Received Date : 19-Nov-2015
Revised Date : 26-Feb-2016
Accepted Date : 02-Mar-2016
Article type : Original Article

Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee

Quinn S. McFrederick[1] and Sandra M. Rehan[2]

[1] Department of Entomology, University of California Riverside, 900 University Avenue, Riverside, CA 92521; email Quinn.mcfrederick@ucr.edu

[2] Department of Biological Sciences, University of New Hampshire, 46 College Road, Durham, NH 03824; email: sandra.rehan@unh.edu

Abstract

Many insects obtain gut microbes from their diet, but how a mother's foraging patterns influences the microbes found in her offspring's food remains an open question. To address this gap, we studied a bee that forages for pollen from multiple species of plants and may therefore acquire diverse bacteria from different plants. We tested the hypothesis that pollen diversity correlates with bacterial diversity by simultaneously characterizing these two communities in bee brood provisions for the first time. We used deep sequencing of the plant RBCL gene and the bacterial 16S rRNA gene to characterize pollen and bacterial diversity. We then tested for associations between pollen and bacterial species richness and community composition, as well

as co-occurrence of specific bacteria and pollen types. We found that both pollen and bacterial communities were extremely diverse, indicating that mother bees visit a wide variety of flowers for pollen and nectar and subsequently bring a diversity of microbes back into their nests. Pollen and bacterial species richness and community composition, however, were not correlated. Certain pollen types significantly co-occurred with the most proportionally abundant bacteria, indicating that the plants these pollen types came from may serve as reservoirs for these bacteria. Even so, the overall diversity of these communities appears to mask these associations at a broader scale. Further study of these pollen and bacteria associations will be important for understanding the complicated relationship between bacteria and wild bees.

Introduction

The diet of an insect and its gut microbiome are tightly linked. The guts of many insects harbor environmental microbes that appear to be acquired from the insect's food (Engel & Moran 2013). For example, sweat bees, a fungus-farming ant, and two species of fire ants all associate with lactobacilli that can also be found in their food (McFrederick *et al.* 2013). Fruit fly gut microbiomes are influenced more by diet than by host phylogeny or geography (Chandler *et al.* 2011). The conditions in the gut select a subset of the microbes found in food, but for many insects food microbiome composition influences gut microbiome composition (Engel & Moran 2013). In contrast, the guts of some insects have been found to harbor bacteria not found in the environment or the insect's food. For example, honey bees and bumble bees associate with host-specific microbes (Kwong & Moran 2015). These host-specific bacteria may influence host nutrition, as they contain genes involved in carbohydrate digestion (Engel *et al.* 2012). Whether

gut microbes are environmentally acquired or host specific, they are integral to understanding of bee nutrition and health.

Bacteria found in bee food dominate wild and solitary bee gut microbiomes. Outside the corbiculate apids (honey, bumble, and stingless bees), the other bee species studied to date (mostly halictid and megachilid bees) largely associate with environmental bacteria (Martinson *et al.* 2011; Keller *et al.* 2013), especially bacteria that have either been found on flowers or have relatives that are found on flowers (McFrederick *et al.* 2012; 2014a, b). In sweat bees in the genus *Megalopta,* these bacteria appear to be transmitted at flowers, and introduced into bee nests when the mother or workers place pollen and nectar into brood cells (= sealed chamber within which a single juvenile bee develops) (McFrederick *et al.* 2014b). The next generation of bees contain these bacteria in the larval gut, lose them during metamorphosis, but then appear to regain them as they begin foraging on flowers as adults (McFrederick *et al.* 2014b). As female bees carry nectar (and in some cases pollen) bound for the brood provision in the crop (Michener 1974), which is the first section of the gut, it is perhaps not surprising that the brood provisions harbor the same bacteria as adult bees. Many brood provisions are dominated by an undescribed *Lactobacillus* species that is closely related to *L. kunkeei* and is part of the WCFS *Lactobacillus* clade (hereafter referred to as *L.* sp. aff. *kunkeei*), while all provisions studied to date harbor a diversity of microbes (McFrederick *et al.* 2012; 2014b).

Given that adult bees pick up microbes from flowers and further transmit these bacteria to the next generation via brood provisions, an unanswered question is how the foraging behavior of a mother bee affects the microbiome of the food with which she provisions her offspring. Oligolectic bees, i.e. bees that collect pollen from a diversity of flowers, vary in the diversity of

pollen that constitutes an individual provision. *Megalopta* bees often use a single pollen source for one provision (Smith *et al.* 2012), while *Osmia lignaria* produces mixed provisions (Williams & Tepedino 2003). Different microhabitats in flowers can harbor distinct microbiomes (Junker & Keller 2015), and flower microbiomes show distinct phenological shifts as apple flowers go from bud to open flowers (Shade *et al.* 2013). The brood provision microbiome may reflect the microbiomes of the flowers from whence that provision came.

Here we test the hypothesis that the diversity of pollen in a provision predicts the diversity of that provision's microbiome. We asked two related questions: (A) Is there a relationship between community diversity measurements of pollen and bacteria in brood provisions of a pollen-generalist bee? and (B) Do certain pollen types associate with certain bacterial types? To answer these questions, we used deep sequencing of partial 16S rRNA bacterial genes and partial ribulose biphosphate carboxylase large chain (RBCL) plant genes from brood provisions of the solitary, oligolectic bee *Ceratina calcarata.* To our knowledge, this is the first time that pollen and bacterial communities in bee brood provisions have been simultaneously characterized.

Materials and Methods

Study organisms

Ceratina calcarata is a stem nesting bee endemic to eastern North America (Rehan & Sheffield 2011; Shell & Rehan 2016). Both males and females of this species overwinter as adults in their natal nest and in spring individuals disperse and mate (Rehan & Richards 2010a). Females establish new nests solitarily by excavating dead, broken stems. *Ceratina calcarata* are nest loyal and produce a single nest in their one-year life span (Rehan & Richards 2013). After

hollowing out of the pithy core, females forage for nectar and pollen to provide a mass provision prior to laying an egg. This species is a generalist pollinator observed foraging on a wide variety of floral hosts (reviewed in Kennedy *et al.* 2013). Females provision and lay a single cell per day (Rehan & Richards 2010b). Following mass provisioning and oviposition, females scrape the inner walls of the nest and form a pith partition to cap the brood cell. This is repeated in a serial manner and females provision on average eight brood cells (range 1-14 brood cells; Rehan & Richards 2010b).

We collected 12 *C. calcarata* nests along roadsides in Durham, New Hampshire (43.1339° N, 70.9264° W) in July 2014. The number of brood cells averaged 8 per nest, with a range of 1-12 brood cells. All brood cells contained eggs or first instar larvae, meaning that we used only complete or nearly complete pollen provisions in our study. Nests were collected from dead broken stems of staghorn sumac, *Rhus typhinia,* and brought the nests back to the lab for processing. To collect brood provisions, we split stems longitudinally and used flame sterilized tools to transfer pollen balls to cryovials and immediately stored them in a -80 freezer until we conducted DNA extraction.

DNA extractions and deep sequencing of pollen and bacterial barcoding genes

To thoroughly lyse pollen and bacterial cells, we used DNA extraction methods recommended for recalcitrant gram-positive bacteria (Engel *et al.* 2013). Using the DNeasy Blood and tissue kit (Qiagen, Valencia, CA), we first added 180 µL of buffer ATL, a sterile 5 mm stainless steel bead, and approximately 100 µL of 0.1 mm glass beads to each sample. We then used a Qiagen tissue lyser to bead beat each sample for 3 minutes at 30 hz. We added 20 µL

of Proteinase K to each sample, incubated the samples at 57 °C overnight, and then followed the DNeasy standard extraction protocol.

We used a dual-index inline barcoding approach to prepare libraries for sequencing on the Illumina MiSeq. First, we designed primers that included either the forward or reverse Illumina sequencing primer, an eight nucleotide long barcode, and the forward or reverse genomic oligonucleotide, as in Kembel *et al.* (2014). For bacterial 16S rRNA gene sequence, we used 799F-mod3 CMGGATTAGATACCCKGG (Hanshew *et al.* 2013) and modified 1115R AGGGTTGCGCTCGTTG (Kembel *et al.* 2014), as both of these primers have been shown to minimize plastid contamination. For pollen metabarcoding, we used RBCL7 CTCCTGAMTAYGAAACCAAAGA and RBCL8 GTAGCAGCGCCCTTTGTAAC, which amplifies 180-220 bases of the plant ribulose bisphosphate carboxylase large chain gene. These unpublished primers were developed for next-generation sequencing of fossil pollen. See Table S1 for the complete primer set. Although it has been suggested that chloroplasts are missing in pollen (Willerslev *et al.* 2003), recent pollen metabarcoding studies have successfully used plastid markers such as RBCL (Hawkins *et al.* 2015; Richardson *et al.* 2015a). In one study, rank abundance of RBCL reads significantly correlated with microscopy-based estimates of abundance in five out of six samples, suggesting that RBCL may provide better relative abundance estimates compared to ribosomal loci (Richardson *et al.* 2015a).

To generate amplicons for Illumina sequencing, we used the above primers to perform triplicate PCRs for each gene. We used 10 μL ultrapure water, 10 μL HotMasterMix (5 Prime, Gaithersburg, MD), 0.5 μL each of 10 μM primer stock, and 4 μL of DNA. For both 16S rRNA and RBCL genes, we used a 52 °C annealing temperature, 35 cycles, and negative controls for each replicate reaction. To remove unincorporated primers and dNTPs, we first combined the

triplicate reactions and then used the Ultraclean PCR clean up kit (MoBio, Carlsbad, CA). We used 1 μL of the clean PCR product as template for a second PCR, using HPLC purified primers to complete the Illumina sequencing construct as in (Kembel *et al.* 2014): CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGC and AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG. For these reactions, we used a 58 °C annealing temperature, 35 cycles, and negative controls. To normalize the amount of PCR product from each reaction, we used 18 μL of these reactions and SequalPrep Normalization plates (ThermoFisher Scientific, Waltham, MA). We pooled 5 μL of each of the normalized samples, performed a second clean up, and then assessed the quality of our libraries using a 2100 Bioanalyzer (Agilent, Santa Clara, CA). After quality control, we sequenced the libraries using the MiSeq Reagent Kit v3 with 600 cycles. Raw data are available on the NCBI Sequence Read Archive (SRA) under accession number SRP066483.

Bioinformatics

For initial quality control and demultiplexing, we processed both 16S rRNA gene and RBCL gene sequences using MacQIIME 1.9.1 (Caporaso *et al.* 2010b). We used USEARCH 8 (v8.1.1831) for OTU clustering, assignment of reads to OTUs, open and reference based chimera checking for 16S rRNA gene sequences and open chimera checking for RBCL sequences, and removal of OTUs with fewer than four sequences (Edgar 2010). To bin sequences into OTUs, we used 97% sequence identity for bacteria and 99% sequence identity for plants. We then picked representative sequences from each OTU cluster, at which point our 16S rRNA and RBCL gene analyses diverged.

For 16S rRNA gene data, we performed standard alpha and beta diversity analyses in QIIME. We used the Greengenes taxonomy (McDonald *et al.* 2012) and the RDP Naïve Bayesian Classifier (Wang *et al.* 2007) to assign taxonomy to the OTUs. As training set can influence these taxonomic assignments (Newton & Roeselers 2012), we also performed local BLASTn searches against NCBI's Nucleotide Collection (nt) database (accessed June 9, 2015). We used this taxonomy to identify 891 OTUs that were assigned to mitochondria or chloroplast, which we then removed from our dataset. To build a phylogeny, we first aligned the quality-filtered dataset using the pynast aligner (Caporaso *et al.* 2010a) and the Greengenes database (McDonald *et al.* 2012). We then refined the alignment by eye in Mesquite version 3.04 (Maddison & Maddison 2015), and reconstructed the phylogeny of the bacterial OTUs using Fast Tree version 2.1.3 (Price *et al.* 2010). We used this phylogeny as input for both weighted and unweighted UniFrac analyses (Hamady & Lozupone 2009). Using these distance matrices, we performed Non-metric Multi-Dimensional Scaling (NMDS), and plotted the resulting ordination in R (R Core Development Team 2015). For alpha diversity, we plotted rarefaction curves in QIIME, and used gplots (Warnes *et al.* 2015) to create a heatmap of the most abundant OTUs, arranged by hierarchical clustering.

To assign taxonomy to the RBCL gene OTUs, we used local BLAST searches (Altschul *et al.* 1990) and MEGAN5 (Huson & Mitra 2012). As described above, we used BLASTn to search for matches, saving the top 100 matches with e-values less than 1e-80. To assign representative sequences from the RBCL OTUs to taxonomy, we used the subsequent BLAST output to perform lowest common ancestor (LCA) analysis in MEGAN5. LCA assignment uses the set of BLAST hits to assign sequences to their lowest possible node in the NCBI taxonomy. We imported this taxonomy into the QIIME OTU table. To align the RBCL gene data, we used

MUSCLE version 3.8.31 (Edgar 2004), and refined the alignment by eye in Mesquite (Maddison & Maddison). We then continued with standard alpha and beta diversity analyses, as outlined above for bacterial 16S rRNA gene sequences.

We additionally binned the bacterial and pollen OTUs at different levels. To explore strain diversity in bacteria, we used USEARCH8 as described above to bin sequences at 99% sequence identity. We also binned bacteria at the genus level using the summarize taxa QIIME script. For RBCL, we manually curated OTUs into genus-level bins based on the before-mentioned BLAST results. Best BLAST hits were verified as confirmed genera occurring in New Hampshire using the USDA Plants Database (http://plants.usda.gov/checklist.html). For the genus-level classifications, we used QIIME to calculate Bray-Curtis community distance matrices instead of UniFrac distance matrices, as selecting reference sequences for tree-building was more ambiguous compared to the sequence-similarity based OTUs. We ran beta-diversity analyses on the full OTU tables as well as 'core OTU' tables, which we generated by paring down the tables to only OTUs that occurred in greater than 80% of our samples.

To compare pollen and bacterial distance matrices, we performed mantel tests on the UniFrac distance matrices (both full and 'core' matrices) using the ADE4 package (Dray & Dufour 2007) in R. As abundance artifacts introduced by primer or other PCR biases could confound these analyses, we used both weighted (which includes abundances) and unweighted (which is based on presence/absence only) UniFrac distance matrices. As additional tests for associations between bacteria and pollen communities, we also ran Procrustes analyses in QIIME on the first three principal coordinate axes of pollen and bacterial ordinations. Procrustes analyses tests goodness of fit by rotating and scaling the configuration of ordination points (Gower 1975). We used cor.test in R to conduct Kendall's correlations between plant OTU

proportional abundance and bacterial OTU proportional abundance, order in which a brood cell was produced (which is an indicator of how old the brood cell is), and proportional abundance of dominant OTUs and axes from our NMDS analyses.

To explore co-occurrence of OTUs, we used the program SparCC, which uses the absolute abundances of OTUs to identify correlations (Friedman & Alm 2012). To test for co-occurrence using rarified data, we used CoNet, which combines p-values from a series of tests and adjust for multiple comparisons (Faust *et al.* 2012). For SparCC, we used 1000 bootstrap replicates to calculate significance values, and considered correlation coefficients greater or less than 0.2 and -0.2 respectively and p-values less than 0.001 (i.e. lower then occurred in any of our bootstrap replicates). We also excluded OTUs found at fewer than two reads per sample on average, as in Friedman & Alm (2012). For CoNet, we rarefied all samples to 1175 reads (the smallest bacterial read depth with enough coverage as determined by rarefaction curves) and used Mutual Information, Spearman, Pearson, Bray Curtis and Kullback-Leibler dissimilarity. We used Brown's method to combine p-values and the Benjamini-Hochberg correction for multiple comparisons.

Results

From 96 brood provisions, we obtained 5,972,645 quality-filtered reads for the RBCL gene and 581,175 quality-filtered reads for 16S rRNA gene amplicons. For the RBCL gene, we obtained an average of 63,544 and a standard deviation of 15,128 reads per sample, while we obtained an average of 6,183 and standard deviation of 2462 reads per sample for the 16S rRNA gene. Two samples failed to sequence for the RBCL gene while two different samples failed for the 16S rRNA gene. As the two failed samples from each community did not overlap, we used

92 samples for analyses in which we compared plant and bacterial communities. Rarefaction curves indicated that we were able to accurately characterize plant and bacterial communities in most samples (Fig. S1).

Ceratina calcarata brood provisions contained pollen from 110 genera of plants, but eight of these genera accounted for 94% of the RBCL gene reads. On average, each provision contained RBCL reads from 34 genera of plants, with a minimum of 16 and a maximum of 61 genera. *Rhamnus*, *Rhus*, and *Rubus* were the most abundant pollen types (Fig. 1). While some provisions represented a mix of pollen types, a single pollen type dominated other provisions (Fig. 1, Table S2). Both well mixed provisions and provisions dominated by a single pollen type were present within a single nest (Fig. S2, Table S2). These variable patterns were also present in NMDS ordination analyses of brood provisions, in which provisions did not cluster by age of the cell (Fig. 2) or by nest (Fig. 3). Proportional abundances of *Rhamnus* (Kendall $\tau = 0.42$, $p = 1.776e\text{-}09$) and *Rhus* (Kendall $\tau = 0.519$, $p = 2.20e\text{-}13$) pollen were positively correlated with NMDS axis 1 while *Rubus* (Kendall $\tau = -0.576$, $p = 4.47e\text{-}13$) was negatively correlated with NMDS axis 1 (Figs. 2 & 3). *Rhamnus* was positively correlated with NMDS axis two (Kendall $\tau = 0.494$, $p = 3.06e\text{-}12$) while *Rhus* was negatively correlated with NMDS axis two (Kendall $\tau = -0.371$, $p = 1.62e\text{-}7$).

Across all samples we found 1270 bacterial OTUs using 97% sequence identity. Bacterial communities were diverse and included many rare OTUs, for example no OTUs accounted for more than 10% of all reads, and only 15 OTUs accounted for more than 1% of all reads. The most abundant OTU shared 99% sequence identity with *Lactobacillus* sp. aff. *kunkeei*, which we have previously reported associated with multiple halictid and megachilid bees and their brood provisions (McFrederick *et al.* 2012; 2014a, b). The endosymbionts *Wolbachia* and *Sodalis* were

also found in many brood provisions (Fig. 1). Bacterial communities, like pollen communities, did not cluster by cell age or by nest (Figs. 2 & 3).

Mantel tests between pollen and bacteria community distance matrices indicated that the two communities are not correlated. This was true whether we ran the tests with Bray-Curtis community distance matrices of plant genera and 97% sequence identical bacterial OTUs ($p = 0.323$), weighted UniFrac distance matrices of 99% sequence identical plant and bacterial OTUs ($p = 0.300$) or weighted UniFrac distance matrices of 97% sequence identical plant and bacterial OTUs ($p = 0.157$). This pattern also held with unweighted UniFrac distance matrices of 97% sequence identical plant and bacterial OTUs ($p = 0.253$). Mantel tests on weighted ($p=0.101$) and unweighted ($p = 0.254$) 'core OTU' distance matrices also failed to detect significant correlations between the two matrices. Procrustes analyses of the full ($M^2 = 0.96$, $p = 0.127$) and core OTU tables (weighted UniFrac $M^2 = 0.97$, $p = 0.339$, unweighted UniFrac $M^2 = 0.95$, $p = 0.068$) further suggested that the two matrices are not related. Likewise, we found no significant correlation between the number of plant genera in a brood provision and the number of bacterial 97% sequence identity OTUs (Kendall $\tau = 0.019$, $p = 0.787$) or 99% sequence identity plant and bacterial OTUs (Kendall $\tau = -0.030$. $p = 0.673$). There was a weak but significant negative correlation between *L.* sp. aff. *kunkeei* proportional abundance and the order in which a brood cell was built (Fig. S3, Kendall $\tau = -0.179$, $p = 0.013$).

SparCC co-occurrence analysis suggested that several plant genera significantly co-occur with some focal bacteria. For example, *L.* sp. aff. *kunkeei* was found to positively correlate with *Celastrus*, *Gleditsia, Rosa* and *Sanguisorba* (Table 1). *Rubus* was the only plant that we found to significantly correlate with two bacterial OTUs: *Acinetobacter* and *Sodalis*. While the CoNet analysis failed to detect all of the same associations (Fig. S4), several positive associations were

significant in both analyses: *Acinetobacter* and *Rubus*, *Acinetobacter* and *Prunus*, and *Lactobacillus* sp. aff. *kunkeei* and *Sanguisorba*.

Discussion

The diversity of pollen found in a *C. calcarata* brood provision does not predict the diversity of bacteria found within that same provision. When a mother bee visits many different flowers to collect the nectar and pollen with which she provisions her young, bacterial communities in that brood provision are no more species rich than when she visits fewer flowers. Additionally, pollen community metrics do not correlate with bacterial community metrics. The foraging decisions that a mother makes, however, may still influence her brood's microbiome. We found that certain pollen types significantly co-occur with certain bacteria, meaning that pollen source may predict whether a certain bacteria is found in a brood provision. Three associations between pollen and bacterial types were detected in two independent analyses, suggesting that these relationships are likely important. These correlations, however, may be driven by many possible factors, and empirical confirmation of these associations from floral hosts is needed.

The lack of associations between pollen and bacterial communities may be caused by the surprising diversity of plant genera found in brood provisions, which likely reflects both pollen and nectar sources. Honey contains enough pollen for metabarcoding and morphological study (Hawkins *et al.* 2015), suggesting that incidental pollen may be carried in nectar by other bees besides *Apis mellifera*. An alternative explanation may relate to the internal carriage of pollen by *Ceratina* adults (Michener 2007). Processing of the pollen and selection for certain microbes

may occur before the pollen is deposited inside the nest, erasing signatures of variation in flower microbiomes.

Recent studies have also used a metabarcoding approach to identify pollen in bee provisions (Richardson *et al.* 2015b; Keller *et al.* 2015; Sickel *et al.* 2015) and honey (Hawkins *et al.* 2015). For example, Sickel et al. (2015), found that two *Osmia* species also have extremely diverse brood provisions, with pollen from up to 85 plant species found in one provision. As expected, an oligolectic (pollen specialist) bee, *Osmia truncorum,* collected mostly Asteraceae pollen while a polylectic (pollen generalist) bee, *O. bicornis,* foraged on more diverse pollen types (Sickel *et al.* 2015). Much like other studies on polylectic bees (Keller *et al.* 2015; Sickel *et al.* 2015), we found that *C. calcarata* brood provisions can be extremely diverse.

Our pollen barcoding data revealed interesting foraging patterns by the bees. As *C. calcarata* nests are provisioned serially, the position of the brood cell indicates the relative age of the cell (Rehan & Richards 2010b). This serial provisioning allows us to examine how a mother's foraging patterns change across the provisioning cycle of a nest. For example, the oldest six brood cells in nest M67 contained 50-90% *Rubus* pollen, but the following four brood cells contained 50-90% *Rhamnus* pollen, indicating that the mother switched between these plants in the middle of the nest provisioning process. Nest M80, in contrast, had more diverse provisions, with only one provision containing more than 50% pollen from a single plant genus. There appears to be a large amount of variation in foraging strategies in *C. calcarata*. Coupling next-generation sequencing barcoding approaches to surveys of resource availability in nearby communities may prove to be a particularly powerful approach for investigating foraging behavior in bees and other central-place foragers. Bumble bees and honey bees avoid consuming artificial nectar containing bacteria, suggesting that bacteria can influence foraging patterns

(Junker *et al.* 2014; Good *et al.* 2014). Simultaneous sequencing of bacteria and pollen in bee provisions and flower microbiomes, coupled with surveys of floral availability, would provide insight into the effects of bacteria on flower visitation rates.

Like pollen, the bacterial communities in *C. calcarata* brood provisions are surprisingly diverse. Many of the bacteria were assigned to genera that are commonly found closely associated with plants, such as *Gluconoacetobacter, Erwinia* and *Rhizobium* (Gnanamanickam 2007; Junker *et al.* 2011). Much of this diversity may be incidental and likely related to the diversity of flowers and other environments that the mother bees visit. Common flower- and/or bee-associated bacteria such as *Acinetobacter nectaris* (Alvarez-Perez *et al.* 2013), *Saccharibacter floricola* (Jojima *et al.* 2004), and *Fructobacillus fructosus* (Endo & Okada 2008) were also present, confirming the ubiquity of these bacteria in the pollination landscape across North America. Other notable bacteria included a close relative of the honey-bee specific *Gilliamella apicola* and an OTU with 97% sequence identity to *Snodgrasella alvi,* one of the honey bee 'core' worker gut bacteria (Kwong & Moran 2013).

The flower- and wild bee-associated *L.* sp. aff. *kunkeei* (McFrederick *et al.* 2012; 2014a, b) was the most proportionally abundant bacterium across all brood provisions. *Lactobacillus* sp. aff. *kunkeei* significantly co-occurred with two pollen types in two independent analyses. These plant genera may either serve as particularly suitable reservoirs for *L.* sp. aff. *kunkeei* or perhaps these plants provide resources that promote *L.* sp. aff. *kunkeei* fitness in the brood provision.

Lactobacillus sp. aff. *kunkeei* was weakly correlated with brood cell age, suggesting that as the brood provisions age *L.* sp. aff. *kunkeei* increases in proportional abundance. These proportional data do not allow us to determine if this increase is due to competitive effects, or simply because *L.* sp. aff. *kunkeei* is able to persist in the brood cell environment while other

bacteria are not. We are currently conducting competitive assays between *L.* sp. aff. *kunkeei* and spoilage fungi as well as nutritional experiments to try to untangle the role of *L.* sp. aff. *kunkeei* in the brood cell environment.

Two insect-associated endosymbiotic bacteria, *Wolbachia* and *Sodalis*, were surprisingly common in *C. calcarata* brood provisions. The presence of endosymbionts outside the host is not entirely unprecedented, as *Wolbachia* has been previously reported in the foregut and feces of leafcutting ants in the genus *Acromyrmex* (Andersen *et al.* 2012). *Ceratina calcarata* has reduced body hairs and carries nectar and pollen back to the nest within the crop (Michener 2007), which may account for the presence of these endosymbiotic bacteria in the brood provisions. If so, transmission from mother to offspring via brood provisions may represent a secondary route of vertical transmission for these bacteria. This also opens up the possibility that *Wolbachia* in the bee foregut may be deposited on flowers and transferred between species. *Wolbachia*, however, has not yet been reported in flower microbiomes (Fridman *et al.* 2011; Shade *et al.* 2013; Junker & Keller 2015).

In summary, we found that pollen composition does not correlate with microbiome composition in *C. calcarata* brood provisions. Certain plant species, however, appear to predict the abundance of common bee- and flower-associated bacteria in the brood provision, suggesting that these plants may serve as particularly suitable reservoirs for these bacteria. If these bacteria are important in bee fitness, investigation of specific plants as means for either increasing or decreasing their availability to wild bee pollinators may be a promising avenue for applied research. Our research also suggests a novel route of maternal inheritance for endosymbiotic bacteria, i.e. via brood provisions. Whether this transmission route is found only in animals that carry food destined for their brood in the gut or orally merits further investigation.

Acknowledgements:

We thank Sean Lombard and Nicholas Pizzi for assistance with nest collections, Krista Ciaccio and Wyatt Shell for brood cell processing, Kaleigh Russell for help with DNA extractions and library preparation, Glenn Hicks of UCR's IIGB for advice on primer design, and Jason Rothman, Peter Graystock, Hoang Vuong, Kaleigh Russell, Alexander Keller, and an anonymous reviewer for comments on an earlier draft. 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.

References cited

Altschul SFS, Gish WW, Miller WW, Myers EWE, Lipman DJD (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Alvarez-Perez S, Lievens B, Jacquemyn H, Herrera CM (2013) *Acinetobacter nectaris* sp. nov. and *Acinetobacter boissieri* sp. nov., isolated from floral nectar of wild Mediterranean insect-pollinated plants. *International Journal of Systematic and Evolutionary Microbiology*, **63**, 1532–1539.

Andersen SB, Boye M, Nash DR, Boomsma JJ (2012) Dynamic *Wolbachia* prevalence in *Acromyrmex* leaf-cutting ants: potential for a nutritional symbiosis. *Journal of Evolutionary Biology*, **25**, 1340–1350.

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

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chandler JA, Morgan Lang J, Bhatnagar S, Eisen JA, Kopp A (2011) Bacterial communities of diverse *Drosophila* species: Ecological context of a host–microbe model system. *PLoS Genetics*, **7**, e1002272.

Dray S, Dufour A-B (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Endo A, Okada S (2008) Reclassification of the genus *Leuconostoc* and proposals of *Fructobacillus fructosus* gen. nov., comb. nov., *Fructobacillus durionis* comb. nov., *Fructobacillus ficulneus* comb. nov. and *Fructobacillus pseudoficulneus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, **58**, 2195–2205.

Engel P, Moran NA (2013) The gut microbiota of insects - diversity in structure and function. *FEMS Microbiology Reviews*, **37**, 699–735.

Engel P, James RR, Koga R *et al.* (2013) Standard methods for research on *Apis mellifera* gut symbionts. *Journal of Apicultural Research*, **52**, 1–24.

Engel P, Martinson VG, Moran NA (2012) Functional diversity within the simple gut microbiota of the honey bee. *Proceedings Of The National Academy Of Sciences*, **109**, 11002–11007.

Faust K, Sathirapongsasuti JF, Izard J *et al.* (2012) Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology*, **8**, e1002606.

Fridman S, Izhaki I, Gerchman Y, Halpern M (2011) Bacterial communities in floral nectar. *Environmental Microbiology Reports*, **4**, 97–104.

Friedman J, Alm EJ (2012) Inferring correlation networks from genomic survey data. *PLoS Computational Biology*, **8**, e1002687.

Gnanamanickam SS (2007) *Plant-associated bacteria*. Springer, Dordrecht, The Netherlands.

Good AP, Gauthier M-PL, Vannette RL, Fukami T (2014) Honey bees avoid nectar colonized by three bacterial species, but not by a yeast species, isolated from the bee gut. *PLoS ONE*, **9**, e86494.

Gower JC (1975) Generalized Procrustes analysis. *Psychometrika*, **40**, 33–51.

Hamady M, Lozupone C (2009) Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *The ISME Journal*, **4**, 17–27.

Hanshew AS, Mason CJ, Raffa KF, Currie CR (2013) Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *Journal of Microbiological Methods*, **95**, 149–155.

Hawkins J, de Vere N, Griffith A *et al.* (2015) Using DNA metabarcoding to identify the floral composition of honey: a new tool for Investigating honey bee foraging preference. *PLoS ONE*, **10**, e0134735.

Huson DH, Mitra S (2012) Introduction to the analysis of environmental sequences: metagenomics with MEGAN. *Methods in Molecular Biology*, **856**, 415–429.

Jojima Y, Mihara Y, Suzuki S (2004) *Saccharibacter floricola* gen. nov., sp. nov., a novel osmophilic acetic acid bacterium isolated from pollen. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 2263–2267.

Junker RR, Keller A (2015) Microhabitat heterogeneity across leaves and flower organs promotes bacterial diversity. *FEMS Microbiology Ecology*, **91**, fiv097.

Junker RR, Loewel C, Gross R *et al.* (2011) Composition of epiphytic bacterial communities differs on petals and leaves. *Plant biology*, **13**, 918–924.

Junker RR, Romeike T, Keller A, Langen D (2014) Density-dependent negative responses by bumblebees to bacteria isolated from flowers. *Apidologie*, **45**, 467–477.

Keller A, Danner N, Grimmer G *et al.* (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology*, **17**, 558–566.

Keller A, Grimmer G, Steffan-Dewenter I (2013) Diverse microbiota Identified in whole intact nest chambers of the red mason bee *Osmia bicornis* (Linnaeus 1758). *PLoS ONE*, **8**, e78296.

Kembel SW, O'Connor TK, Arnold HK *et al.* (2014) Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings Of The National Academy Of Sciences*, **111**, 13715–13720.

Kennedy CM, Lonsdorf E, Neel MC *et al.* (2013) A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. *Ecology Letters*, **16**, 584–599.

Kwong WK, Moran NA (2013) Cultivation and characterization of the gut symbionts of honey bees and bumble bees: *Snodgrassella alvi* gen. nov., sp. nov., a member of the Neisseriaceae family of the Betaproteobacteria; and *Gilliamella apicola* gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the Enterobacteriales order of the Gammaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology*, **63**, 2008–2018.

Kwong WK, Moran NA (2015) Evolution of host specialization in gut microbes: the bee gut as a model. *Gut Microbes*, **6**, 214–220.

Maddison W, Maddison DR Mesquite: a modular system for evolutionary analysis.

Martinson VG, Danforth BN, Minckley RL *et al.* (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, **20**, 619–628.

McDonald D, Price MN, Goodrich J *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, **6**, 610–618.

McFrederick QS, Cannone JJ, Gutell RR *et al.* (2013) Specificity between lactobacilli and hymenopteran hosts is the exception rather than the rule. *Applied and Environmental Microbiology*, **79**, 1803–1812.

McFrederick QS, Mueller UG, James RR (2014a) Interactions between fungi and bacteria influence microbial community structure in the *Megachile rotundata* larval gut. *Proceedings Of The Royal Society B-Biological Sciences*, **281**, 20132653.

McFrederick QS, Wcislo WT, Hout MC, Mueller UG (2014b) Host species and developmental stage, but not host social structure, affects bacterial community structure in socially polymorphic bees. *FEMS Microbiology Ecology*, **88**, 398–406.

McFrederick QS, Wcislo WT, Taylor DR *et al.* (2012) Environment or kin: whence do bees obtain acidophilic bacteria? *Molecular Ecology*, **21**, 1754–1768.

Michener CD (1974) *The Social Behavior of the Bees*. Harvard University Press, Cambridge, Massachusetts.

Michener CD (2007) *The Bees of the World*. John Hopkins University Press, Baltimore, Maryland.

Newton ILG, Roeselers G (2012) The effect of training set on the classification of honey bee gut microbiota using the Naïve Bayesian Classifier. *BMC Microbiology*, **12**, 221.

Price MNM, Dehal PSP, Arkin APA (2010) FastTree 2--approximately maximum-likelihood

trees for large alignments. *PLoS ONE*, **5**, e9490–e9490.

R Core Development Team (2015) A language and environment for statistical computing.

Rehan SM, Richards MH (2010a) Nesting biology and subsociality in *Ceratina calcarata* (Hymenoptera: Apidae). *Canadian Entomologist*, **142**, 65–74.

Rehan SM, Richards MH (2010b) The Influence of maternal quality on brood sex allocation in the small carpenter bee, *Ceratina calcarata*. *Ethology*, **116**, 876–887.

Rehan SM, Richards MH (2013) Reproductive aggression and nestmate recognition in a subsocial bee. *Animal Behaviour*, **85**, 733–741.

Rehan SM, Sheffield CS (2011) Morphological and molecular delineation of a new species in the *Ceratina dupla* species-group (Hymenoptera: Apidae: Xylocopinae) of eastern North America. *Zootaxa*, 35–50.

Richardson RT, Lin C-H, Quijia JO *et al.* (2015a) Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in Plant Sciences*, **3**, 1500043–10.

Richardson RT, Lin C-H, Sponsler DB *et al.* (2015b) Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences*, **3**, 1400066.

Shade A, McManus PS, Handelsman J (2013) Unexpected diversity during community succession in the apple flower microbiome. *mBio*, **4**, e00602–12–e00602–12.

Shell WA, Rehan SM (2016) Recent and rapid diversification of the small carpenter bees in eastern North America. *Biological Journal of the Linnean Society*, **117**, 633–645.

Sickel W, Ankenbrand MJ, Grimmer G *et al.* (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, **15**, 20.

Smith AR, López Quintero IJ, Moreno Patiño JE, Roubik DW, Wcislo WT (2012) Pollen use by *Megalopta* sweat bees in relation to resource availability in a tropical forest. *Ecological Entomology*, **37**, 309–317.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian Classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Warnes GR, Bolker B, Bonebakker L *et al.* (2015) gplots: Various R programming tool for plotting data.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

Williams NM, Tepedino VJ (2003) Consistent mixing of near and distant resources in foraging bouts by the solitary mason bee Osmia lignaria. *Behavioral Ecology*, **14**, 141–149.

Data accessibility: Metabarcoding amplicon data and associated metadata are available on the NCBI Sequence Read Archive (SRA) under accession number SRP066483.

Author contributions: SMR collected the samples, QSM and SMR designed the experiment, QSM performed the laboratory work, and QSM and SMR analyzed the data and wrote the manuscript.

Table 1. SparCC correlation coefficients for co-occurrence of abundant bacteria and plant pollen sources. All associations were significant at $p < 0.01$. SparrCC assesses all possible pairwise comparisons; non-significant and negative correlations are not reported.

	Celastrus	*Franklinia*	*Gleditsia*	*Lamium*	*Prunus*	*Rosa*	*Rubus*	*Sanguisorba*	*Vicia*
Acinetobacter		0.226		0.251	0.345		0.287		
Lactobacillus	0.267		0.231			0.225		0.366	
Pantoea									0.236
Sodalis							0.249		

Figure 1. Heatmap of the top-six plants and bacteria from each sample. Each column is a single brood provision, while row represents pollen type or bacterial OTU, in descending order of overall proportional abundance within plants or bacteria. The dendrogram groups samples on hierarchical clustering based on community similarity across both plants and bacteria.

Figure 2. Non-metric Multi Dimensional Scaling (NMDS) analysis of Bray-Curtis distance matrices representing plant and bacterial communities from each brood provision. Each sample is color-coded as brood cell, with one being the first cell provisioned in the nest. The plant communities are based on the binned plant genera dataset, while the bacterial communities are from the 97% sequence identity dataset.

Figure 3. Non-metric Multi Dimensional Scaling (NMDS) analysis of Bray-Curtis distance matrices representing plant and bacterial communities from each brood provision. Each sample

is color-coded by nest. The plant communities are based on the binned plant genera dataset, while the bacterial communities are from the 97% sequence identity dataset.

Supplemental figure captions:

Fig. S1. Rarefaction curves of plant OTUs binned by genera (Fig. A), plant OTUs binned at 99% sequence identity (Fig. B), and bacterial OTUs binned at 97% sequence identity (Fig. C).

Fig. S2. Heatmap of the top six plants and bacteria from each sample. Each column is a single brood provision, while row represents pollen type or bacterial OTU, in descending order of overall proportional abundance within plants or bacteria. Samples are arranged by nest, with oldest brood cell on the left. Vertical lines separate samples from different nests.

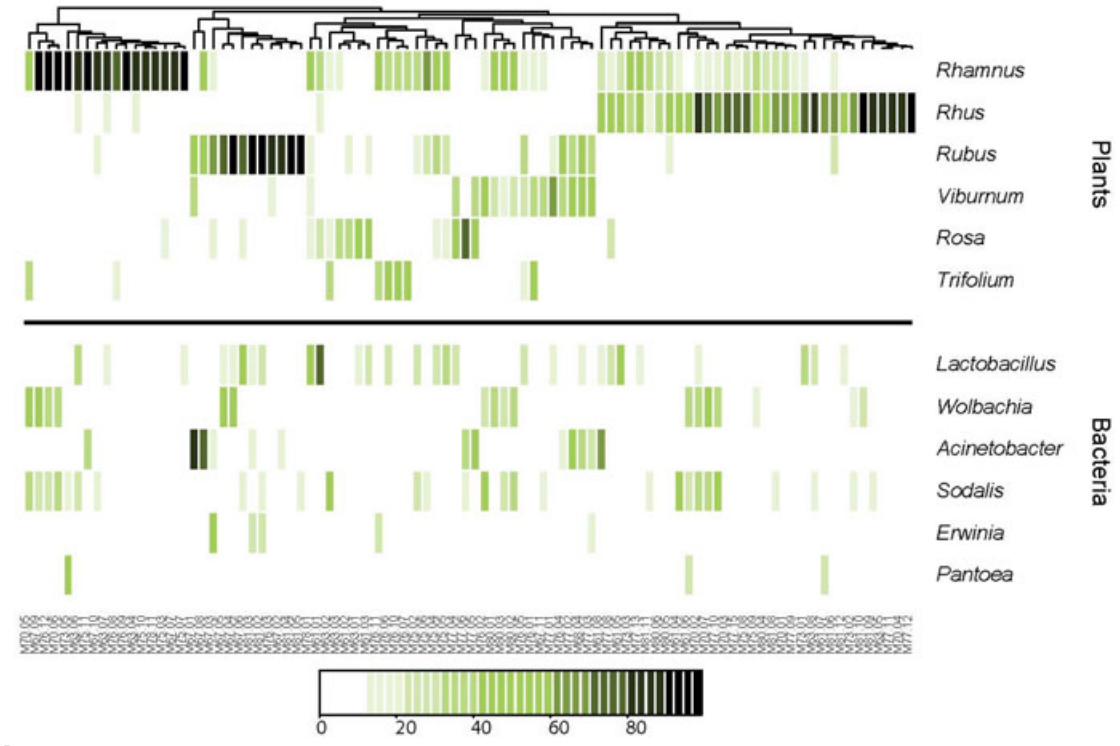
Fig. S3. Scatterplot of *L.* sp. aff. *kunkeei* against placement of brood cell in the nest, where the first brood cell is the oldest in the nest. *Lactobacillus* sp. aff. *kunkeei* proportional abundance shows a weak, but significant relationship with brood cell age (Kendall $\tau = -0.179$, $p = 0.013$), suggesting that *L.* sp. aff. *kunkeei* increases in proportional abundance as the brood cell ages.

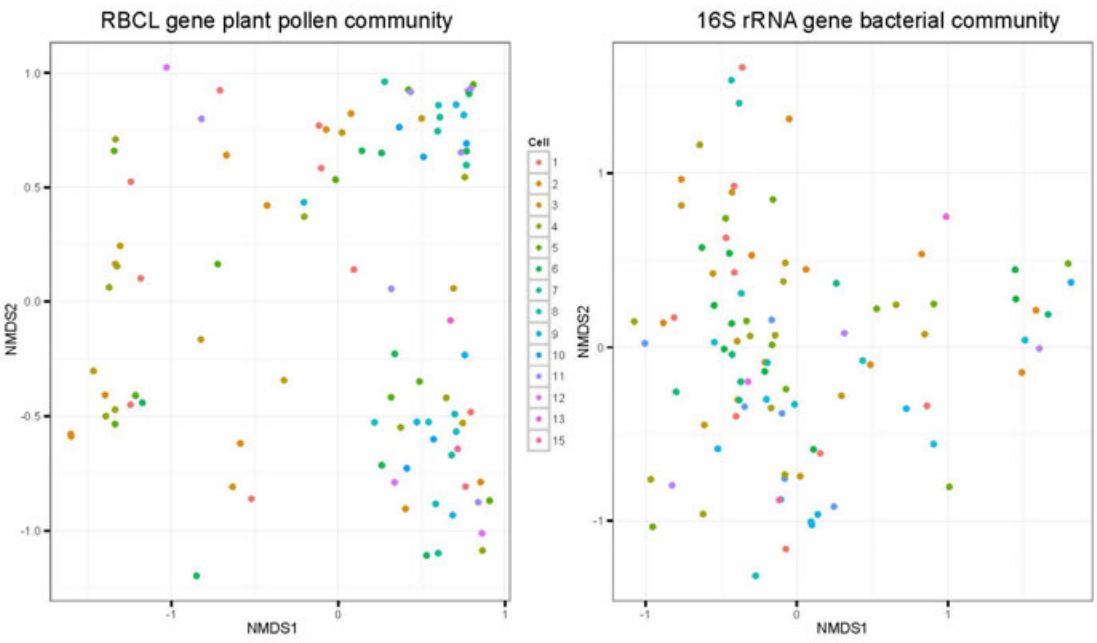
Fig. S4. Co-occurrence network of positive correlations between plant and bacterial OTUs, as determined by CoNet analysis. Only significant correlations are depicted.

Supplemental Table captions:

Table S1. 16S rRNA gene and RBCL gene primers used in this study.

Table S2. OTU table of 16S rRNA gene OTUs binned at 97% sequence identity and RBCL gene OTUs binned by genera.





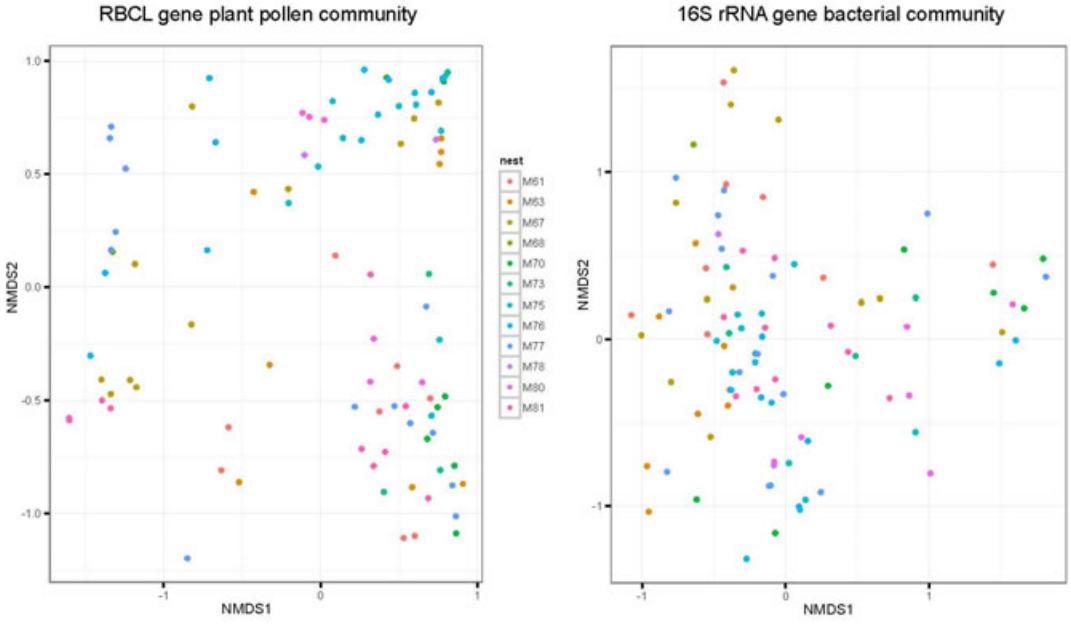


Fig. S1A

RBCL Plant Genus

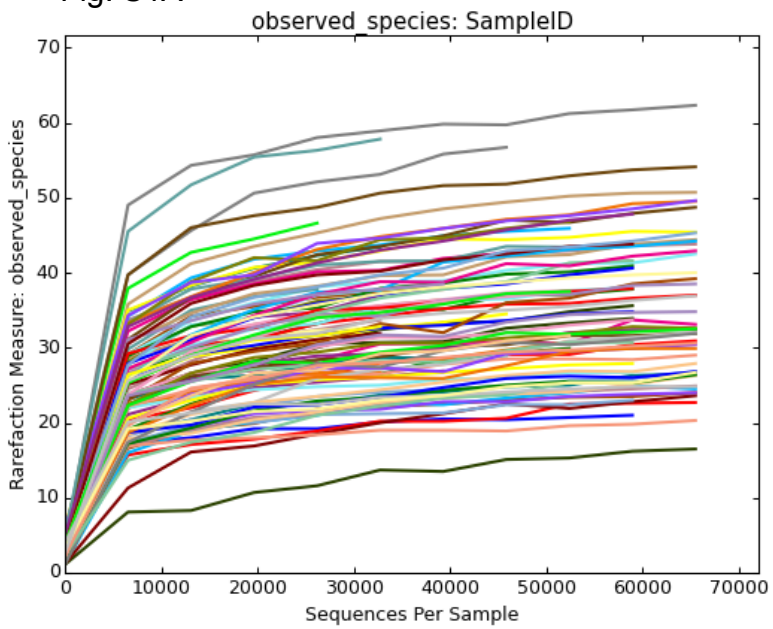


Fig. S1B

RBCL Plant 99% sequence identity

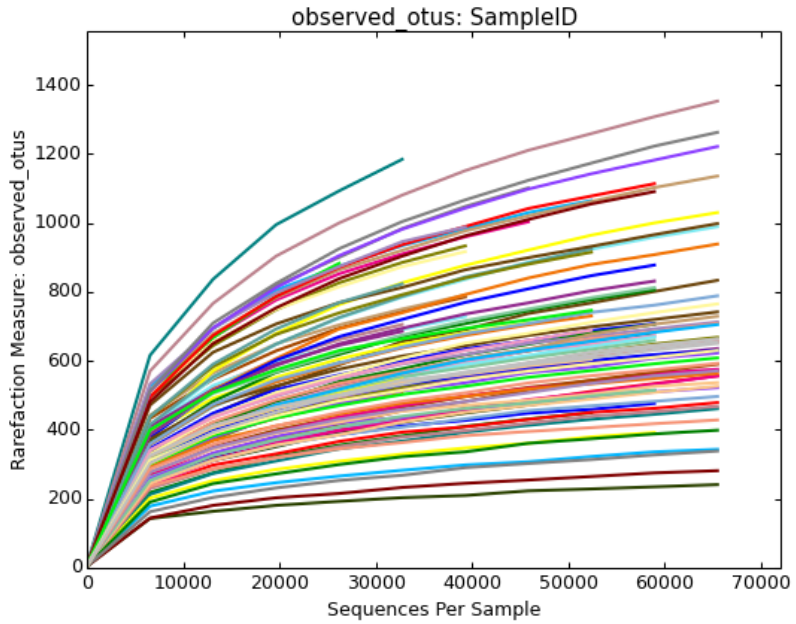


Fig. S1C

Bacterial 97% sequence identity OTUs

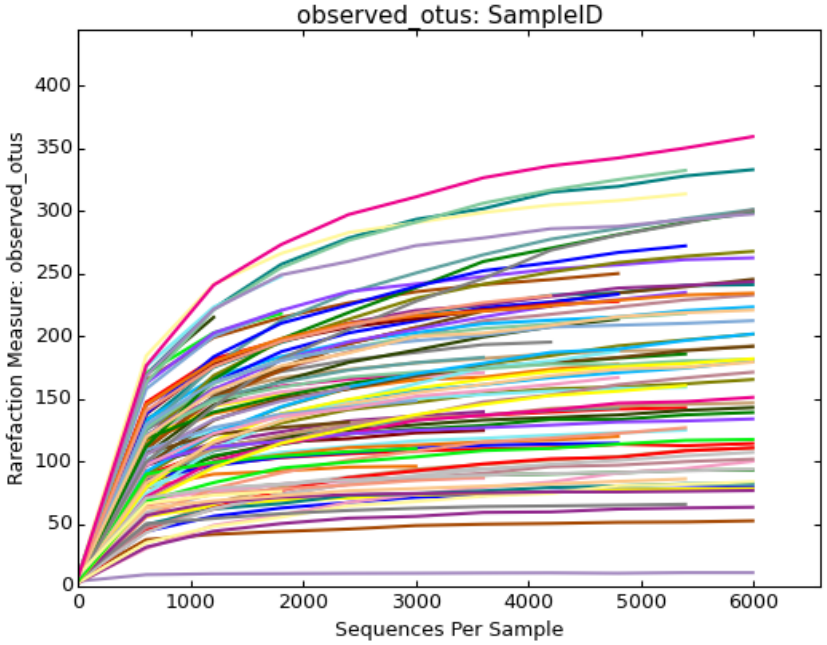


Fig. S2

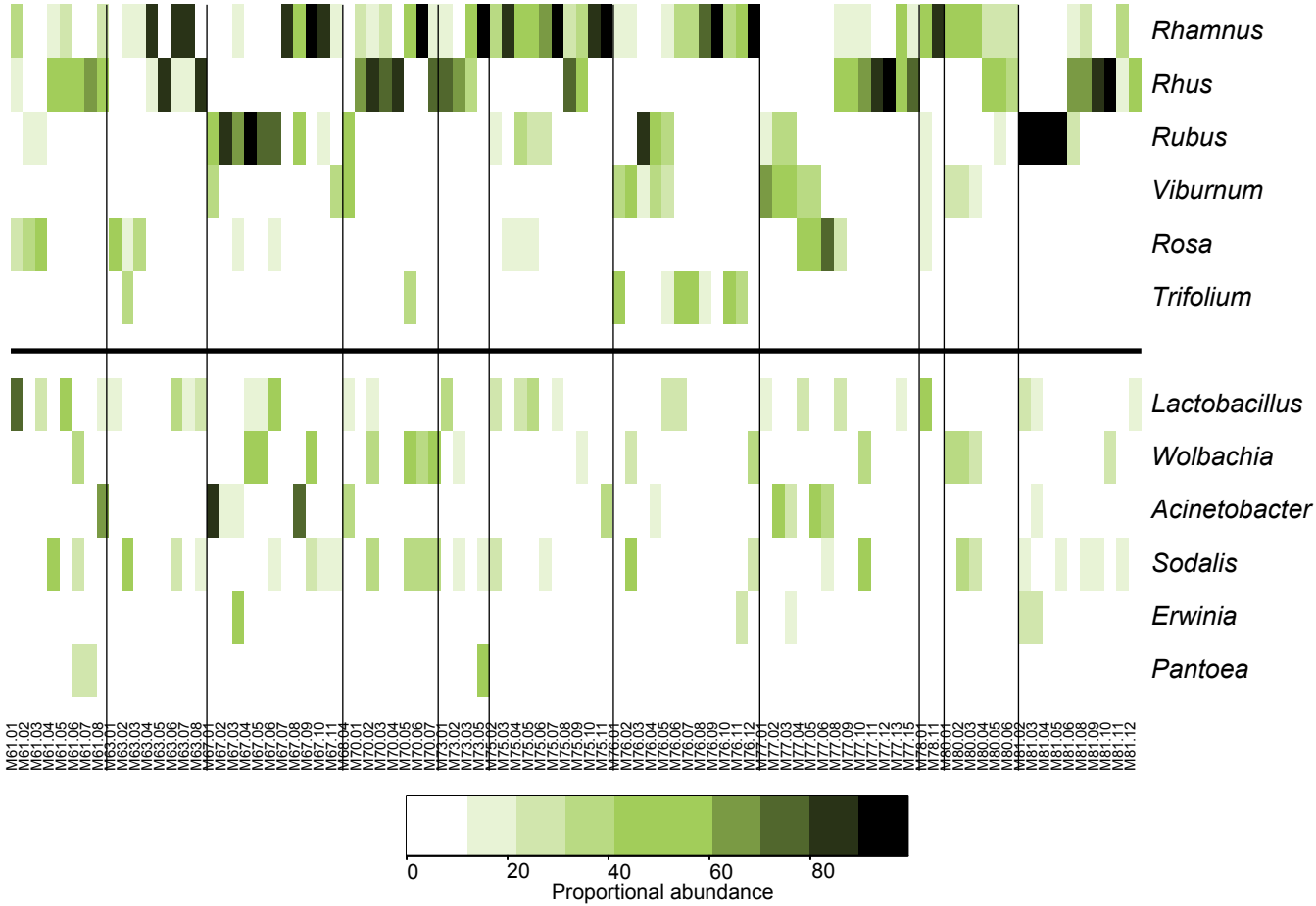

Plants
Rhamnus
Rhus
Rubus
Viburnum
Rosa
Trifolium
Bacteria
Lactobacillus
Wolbachia
Acinetobacter
Sodalis
Erwinia
Pantoea
M61.01
M61.02
M61.03
M61.04
M61.05
M61.06
M61.07
M61.08
M63.01
M63.02
M63.03
M63.04
M63.05
M63.06
M63.07
M63.08
M67.01
M67.02
M67.03
M67.04
M67.05
M67.06
M67.07
M67.08
M67.09
M67.10
M67.11
M67.12
M70.01
M70.02
M70.03
M70.04
M70.05
M70.06
M70.07
M73.01
M73.02
M73.03
M73.05
M75.02
M75.03
M75.04
M75.05
M75.06
M75.07
M75.08
M75.09
M75.10
M75.11
M76.01
M76.02
M76.03
M76.04
M76.05
M76.06
M76.07
M76.08
M76.09
M76.10
M76.11
M76.12
M77.01
M77.02
M77.03
M77.04
M77.05
M77.06
M77.08
M77.09
M77.10
M77.11
M77.12
M77.13
M77.15
M78.01
M78.11
M80.01
M80.02
M80.03
M80.04
M80.05
M80.06
M81.02
M81.03
M81.04
M81.05
M81.06
M81.08
M81.09
M81.10
M81.11
M81.12
0
20
40
60
80
Proportional abundance

Fig. S3

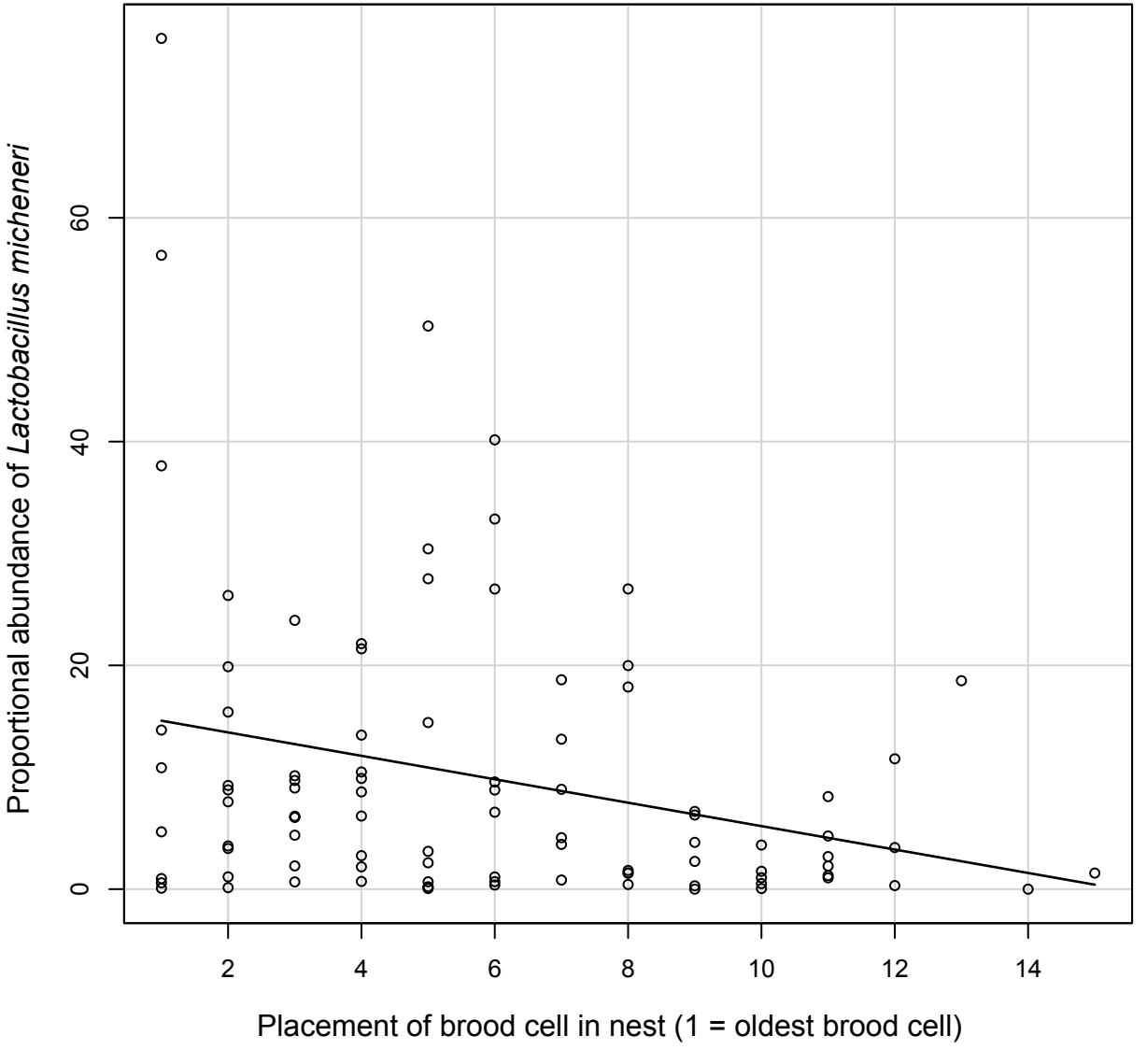
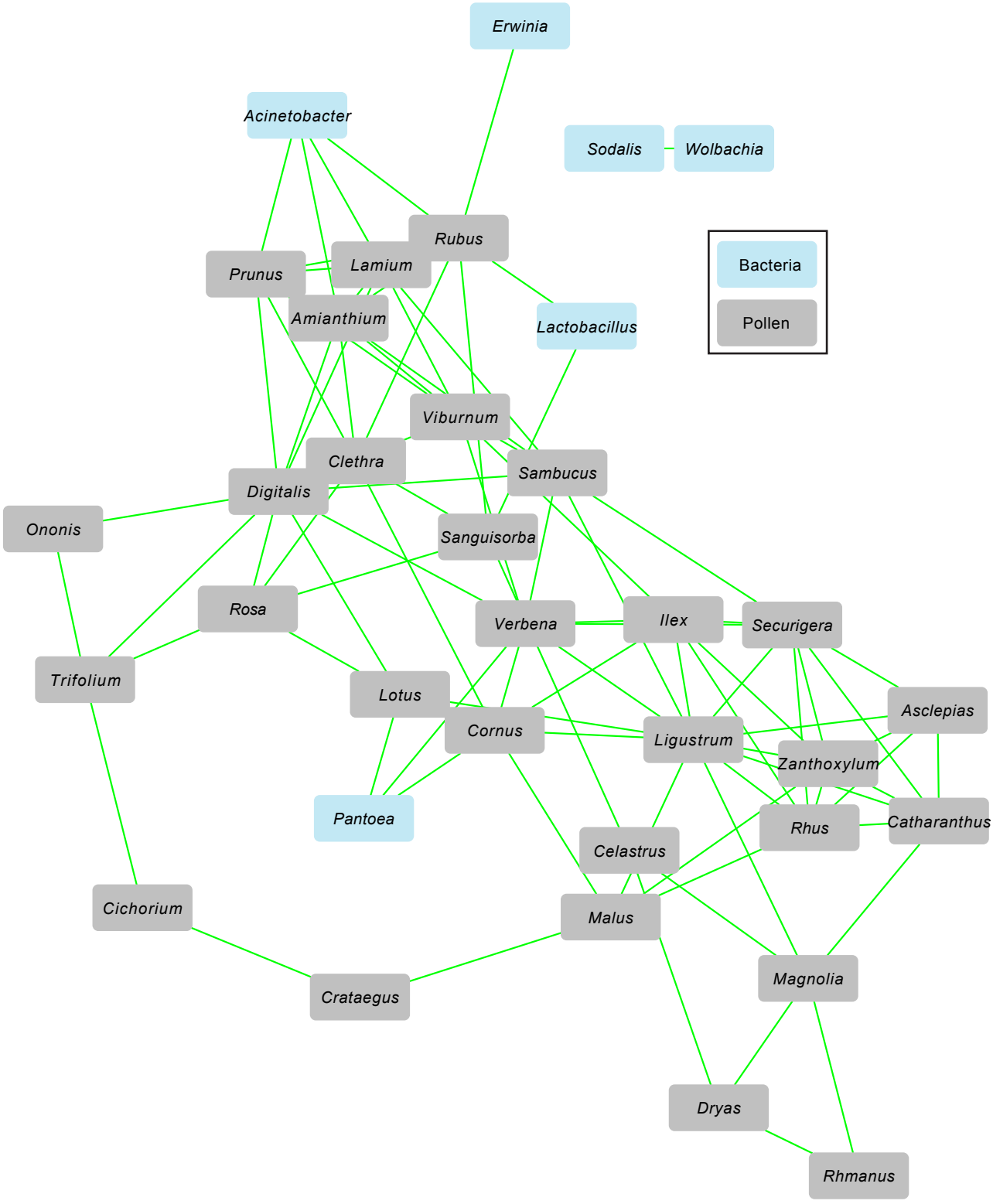

Proportional abundance of *Lactobacillus micheneri*
0
20
40
60
2
4
6
8
10
12
14
Placement of brood cell in nest (1 = oldest brood cell)

Fig. S4.


Erwinia
Acinetobacter
Sodalis
Wolbachia
Rubus
Prunus
Lamium
Amianthium
Lactobacillus
Bacteria
Pollen
Viburnum
Clethra
Digitalis
Sambucus
Ononis
Sanguisorba
Rosa
Verbena
Ilex
Securigera
Trifolium
Lotus
Asclepias
Cornus
Ligustrum
Zanthoxylum
Pantoea
Rhus
Catharanthus
Celastrus
Cichorium
Malus
Magnolia
Crataegus
Dryas
Rhmanus