for research on *Apis mellifera* gut symbionts. *J Apic Res* 52:1–24. doi: 10.3896/IBRA.1.52.4.07

Engel P, Kwong WK, McFrederick QS, et al (2016) The bee microbiome: impact on bee health and model for evolution and ecology of host-microbe interactions. *MBio* 7:1–9. doi: 10.1128/mBio.02164-15

Gallai N, Salles JM, Settele J, Vaissiere BE (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol Econ* 68:810–821. doi: 10.1016/j.ecolecon.2008.06.014

Graystock P, Goulson D, Hughes WOH (2015) Parasites in bloom: flowers aid dispersal and transmission of pollinator parasites within and between bee species. *Proc R Soc B Biol Sci* 282:20151371. doi: 10.1098/rspb.2015.1371

Hamady M, Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* 19:1141–1152

Kaufman MG, Klug MJ (1991) The contribution of hindgut bacteria to dietary carbohydrate utilization by crickets (Orthoptera: Gryllidae). *Comp Biochem Physiol—Part A Physiol* 98:117–123. doi: 10.1016/0300-9629(91)90588-4

Koch H, Schmid-Hempel P (2011a) Bacterial communities in central European bumblebees: low diversity and high specificity. *Microb Ecol* 62:121–133. doi: 10.1007/s00248-011-9854-3

Koch H, Schmid-Hempel P (2011b) Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci USA* 108:19288–19292. doi: 10.1073/pnas.1110474108

- Koch H, Abrol DP, Li J, Schmid-Hempel P (2013) Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Mol Ecol* 22:2028–2044. doi: 10.1111/mec.12209
- Kong HH, Oh J, Deming C et al (2012) Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res* 22:850–859. doi: 10.1101/gr.131029.111
- Kwong WK, Moran NA (2016) Gut microbial communities of social bees. *Nat Rev Microbiol* 14:374–384. doi: 10.1038/nrmicro.2016.43
- Li J, Powell JE, Guo J et al (2015) Two gut community enterotypes recur in diverse bumblebee species. *Curr Biol* 25:R652–R653. doi: 10.1016/j.cub.2015.06.031
- Martins AC, Melo GAR, Renner SS (2014) The corbiculate bees arose from New World oil-collecting bees: implications for the origin of pollen baskets. *Mol Phylogenet Evol* 80:88–94. doi: 10.1016/j.ympev.2014.07.003
- Martinson VG, Danforth BN, Minckley RL et al (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* 20:619–628. doi: 10.1111/j.1365-294X.2010.04959.x
- McFrederick QS, Rehan SM (2016) Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. *Mol Ecol* 25:2302–2311. doi: 10.1111/mec.13608
- McFrederick QS, Wcislo WT, Taylor DR et al (2012) Environment or kin: whence do bees obtain acidophilic bacteria? *Mol Ecol* 21:1754–1768. doi: 10.1111/j.1365-294X.2012.05496.x
- McFrederick QS, Wcislo WT, Hout MC, Mueller UG (2014) Host species and developmental stage, but not host social structure, affects bacterial community structure in socially polymorphic bees. *FEMS Microbiol Ecol* 88:398–406. doi: 10.1111/1574-6941.12302
- McFrederick QS, Thomas JM, Neff JL et al (2016) Flowers and Wild Megachilid Bees Share Microbes. *Microb Ecol*. doi: 10.1007/s00248-016-0838-1

Mercier C, Boyer F, Bonin A, Coissac E (2013) SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. Abstr In: SeqBio 25–26th Nov 2013 27. doi: 10.1002/ejoc.201200111

Mesquite Project Team (2014) Mesquite: A modular system for evolutionary analysis. In: Available from <http://mesquiteproject.wikispaces.com/home>.

Michener CD (2007) Bees of the World. Johns Hopkins University Press, Baltimore

Olofsson TC, Alsterfjord M, Nilson B et al (2014) *Lactobacillus apinorum* sp. nov., *Lactobacillus mellifer* sp. nov., *Lactobacillus mellis* sp. nov., *Lactobacillus melliventris* sp. nov., *Lactobacillus kimbladii* sp. nov., *Lactobacillus helsingborgensis* sp. n. Int J Syst Evol Microbiol 64:3109–3119. doi: 10.1099/ijms.0.059600-0

Powell JE, Martinson VG, Urban-Mead K, Moran N a (2014) Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera*. Appl Environ Microbiol 80:7378–7387. doi: 10.1128/AEM.01861-14

Price MN, Dehal PS, Arkin AP (2009) Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641–1650. doi: 10.1093/molbev/msp077

Rehan SM, Richards MH (2010) Nesting biology and subsociality in *Ceratina calcarata* (Hymenoptera: Apidae). Can Entomol 142:65–74. doi: 10.4039/n09-056

Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ (2006) Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J Math Model Algorithms 5:475–504. doi: 10.1007/s10852-005-9022-1

Rognes T, Flouri T, Nichols B, et al (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ Prepr 4:e2409v1. doi: 10.7287/peerj.preprints.2409v1

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. doi: 10.1016/0377-0427(87)90125-7

Russell J a, Moreau CS, Goldman-Huertas B et al (2009) Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc Natl Acad Sci U S A* 106:21236–21241. doi: 10.1073/pnas.0907926106

Shade A, Handelsman J (2012) Beyond the Venn diagram: The hunt for a core microbiome. *Environ Microbiol* 14:4–12. doi: 10.1111/j.1462-2920.2011.02585.x

Sharon G, Segal D, Ringo JM, et al (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 107:20051–20056. doi: 10.1073/pnas.1009906107

Starr MP, Chatterjee AK (1972) The genus *Erwinia*: enterobacteria pathogenic to plants and animals. *Annu Rev Microbiol* 26:389–426. doi: 10.1146/annurev.mi.26.100172.002133

Turnbaugh PJ, Turnbaugh PJ, Ley RE et al (2007) The human microbiome project. *Nature* 449:804–810. doi: 10.1038/nature06244

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the New bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. doi: 10.1128/AEM.00062-07

Woodard SH, Lozier JD, Goulson D et al (2015) Molecular tools and bumble bees: revealing hidden details of ecology and evolution in a model system. *Mol Ecol* 24:2916–2936. doi: 10.1111/mec.13198

Yun JH, Roh SW, Whon TW et al (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl Environ Microbiol* 80:5254–5264. doi: 10.1128/AEM.01226-14