Published in final edited form as:

Mol Ecol. 2016 September ; 25(18): 4461–4471. doi:10.1111/mec.13787.

Strain diversity and host specificity in bee gut symbionts revealed by deep sampling of single copy protein-coding sequences

J. Elijah Powell[1,*], **Nalin Ratnayeke**[1], and **Nancy A. Moran**[1]

[1]Department of Integrative Biology, University of Texas, NMS building, 2506 Speedway, A5000, Austin, TX 78712, USA

Abstract

High throughput rRNA amplicon surveys of bacterial communities provide a rapid snapshot of taxonomic composition. But strains with nearly identical rRNA sequences often differ in gene repertoires and metabolic capabilities. To assess strain-level variation within *Snodgrassella alvi*, a gut symbiont of corbiculate bees, we performed deep sequencing on amplicons of a single copy coding gene (*minD*) as well as the 16S rDNA V4 region. We surveyed honey bees (*Apis mellifera*) sampled globally and 12 bumble bee species (*Bombus*) sampled from two regions of the USA. The *minD* analyses reveal that *S. alvi* contains far more strain diversity than is evident from 16S rDNA analysis. Many taxa inferred on the basis of 16S rDNA are shared between *A. mellifera* and *Bombus* species, but taxa inferred on the basis of *minD* are never shared and often are restricted to particular *Bombus* species. Clustering based on *minD* revealed that gut communities often reflect host species and geographic location. Both *minD* and 16S rDNA analyses indicate that strain diversity is higher in *A. mellifera* than in *Bombus* species. The *minD* locus flanks a 16S gene, enabling development of strain-specific 16S fluorescent probes to illuminate the spatial relationship of strains within the bee gut.

Introduction

The gut microbiota of corbiculate bees contains several specialized bacterial clades that are shared across host species in the genera *Apis* and *Bombus* (Cox-Foster et al. 2007; Martinson et al. 2011; Koch et al. 2013). Notable among these is *Snodgrassella alvi* (Betaproteobacteria: Neisseriales), which inhabits the ileum, the hindgut region joining midgut to rectum, along with an associated species, *Gilliamella apicola* (Gammaproteobacteria) These two bacterial species occur worldwide in honey bees (*Apis mellifera*) and other members of the genus *Apis* (Martinson *et al.*, 2011; Cox-Foster *et al.*, 2007; Jeyaprakash *et al.*, 2003; Mohr and Tebbe, 2006; Ahn *et al.*, 2012; Disayathanoowat *et al.*, 2012), and also in all species of the genus *Bombus* (Cariveau *et al.*, 2014; Koch and

*Corresponding author (eli.powell@utexas.edu).

Conflict of Interest

The authors state they have no conflicts of interest.

Supplementary information

Supplementary information is available at ISME Journal's website.

Schmid-Hempel, 2011; Koch *et al.*, 2013; Kaltenpoth, 2011; Li *et al.*, 2015; Martinson *et al.*, 2011). *S. alvi* forms a robust biofilm on the epithelium and intima of the ileum, while *G. apicola* grows in a layer over this and into the lumen (Martinson *et al.*, 2012; Kwong and Moran, 2013). *S. alvi* and *G. apicola* are of particular interest because of their close interaction with host tissue, hypothesized pathways for host interaction and their ubiquitous presence in bees worldwide (Kwong *et al.*, 2014; Engel *et al.*, 2012). With the recent declines in bee populations globally, it is critical to investigate potential effects of these bacteria on host health, nutrition and pathogen defense (Kaltenpoth and Engl, 2014). Furthermore, these species are models for investigating evolution, specialization, diversity, distribution and colonization in host-restricted, socially transmitted symbionts. Analogous systems are found in vertebrate guts (Frese *et al.*, 2011; Walter and Ley, 2011).

Genomic analysis of single bacterial cells sorted from honey bee guts has revealed the presence of distinct strains of *S. alvi* with different gene sets (Engel *et al.*, 2014). The extent of this strain variation and its distribution across members of the same and different host species is not yet evident. Analysis of cloned amplicons of nearly full length 16S rDNA from individual bees indicates some level of specificity to host species (Koch *et al.*, 2013) as well as to geographic region and colony (Moran *et al.*, 2012). However, rRNA sequences are highly conserved and often cannot be used to distinguish closely related strains. Thus, in the bee gut, strains with >99% identity of 16S rRNA commonly have only 80–90% identity of homologous protein-coding genes and may have very different gene repertoires (Engel *et al.*, 2014). The strain-level variation in the honey bee gut parallels that recently documented for species in the human gut microbiota (Zhu *et al.*, 2015; Greenblum *et al.*, 2015). The limitations of 16S rRNA surveys are further exacerbated in studies using high throughput sequencing technologies based on short amplicons (<400 bp) typically of the V4 region of 16S rRNA (Caporaso *et al.*, 2012); these short reads are even less likely than full length 16S rRNA surveys to distinguish closely related strains.

Standard surveys based on 16S sequences rely on clustering reads into Operational Taxonomic Units (OTUs) based on sequence similarity, usually at the 97% identity threshold (the conventionally accepted level for species boundaries (Stackebrandt *et al.*, 2002). To date, attempts to resolve finer relationships from 16S data have been limited. Some investigators have examined relationships by clustering OTUs at >97% identity (Moeller *et al.*, 2013) threshold, while another group (Tikhonov *et al.*, 2014) used a ‘dynamical similarity’ approach to examine highly similar (>99.2%) OTUs based on temporal signatures in individual hosts sampled over time.

To explore novel approaches to interrogating strain-level diversity, we have conducted high throughput sequencing of a region of a single copy protein-coding gene (*minD*) in *S. alvi* and compared these results to those from a typical 16S V4 analysis of the same samples. We exploited the greater resolution of the protein-coding sequence to address several biological questions. First, to what extent are individual strains restricted to particular host species? Second, what is the extent of strain diversity within host individuals and species? Third, is strain diversity higher in honey bees than in bumble bees, as expected based on life cycle differences between these hosts? We used strain-specific fluorescent in situ hybridization

(FISH) to determine if coexisting strains live in distinct areas of the gut within an individual host.

Materials and Methods

Bombus samples

Two species of *Bombus* (*B. bimaculatus* and *B. impatiens*) were collected at 4 locations near New Brunswick, New Jersey in spring and summer 2013, and 12 species (*B. appositus, B. balteatus, B. bifarius, B. fervidus, B. flavifrons, B. frigidus, B. huntii, B. mixtus, B. nevadensis, B. occidentalis, B. rufocinctus, and B. sylvicola*) were collected at 15 sites near Crested Butte, Colorado in summer 2014 (see Supplementary Table SI1 for locations, dates and additional information). Workers were captured with nets in the field and visually identified; their guts were aseptically dissected and placed in RNAlater (Ambion, Austin, TX, USA) for preservation. Preserved guts were stored at −80°C until nucleic acid was extracted. Specimen bodies were pinned and maintained as vouchers.

DNA was extracted from these samples using the Epicentre Masterpure kit (Epicentre, Madison, WI, USA) using the manufacturers' protocol with an initial lysozyme treatment (250,000 units 1 hour at 37°C). DNA was suspended in ultrapure water and normalized to a concentration of 20 ng/μl.

Bombus species identifications were verified by amplifying, sequencing and submitting the COI gene to the Bold System V3 identification database as detailed in (Cariveau *et al.*, 2014).

Apis mellifera samples

Collections of adult workers were obtained from several colonies in the USA, Europe and New Zealand. These included samples from colonies with histories of antibiotic treatments that ranged from untreated feral colonies and hives with no history of antibiotic intervention to those from recently treated colonies. Specimens were preserved in 95% ethanol and stored at 4°C until dissection. Whole gut samples were prepared via the methods outlined in (Powell *et al.*, 2014) and archived at −80°C.

Amplification and sequencing

Sample DNA extracts were screened for the presence of bacteria and specifically *S. alvi* using the 27F-short/1507R universal 16S primers and the Beta forward primer with 1507R, using primers and methods from (Martinson *et al.*, 2011). Samples producing amplicons based on visualization on 2% agarose gels were utilized for high throughput microbial diversity screening. Bacterial diversity of the samples was assessed by amplifying the V4 region of the 16S rRNA gene (Powell *et al.*, 2014), and runs were performed at the University of Texas Genome Sequencing and Analysis Facility (UT-GSAF) on MiSeq (Illumina, San Diego, CA, USA). Read processing metrics are detailed in Supplementary Table SI1. We selected *minD* as a single copy protein-coding gene (locus SALWKB2_0320 of GenBank accession CP007446). For a set of sequenced genomes of *S. alvi* isolates, we observed identities of 99.4% +/− 0.6 S.D at the 16S V4 locus compared to 90.6% +/− 7.7

S.D. at *minD*. Most coding genes show similar or greater divergence (Engel *et al*., 2014). This gene was selected because, based on *in silico* analysis of representative genomic sequences of *S. alvi*, it contains appropriately spaced conserved regions for obtaining assembled sequences from a 2 × 300 bp paired-end Illumina MiSeq run. The choice was also motivated by the chromosomal position of *minD* adjacent to a 16S rRNA gene, enabling us to link coding gene diversity to polymorphisms within the rDNA that might serve as targets for FISH probes, thus enabling localization of different strains.

Primers specific to the *minD* of *S. alvi* were designed based on *minD* alignments from several sequenced genomes from isolates of *S. alvi* as well as *minD* sequences obtained by single cell genome sequencing (Engel *et al*., 2014). A 514 bp *minD* region was amplified and prepared for sequencing via a two-step amplification procedure. Samples were amplified in triplicate with Illumina-adapted primers RND1_MinDF1 5′ -TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TGCGGTAATTGATTTTGATG**-3′ and RND1_MinDR2wbl2: 5′-GTCTCGTGGGCTCGGAGATGTGTATAA-GAGACAG**RCGYAAAATATCCTGAATATC**-3′ (*minD* primers are in bold and Illumina platform primers are in standard text). This first round of PCR consisted of an initial 2 min denaturation at 98°C, 30 cycles of 10 s at 98°C, 15 s at 59°C and 15 s at 72°C plus a final 2 min extension at 72°C. Each 20 μl reaction mixture contained 1 ul template, 4 ul component buffer GC, 1 μl of each 10 μM primer, 0.2 μl Phusion Taq (New England Biochemical, Ipswich, MA, USA), 0.4 μl 10mM deoxynucleotide triphosphates (dNTPs), 0.6 μl DMSO and 11.8 ul H_2O. The reactions were pooled and cleaned with Agencourt Ampure beads (Beckman Coulter Genomics, Indianapolis, IN, USA) and resuspended in 25ul dd H_2O. A second PCR reaction was performed in order to add dual indexed barcodes to both ends of the amplicon. This reaction used primers provided by UT-GSAF based on (Wang and Qian, 2009) with the format: Hyb_Fnn_i5, AATGATACGGCGACCACCGAGATCTACAC nnnnnn TCGTCGGCAG CGTC Hyb_Rnn_i7, CAAGCAGAAGACGGCATACGAGAT nnnnnn GTCTCGTGGGCTCG G, where nnnnnn is the variable barcode sequence. The PCR reactions used 1 μl of the cleaned amplicon from the first PCR, 4 μl component buffer HF, 2 μl of each 5 μM primer, 0.2 μl Phusion Taq (New England Biochemical, Ipswich, MA, USA), 0.4 μl 10mM deoxynucleotide triphosphates (dNTPs) and 10.4 ul H_2O. The thermocycling profile consisted of an initial 2 min denaturation at 98°C, 12 cycles of 10 s at 98°C, 15 s at 55°C and 15 s at 72°C and a final 2 min extension step at 72°C. Reactions were cleaned with Ampure beads, quantified with a Qubit dsDNA BR assay (Invitrogen/Life Technologies, Grand Island, NY, USA), and equimolar amounts of all reactions were sent for sequencing. Sequencing was performed as several partial MiSeq 2×300 runs at UT-GSAF.

Bioinformatic analyses

Qiime version 1.9 (Caporaso *et al*., 2010) was used to process and analyze both 16S and *minD* datasets. Reads were assembled by the seqprep method, and OTU clustering was performed at both the 97 and 100% identity thresholds using the summaclust method (Mercier *et al*., 2013). Singletons for both sets, and both plastid and mitochondrial sequences for the 16S set were filtered out. Additionally, OTUs comprising less than 1% of any sample were removed from the table, and the representative sequences of filtered OTUs were

removed. Taxonomic assignments were performed by using the SILVA 119 database (Quast *et al.*, 2013) at the default Qiime settings. An OTU table consisting of filtered 16S reads pertaining to *S. alvi* was constructed and used for comparison to the full *minD* table (read processing statistics are reported in Table SI1).

Heatmaps of shared OTUs within the *Apis* and *Bombus* sets were generated using relative abundances within hosts in OTU tables and visualized in R (R Development Core Team, 2008). Only samples with >90 total reads at a clustering level were used to generate relative abundance tables and heatmaps. Alpha and Beta diversity analysis of the 16S 97% and 100% identity clustered tables was performed by subsampling without replacement at 100 reads, while the *minD* table was subsampled at 500 reads. Sampling level was based on examination of rarified Alpha diversity tables with 10 iterations of subsampling at every 100 reads (shown in Figure SI1). Samples not having sufficient read depth were eliminated from the dataset. Kruskal-Wallis was used in R as a one-way test of variance of the Alpha diversity means, and Kruskal-Wallis multiple comparisons tests were utilized post hoc through the R package PMCMR (Pohlert, 2013) to examine means between host species groups.

Binary Bray-Curtis distances were calculated by performing jackknifed beta analyses at the previously mentioned sampling depths and were used to look at beta diversity between samples and groups. PCoA ordinations were plotted for these distance matrices, and the matrices were utilized for statistical analysis via the ADONIS method of permutational multivariate analysis with 999 permutations (Oksanen *et al.*, 2013). Phylogenetic analysis was performed by retrieving *minD* sequences from available genome sequences of *S. alvi* strains isolated from *Bombus* and *Apis* species including published (Kwong *et al.*, 2014; Engel *et al.*, 2014) and novel genomes sequenced using Illumina HiSeq. These sequences were submitted to the www.phylogeny.fr website (Dereeper *et al.*, 2008) where they were aligned via MUSCLE (Edgar, 2004), screened for informative sites via Gblock (Castresana, 2000), and used to infer a phylogeny via PHYML 3.0 (Guindon *et al.*, 2010). (Reads are deposited with NCBI Sequence Read Bioproject: PRJNA305910)

Strain specific fluorescent in situ hybridization (FISH) probe experiment

To visualize spatial distribution of strains within the bee ileum, *S. alvi* strains wkB2 (Kwong and Moran, 2013) and PE78 were cultured on HIA + 5% sheep's blood at 35°C and 5% CO_2 for 36 hours. Colony PCR was performed on the strains with primers RND1_MinDR2wbl2 and 1507R (Martinson *et al.*, 2011) in order to amplify the sequence running from *minD* through 16S (4.3 KB). Targeting of individual strains in a mixed community would be performed by utilizing a strain-specific primer for the *minD* locus. Full-length amplicon was screened by running on a 2% agarose gel in 1x TAE for 20 minutes at 120V. Product was sent to the Sequencing Core at the University of Texas with 16S primers 27Fshort and 1507R (Martinson *et al.*, 2011). Full-length sequence was returned and aligned in Geneious (Kearse *et al.*, 2012). A short polymorphic region was observed (3 bp of a 4 bp sequence at position 1168–1171 based on *E. coli*). Fluorescent probes were designed for each strain (for wkB2, 5′- Cy3/TCCTCCGGTTTGTCACCGGC -3′ and for PE78, 5′-TEX615/ TCCTCCGGCATATCACCGGC-3′ (Life Technologies, Grand Island, NY, USA). The

probe position was queried for signal intensity via Probe base (Loy *et al.*, 2003) and a low intensity (II) was indicated. Helper oligos, complementary to regions flanking the probe position were designed (H-20 5′-AGTCTCATTAGAGTGCCCA-3′, Hnull20+40 5′-TCCTCCGGCATATCACCGGC-3′, H+20 5′-ACTTGACGTCATCCCCACCT-3′) in order to raise probe intensity. Helper oligos have been demonstrated to reliably raise probe intensity when problematic secondary structure is an issue (Fuchs *et al.*, 2000). A general probe previously used for *S. alvi* was also used (beta-572) (Martinson *et al.*, 2012).

Cohorts of 10 *A. mellifera* workers, 1 day post emergence, were fed 5 μl 25% w/v sucrose in PBS either alone or with an OD_{600} = ~1 of either strain PE78, wkB2 or a 1:1 ratio of both. Bees were fed on sterile sugar solution and gamma irradiated bee bread (Powell *et al.*, 2014) for 5 days. Two randomly selected bees from each cohort were immobilized by refrigeration, dissected, had DNA extracted with Qiagen DNeaAsy kits from ileums and were then screened via PCR with the round 1 *minD* primers and PCR protocol. Positive amplification was observed for all bees receiving *S. alvi* and not for controls, providing evidence of effective inoculation for the group. We aseptically dissected the ileums of three bees from each cohort. The ileums were dehydrated, fixed, paraffin-embedded, sectioned and hybridized with the probes, helper oligos (both at 10 pmol/ml) and DAPI (at 10 ug/ml) as per (Engel *et al.*, 2013). Sections were viewed on a Zeiss 710 LSM confocal microscope, and the resulting images were converted to composite figures with scale bars in imageJ (Abràmoff *et al.*, 2004) and assembled in Photoshop CS5.

Results

Raw read data and processing steps from the MiSeq data are summarized in Table SI1, which also includes samples excluded from analyses due to lack of amplification or low read depth.

Strain distribution across hosts based on minD and 16S rDNA

The numbers and relative abundances of OTUs revealed by 16S and *minD* at the 97% and 100% clustering levels are summarized as heatmaps in Figure 1. The 97% 16S analyses yielded 7 OTUs for *A. mellifera* samples and 7 OTUs for *Bombus* samples (Figure 1a, b). All OTUs were shared between *A. mellifera* and *Bombus*, and no clear patterns of specificity to host species or location were observed. In contrast, the 97% *minD* analysis yielded 29 OTUs in *A. mellifera* samples (Figure 1c, d) and 25 OTUs in *Bombus* samples, and none of the OTUs was shared between the two host genera. The patterns of distribution of *minD* OTUs showed some separation by geographic location for *Apis* (e.g., Texas samples vs. other sites) as well as some separation between different *Bombus* host species (e.g., *B. nevadensis* versus *B. appositus*).

The 100% analyses for 16S yielded 14 OTUs for *Apis* and 7 for *Bombus* (Figure 1e, f). Two OTUs (16S100_OTU3 and 16S100_OTU106) were shared between the two host genera. Again there were no clear patterns reflecting host location. However, *Bombus* samples appeared to exhibit three different strain pattern motifs. The clustering of *Bombus* species is interpreted in more detail in the results reported from the PCoA clustering below. The 100% clustering analysis of 16S V4 for *Apis* samples was hampered by large-scale depletion of

total reads (loss of ~50% of reads, see Table SI1); many low abundance OTUs were generated then screened out based on the 1% filtering threshold. As a result, many samples fell below the cutoff of at least 90 reads for abundance analysis. *Bombus* contained relatively few and more distinct OTUs leading to a much-reduced filtering and depletion effect. This depletion effect was seen in the *minD* data though it did not drop samples below the filter threshold because of the much deeper initial sequencing depth.

The 100% *minD* tables revealed far more OTUs for both *A. mellifera* (total of 213 OTUs) and for *Bombus* (total of 118 OTUs) (Figure 1 g–f). No OTUs were shared between *A. mellifera* and *Bombus*, and some *Bombus* species or groups of species had distinctive OTU composition (including *B. sylvicola, B. fervidus, B. flavifrons*, members of *B. bifarius* and *B. rufocinctus*).

Community analyses of S. alvi strains based on minD and 16S

To address whether and how *S. alvi* strain profiles vary among samples and the effect of using 16S versus *minD* to define OTUs, we carried out analyses based on principal components analysis (PCoA) using binary Bray Curtis distances for ordination. For the whole sample set, including both *A. mellifera* and *Bombus*, clustering patterns differ depending on whether 16S or *minD* data were used (Figure 2a, b). In the PCoA plot based on 97% clustering of 16S sequences, the dim1 vs. dim2 plot accounts for 93.35% of the observed variation in that dataset. The plot shows distinct clustering of *A. mellifera* and *Bombus*, and the host genus category is statistically significant (Adonis PERMANOVA, $R^2 = 0.37$; $F_{1, 110} = 65.39$; $P = 0.001$). No distinctions were evident based on sampling location, and the *Bombus* samples are so homogenous that many of the corresponding points are superimposed on the plot.

The PCoA derived from *minD*-related distances at 97% also shows an impact of host genus (Adonis PERMANOVA, $R^2 = 0.30$; $F_{1, 100} = 42.87$; $P = 0.001$). Furthermore, *Apis* samples from Texas cluster separately from other *Apis* samples, which do not display clear groupings (Figure 2b). Potentially, other location-specific patterns exist but are not evident due to limited sampling. Also, the Texas samples were collected on the same date, and a seasonal component to strain composition may contribute to their clustering. *Bombus* species also show some grouping by species (e.g. *B. fervidus*, *B. impatiens* and *B. appositus*) with a tightly clustered group composed mainly of the two New Jersey species, *B. impatiens* and *B. bimaculatus.* When the *Bombus* samples alone are placed in an ordination, the host species category is highly significant (Adonis PERMANOVA, $R^2 = 0.61$; $F_{11, 50} = 7.11$; $P = 0.001$).

The PCoA derived from binary Bray Curtis for 16S data at 100% clustering show a clear split by genus and also distinct clustering of species groups. *B. nevadensis* and *B. impatiens* cluster together with *B. fervidus, and, B. occidentalis*, although these host species are phylogenetically divergent (Cameron *et al.*, 2007). The *B. appositus* samples are spread throughout the *Bombus* portion of the ordination while *B. bifarius, B. rufocinctus, B. mixtus*, and *B. silvicola* dominate another cluster. Examining an ordination of only the *Bombus* samples, groupings based on species categories are strong and statistically significant (Adonis PERMANOVA, $R^2 = 0.78$; $F_{10,45} = 15.6$; $P = 0.001$) (Fig. 3c)

For the *minD* PCoA for OTUs at 100% identity, Texas samples cluster together, though not as tightly as for the 97% plot. This analysis also yields groupings of *Bombus* by species though these groupings are not well separated within the ordination plot. In an ordination based on *Bombus* samples only, samples cluster by species (Adonis PERMANOVA, R^2 =0.43; $F_{11,41}$ =2.84; P = 0.001) (Fig. 3d).

Diversity of S. alvi strains within samples based on minD versus 16S

The mean number of observed OTUs (alpha diversity) is higher for *minD* than for 16S, at both clustering levels (Figure 3). *A. mellifera* samples have the highest number of observed OTUs in all analysis conditions. In both of the 16S analyses, *A. mellifera* samples contain significantly more OTUs than samples from any *Bombus* species (Kruskal-Wallis, P<0.001). For the *minD* analysis, several *Bombus* species (e.g., *B. appositus*, *B. bifarius*, and *B. occidentalis*) contain elevated numbers of OTUs and resembled *A. mellifera* samples in terms of OTU richness; however, within both analyses the only statistically different groups are *A. mellifera* against *B. fervidus, B. mixtus and B. impatiens* (Kruskal-Wallis *post hoc, P* < 0.05). Much of the lack of significance can be attributed to low replicate number for most species.

Direct comparison between numbers of observed OTUs from the 16S versus *minD* datasets was performed using samples with observed values at both sampling levels for the two targets. No groups were significantly different between the two loci (Kruskal-Wallis *post hoc, P* > 0.05), though again this may be due to the low number of replicates.

Phylogenetic relationships of strains based on minD

Inferred phylogenetic relationships based on *minD* sequences from OTU representatives of the 97% clusters as well as full length sequences from isolate genomes indicate that *S. alvi* clades are largely separate between *Apis* and *Bombus* (Figure 4). A surprising exception is *minD* 97_OTU_7, represented mainly in *B. impatiens* samples but occurring on a long branch within the *Apis*-associated clade. Several *Bombus*-associated clades are specific to *Bombus* species (noted as clades *i* and *ii*), whereas others appear to be more generalized across *Bombus* species and subgenera. The most widespread clade is noted as broad host range clade *gen_i*.

FISH probes based on minD strain sequence divergence

Genotypic strain difference may be reflective of the utilization of alternative ecological strategies of the bacteria within the host. FISH probes, designed by associating *minD* sequence divergence with polymorphisms in 16S probe targets, showed small isolated patches of *S. alvi* in bees inoculated with each single strain (Figure 5a). When coinoculated, the two strains developed a robust composite biofilm (Figure 5b).

Discussion

This study adds weight to previous investigations that have indicated presence of multiple strains of *S. alvi* within a single host (Moran *et al.*, 2012). This condition is likely even more

extensive for *Gilliamella* and *Lactobacillus* which show higher diversity of 16S OTUs in metagenomic analyses (Engel *et al.*, 2012).

Both the 16S and *minD* analyses show that *A. mellifera* samples contain higher diversity of *S. alvi* strains than do any *Bombus* species (Figure 1 and Figure 2). However, the use of the protein-coding locus illuminates patterns of bacterial distribution not apparent from the standard 16S method. Most prominently, the *minD* analysis reveals extensive specificity of strains to host species (Figure 4). This specificity could result from inability to colonize alternate hosts, or from lack of opportunity. However, opportunities for symbiont transmission likely abound; for example, *B. nevadensis* forages at the same flowers as other bees in the sampled Colorado sites, yet has distinctive OTUs. Therefore, some barriers due to coevolution with specific hosts appear likely. Such barriers have been shown experimentally for *S. alvi* strains from *Apis* versus *Bombus* (Kwong *et al.*, 2014).

The higher *S. alvi* diversity in *A. mellifera* as compared to *Bombus* species is potentially the result of life cycle differences. Honey bees form new colonies by swarming: the queen is accompanied by thousands of workers which provide a diverse gut community inoculum for the new colony, which can live for years. *Bombus* colonies, on the other hand, are restarted each spring by a single overwintered queen, who is likely the sole source of gut symbionts for her colony. Thus, the *Bombus* life cycle imposes a transmission bottleneck expected to lower diversity, whereas the *Apis* life cycle does not.

A. mellifera colonies do exhibit differences in strain composition, raising the possibility that differences in strain composition contribute to biological differences among colonies. Unexpectedly, our results show a slight trend for *A. mellifera* colonies with recent antibiotic exposure to have higher *S. alvi* strain diversity (Figure SI2). Diversity levels could also reflect other management practices; for example, antibiotic use may be offset by pooling workers from different colonies, as both practices are common in larger commercial apiaries. Thus, management practices may determine levels of strain diversity in honey bee guts. To date, no direct evidence supports the hypothesis that higher strain diversity of gut bacteria benefits bee health, although genomic studies do indicate that distinct strains vary in capacity to utilize different lignocellulose components and sugars (Lee *et al.*, 2015; Engel *et al.*, 2012; Kwong and Moran, 2013).

Our study provides potential tools for visualizing strain distributions within guts via the use of FISH probes. These probes exploit the diversity information provided by a protein coding gene by associating it with polymorphic probe targets in rRNA sequences. In the case we examined, two *S. alvi* strains showed similar colonization patterns and were intermixed when co-inoculated. Our strain-specific probes demonstrated the potential to visualize spatial relationships of strains within the gut.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dan Cariveau, Howard Ochman, Jessica Zung, James Thomson, Dee Lusby, Kerry Gentleman, Jerry Shue, Eva Nováková, Peter Gallman and Sooja Nehrlich, for sample collections. Kim Hammond, for maintaining research colonies and Ryuichi Koga for advice concerning FISH. We also thank Waldan Kwong for *Apis* species *minD* sequences and Hauke Koch for *Bombus* species *minD* sequences. The Rocky Mountain Biological Laboratories provided support for *Bombus* collecting. This work was supported by National Science Foundation award 1415604 to NAM and by the University of Texas at Austin.

References

Abràmoff MD, Magalhães PJ, Ram SJ. Image processing with imageJ. Biophotonics Int. 2004; 11:36–41.

Ahn JH, Hong IP, Bok JI, Kim BY, Song J, Weon HY. Pyrosequencing analysis of the bacterial communities in the guts of honey bees *Apis cerana* and *Apis mellifera* in Korea. J Microbiol. 2012; 50:735–745. [PubMed: 23124740]

Cameron SA, Hines HM, Williams PH. A comprehensive phylogeny of the bumble bees (Bombus). Biol J Linn Soc. 2007; 91:161–188.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. correspondence QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. Nat Publ Gr. 2010; 7:335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012; 6:1621–1624. [PubMed: 22402401]

Cariveau DP, Powell JE, Koch H, Winfree R, Moran NA. Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (Bombus). ISME J. 2014; 8:1–11. [PubMed: 24152717]

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17:540–552. [PubMed: 10742046]

Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. Science. 2007; 318:283–7. [PubMed: 17823314]

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008; 36:W465–W469. [PubMed: 18424797]

Disayathanoowat T, Young JPW, Helgason T, Chantawannakul P. T-RFLP analysis of bacterial communities in the midguts of *Apis mellifera* and *Apis cerana* honey bees in Thailand. FEMS Microbiol Ecol. 2012; 79:273–281. [PubMed: 22092273]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acid Res. 2004; 32:1792–1797. [PubMed: 15034147]

Engel P, James R, Koga R, Kwong WK, McFrederick QS, Moran NA. Standard methods for research on Apis mellifera gut symbionts. J Apic Res. 2013; 52:1–24.

Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A. 2012; 109:11002–11007. [PubMed: 22711827]

Engel P, Stepanauskas R, Moran NA. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. PLoS Genet. 2014; 9:e1004596.

Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, et al. The Evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. PLoS Genet. 2011; 7:e1001314. [PubMed: 21379339]

Fuchs BM, Glöckner FO, Wulf J, Glo FO. Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. Appl Environ Microbiol. 2000; 66:3603–3607. [PubMed: 10919826]

Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. Cell. 2015; 160:583–594. [PubMed: 25640238]

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol. 2010; 59:307–321. [PubMed: 20525638]

Jeyaprakash A, Hoy MA, Allsopp MH. Bacterial diversity in worker adults of *Apis mellifera capensis* and *Apis mellifera scutellata* (Insecta: Hymenoptera) assessed using 16S rRNA sequences. J Invertebr Pathol. 2003; 84:96–103. [PubMed: 14615218]

Kaltenpoth M. Honeybees and bumblebees share similar bacterial symbionts. Mol Ecol. 2011; 20:439–440. [PubMed: 21298852]

Kaltenpoth M, Engl T. Defensive microbial symbionts in Hymenoptera. Funct Ecol. 2014; 28:315–327.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28:1647–9. [PubMed: 22543367]

Koch H, Abrol DP, Li J, Schmid-Hempel P. Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. Mol Ecol. 2013; 7:2028–2044.

Koch H, Schmid-Hempel P. Bacterial Communities in Central European Bumblebees: Low Diversity and High Specificity. Microb Ecol. 2011; 62:121–133. [PubMed: 21556885]

Kwong W, Moran NA. 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*. Int J Syst Evol Microbiol. 2013; 63:2008–2018. [PubMed: 23041637]

Kwong WK, Engel P, Koch H, Moran NA. Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc Natl Acad Sci. 2014; 111:1405838111.

Lee FJ, Rusch DB, Stewart FJ, Mattila HR, Newton ILG. Saccharide breakdown and fermentation by the honey bee gut microbiome. Environ Microbiol. 2015; 17:796–815. [PubMed: 24905222]

Li J, Powell JE, Guo J, Evans JD, Wu J, Williams P, et al. Two gut community enterotypes recur in diverse bumblebee species. Curr Biol. 2015; 25:R652–R653. [PubMed: 26241138]

Loy A, Horn M, Wagner M. probeBase: an online resource for rRNA-targeted oligonucleotide probes. Nucleic Acids Res. 2003; 31:514–6. [PubMed: 12520066]

Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA. A simple and distinctive microbiota associated with honey bees and bumble bees. Mol Ecol. 2011; 20:619–628. [PubMed: 21175905]

Martinson VG, Moy J, Moran NA. Establishment of characteristic gut bacteria during development of the honeybee worker. Appl Environ Microbiol. 2012; 78:2830–2840. [PubMed: 22307297]

Mercier, C., Boyer, F., Bonin, A., Coissac, E. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. 2013. Available: http://metabarcoding.org/sumatra

Moeller AH, Peeters M, Ndjango J, Li Y, Hahn BH, Ochman H. Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. Genome Res. 2013; 10:1715–1720.

Mohr KI, Tebbe CC. Diversity and phylotype consistency of bacteria in the guts of three bee species (Apoidea) at an oilseed rape field. Environ Microbiol. 2006; 8:258–272. [PubMed: 16423014]

Moran NA, Hansen AK, Powell JE, Sabree ZL. Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. PLoS One. 2012; 7:1–10.

Oksanen, J., Blanchet, FG., Kindt, R., Legendre, P., Minchin, PR., O'Hara, RB., et al. vegan: Community Ecology Package. 2013. Available: http://CRAN.R-project.org/package=vegan

Pohlert, T. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). 2013. Available: http://CRAN.R-project.org/package=PMCMR

Powell JE, Martinson VG, Urban-Mead K, Moran NA. Routes of acquisition of the gut microbiota of *Apis mellifera*. Appl Environ Microbiol. 2014; 80:7378–7387. [PubMed: 25239900]

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013; 41:590–596.

R Development Core Team. R: A language and environment for statistical computing. The R Foundation for Statistical Computing; Vienna: 2008.

Stackebrandt E, Stackebrandt E, Frederiksen W, Frederiksen W, Garrity GM, Garrity GM, et al. Report of the ad hoc committee for the reevaluation of the species definition in bacteriology. Int J Syst Evol Microbiol. 2002:1043–1047. [PubMed: 12054223]

Tikhonov M, Leach RW, Wingreen NS. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. ISME J. 2014; 9:1–13. [PubMed: 24936764]

Walter J, Ley R. The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. Annu Rev Microbiol. 2011; 65:411–429. [PubMed: 21682646]

Wang Y, Qian PY. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS One. 2009; 4doi: 10.1371/journal.pone.0007401

Zhu A, Sunagawa S, Mende DR, Bork P. Inter-individual differences in the gene content of human gut bacterial species. Genome Biol. 2015; 16:82. [PubMed: 25896518]

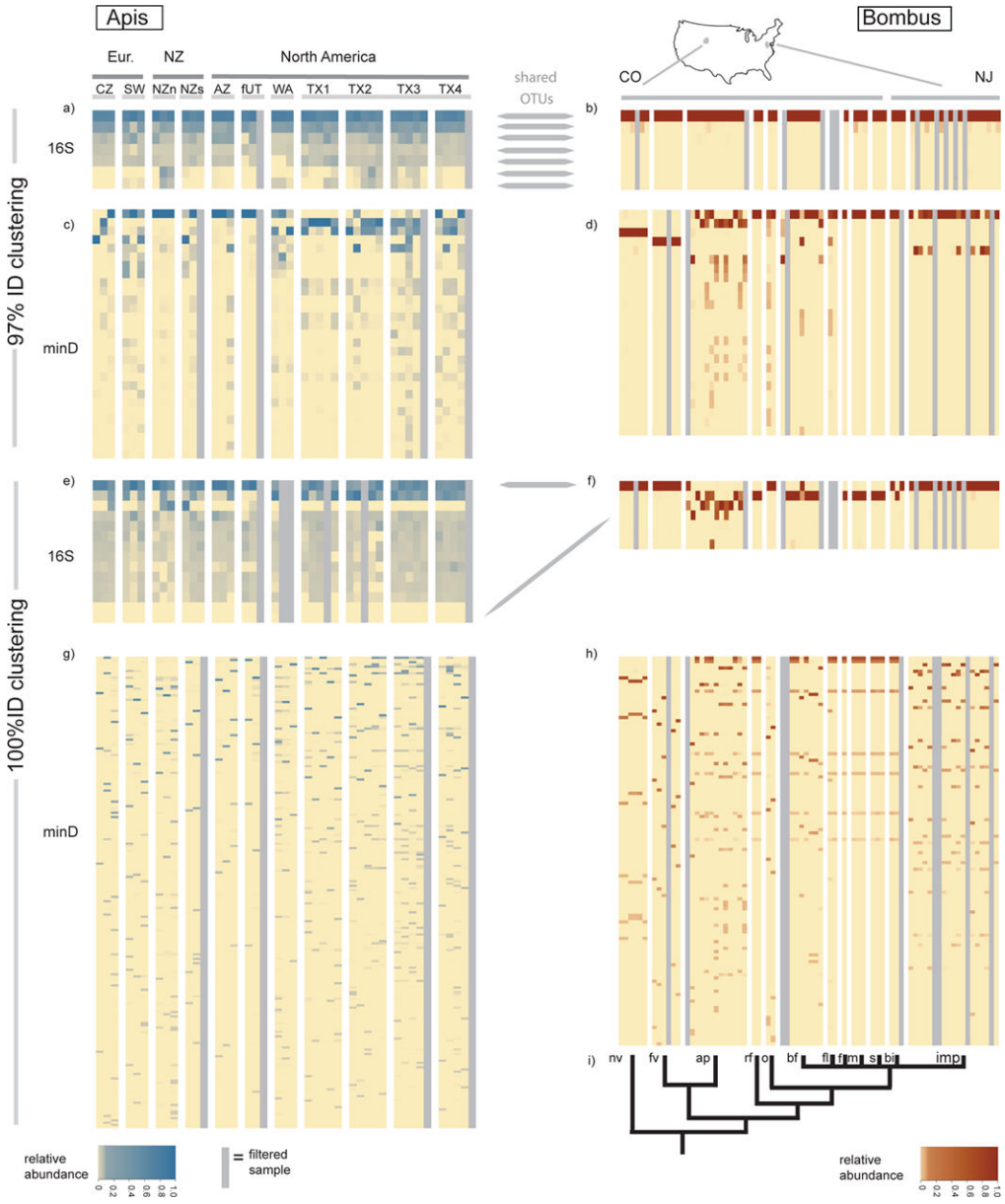

Figure 1.

Relative abundances of *S. alvi* OTUs assigned by clustering at the 97% or 100% threshold based on the 16S rDNA V4 region or on the *minD* coding region. Each column represents one bee, and intensity of color indicates relative abundance of an OTU within that bee. For *A. mellifera*, columns are grouped by locality; Texas samples are further separated by hive of origin. For *Bombus*, columns are grouped by species. Grey columns indicate samples for which all OTUs contained fewer than 90 reads. Grey arrows indicate cases in which OTUs were shared between *A. mellifera* and *Bombus.* The phylogram indicates relationships of *Bombus* subgenera based on (Cameron *et al.*, 2007). The species are abbreviated as follows: nv = *B. nevadensis*, fv = *B. fervidus*, ap = *B. appositus*, rf = *B. rufocinctus*, o = *B. occidentalis*, bf = *B. bifarius*, fl = *B. flavifrons*, f = *B. frigidus*, m = *B. mixtus*, s = *B. sylvicola*, bi = *B. bimaculatus*, imp = *B. impatiens*.

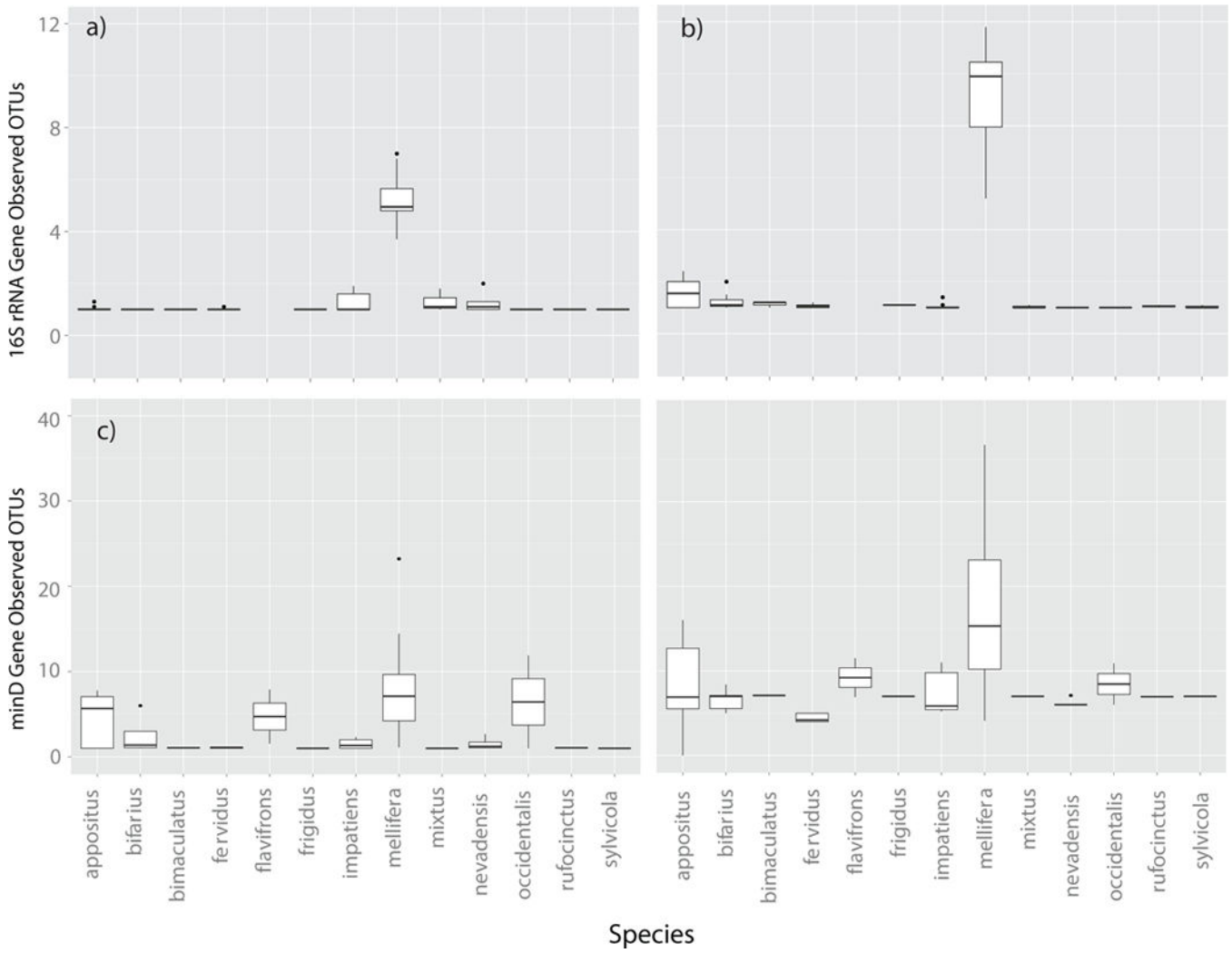

Figure 2.

Average number of observed OTUs by bee host species from 10 subsamplings based on a) 16S *S. alvi*-filtered data at 97% clustering and 100 sequence read depth, b) 16S *S. alvi*-filtered data at 100% clustering and 100 sequence read depth. c) Average observed OTUs for *minD* data at 97% clustering and 500 sequence read depth. d) Average observed OTUs for *minD* data at 100% clustering and 500 sequence read depth. Levels of OTU richness for multiple *Bombus* species are higher for *minD* than for 16S when compared at the same clustering level.

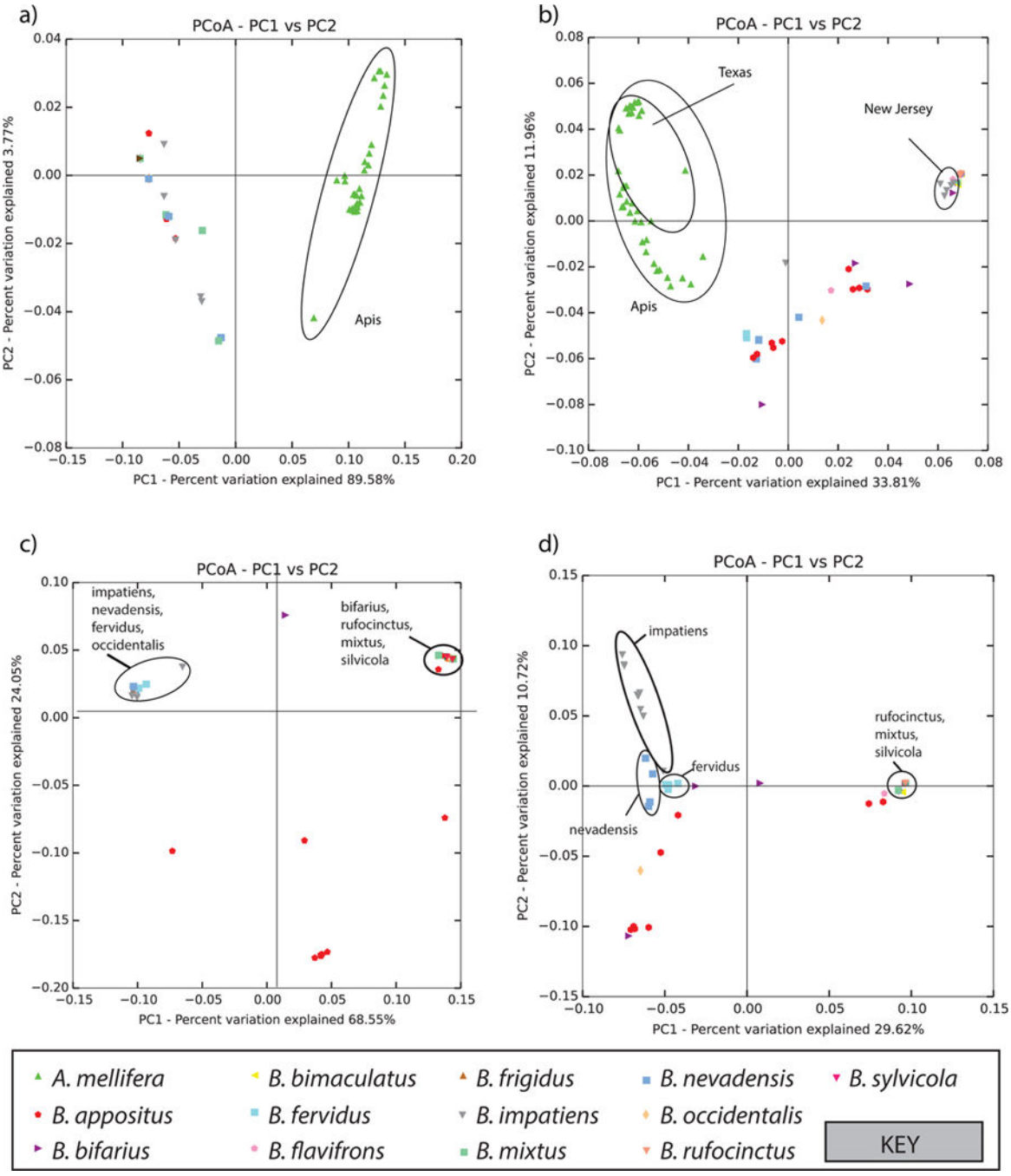

Figure 3.

PCoA plots of *S. alvi* strain composition of *A. mellifera* and *Bombus* (a and b) and of *Bombus* only (c and d). a) 16S rRNA gene clustered at 97% identity. *Apis* and *Bombus* samples cluster separately, but *Bombus* samples do not cluster by species, geographic location or sampling site. b) *minD* clustered at 97% identity. *Apis* and *Bombus* samples cluster separately; furthermore, *Apis* Texas samples cluster strongly as do *Bombus* New Jersey samples (*B. impatiens* and *B. bimaculatus*). c) 16S rRNA gene at 100% identity, for *Bombus* samples only. Clusters include samples from multiple host species. d) *minD*

sequences at 100% identity, for *Bombus* samples only. Most host species form distinct clusters, revealing species-specificity of *S. alvi* strains.

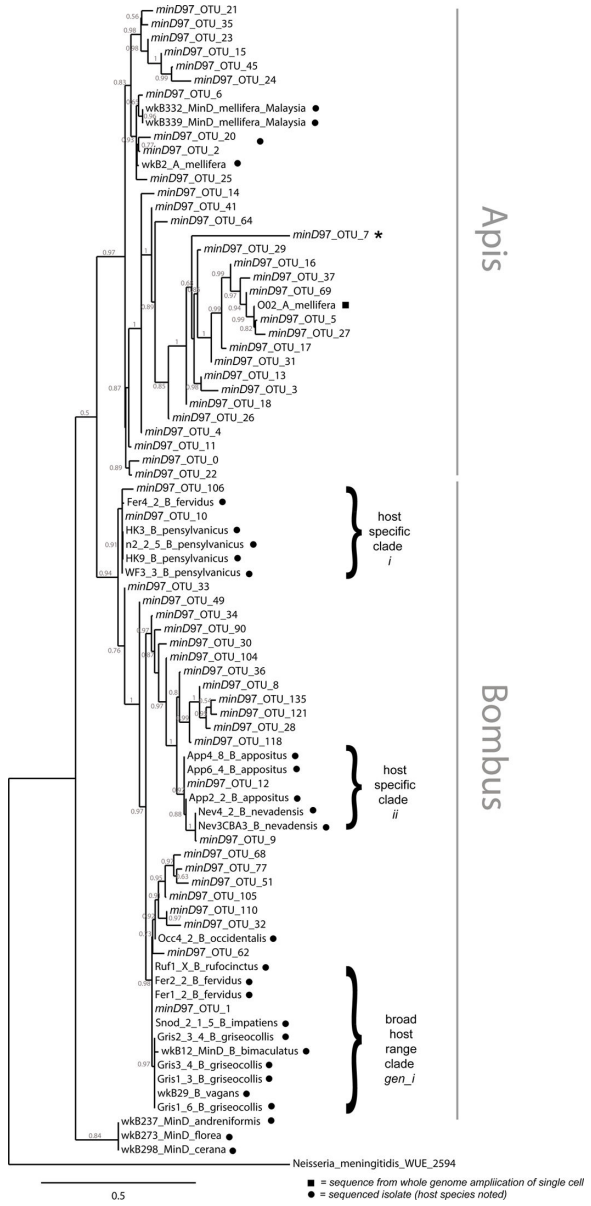

Figure 4.

Maximum likelihood tree inferred from representative *minD* sequences from 97% OTUs plus sequences derived from genomic sequences of isolates. The host species is included in each isolate's title. *Apis*- and *Bombus*-associated OTUs and sequences form two distinct clades with the exception of the *B. impatiens*-associated *minD* 97_OTU_7 (marked "*") within the *Apis* grouping. The *Bombus*-related groups contain several host-specific clusters (noted as clusters *i*, and *ii*.) as well as a lineage that appears to be a generalist across species and subgenera (noted as broad host range clade *gen_i*). Bootstrap values above 0.5 are indicated.

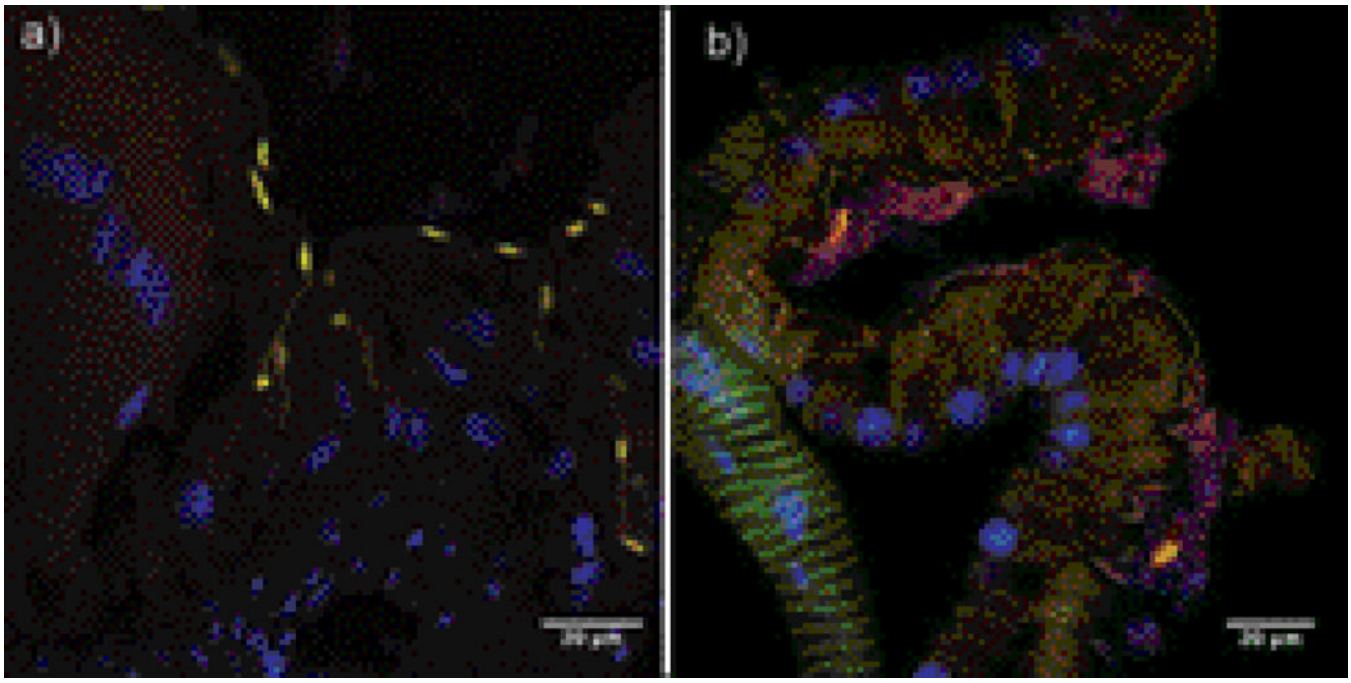

Figure 5.

Strain-specific *S. alvi* FISH probes of individual and combined strains in *A. mellifera* ileums. a) Bee ileum inoculated with strain wkB2 only (yellow, Cy3 probe) forming small assemblages on the epithelial surface and within the folds of the ileum. Host nuclei are stained with DAPI (blue). b) Biofilm from ileum of a bee inoculated with both strains (wkB2, yellow and PE78 Texas Red, red). Host nuclei are stained with DAPI (blue). These two strains form a composite biofilm on the epithelial surface. Host tissue, specifically the trachea in the lower left of the image, has strong auto fluorescence and appears as green.