Comparative Genomics of Wild Bee and Flower Isolated *Lactobacillus* Reveals Potential Adaptation to the Bee Host

Hoang Q. Vuong[1,2] and Quinn S. McFrederick[1,*]

[1]Department of Entomology, University California Riverside
[2]Department of Plant Pathology and Microbiology, University California Riverside

*Corresponding author: E-mail: quinnmc@ucr.edu.

Abstract

Symbiosis with bacteria is common across insects, resulting in adaptive host phenotypes. The recently described bacterial symbionts *Lactobacillus micheneri*, *Lactobacillus timberlakei*, and *Lactobacillus quenuiae* are found in wild bee pollen provisions, bee guts, and flowers but have small genomes in comparison to other lactobacilli. We sequenced, assembled, and analyzed 27 new *L. micheneri* clade genomes to identify their possible ecological functions in flower and bee hosts. We determined possible key functions for the *L. micheneri* clade by identifying genes under positive selection, balancing selection, genes gained or lost, and population structure. A host adherence factor shows signatures of positive selection, whereas other orthologous copies are variable within the *L. micheneri* clade. The host adherence factors serve as strong evidence that these lactobacilli are adapted to animal hosts as their targets are found in the digestive tract of insects. Next, the *L. micheneri* clade is adapted toward a nutrient-rich environment, corroborating observations of where *L. micheneri* is most abundant. Additionally, genes involved in osmotolerance, pH tolerance, temperature resistance, detoxification, and oxidative stress response show signatures of selection that allow these bacteria to thrive in pollen and nectar masses in bee nests and in the bee gut. Altogether, these findings not only suggest that the *L. micheneri* clade is primarily adapted to the wild bee gut but also exhibit genomic features that would be beneficial to survival in flowers.

Key words: *Lactobacillus*, symbiosis, bees, host–microbe, population genomics, comparative genomics.

Introduction

Bees are important for wild and agricultural ecosystems but are also in decline (Potts et al. 2010; Burkle et al. 2013). Research on the bee microbiome has increased in the past decade, due to the new interest in bee health (Engel et al. 2016). Eusocial corbiculate bees, a monophyletic clade of bees in family Apidae which notably includes bumble bees and honey bees, have a specific core microbiome that is important for bee health (Martinson et al. 2011; Li et al. 2015; Kwong et al. 2017). The core microbes confer pathogen defense by inducing host immune functions and directly antagonizing pathogenic microorganisms (Vojvodic et al. 2013; Engel et al. 2016). Strains of *Gilliamella apicola* have the capability to metabolize carbohydrates toxic to the host (Zheng et al. 2016). Honey bee and bumble bee core gut lactobacilli ferment a wide variety of simple carbohydrates in the gut, whereas *Snodgrassella* can convert fermented products to pyruvate for metabolism (Kwong, Engel, et al. 2014; Kwong, Mancenido, et al. 2014).

Bumble bees and honey bees serve as powerful models for microbiome studies (Kwong et al. 2017), but findings of bumble bees and honey bees do not apply to all species of bees. Noneusocial apids (family: Apidae) and noncorbiculate bees (families: Andrenidae, Colletidae, Halictidae, Megachilidae, Melittidae, and Stenotritidae) (all hereafter referred to as wild bees) harbor microbiomes with fewer taxa, are less dense with bacteria, and have less consistent communities compared with honey bees and bumble bees (Martinson et al. 2011; McFrederick et al. 2014). For example, the dominant taxon that associates with megachilid bees (Megachilidae) and halictid bees (Halictidae) is the heterofermentative *Lactobacillus micheneri* (McFrederick et al. 2017). Initially observed as a single species, *L. micheneri* has since been described as three distinct species: *L. micheneri*, *Lactobacillus*

quenuiae, and *Lactobacillus timberlakei* that we refer to collectively as the *L. micheneri* clade (McFrederick et al. 2018). To avoid confusion, the *L. micheneri* species will be referred as "*L. micheneri*."

Despite having reduced genomes compared with other lactobacilli, *L. micheneri* clade bacteria associate with multiple flower and wild bee species (McFrederick et al. 2017, 2018). There is no pattern in host-specificity, as evidenced by the detection of *L. micheneri* and *L. timberlakei* in both megachilid and halictid bees, which is a likely result of horizontal transmission (McFrederick et al. 2017). The ecological function and genomic capabilities of the *L. micheneri* clade are, however, unknown. *Lactobacillus kunkeei* is sister to the *L. micheneri* clade, but *L. kunkeei* is absent or at most present in scarce densities in wild bees (McFrederick et al. 2017; Rothman et al. 2018). *Lactobacillus kunkeei* is typically only detected from flowers and the crops and colony surfaces of honey bees and is not considered part of the honey bee core microbiome (Anderson et al. 2013; Kwong et al. 2017). The lactobacilli that are considered part of the core microbiome in honey bees consists of the homofermentative *Lactobacillus apis* and *Lactobacillus mellis* and their relatives, which are distantly related to *L. micheneri* and *L. kunkeei* (McFrederick et al. 2013). Wild bees are predominantly associated with heterofermentative lactobacilli, whereas honey bees are transiently colonized by heterofermentative lactobacilli but colonized at high density with homofermentative lactobacilli. The closest relatives to *L. micheneri* and *L. kunkeei* are lactobacilli that have been isolated from flowers (Kawasaki et al. 2011), fermented vegetable drinks (Chiou et al. 2018), and sourdough starter (Vogel et al. 2011), but are suggested to be associated with insects (Duar et al. 2017). Because *L. micheneri* inhabits a novel niche, it has likely evolved new functions that facilitate its association with flowers and wild bees. We compared 27 new *L. micheneri* clade genomes with publicly available *Lactobacillus* genomes to detect signatures of adaption which led to the association with wild bees. We identified genes under positive selection, balancing selection, genes gained or lost, and population structure. These genes are candidates of important functions involved in colonizing wild bees and surviving in flowers.

Materials and Methods

Sample Collection

We used DNA from cultures originally isolated in McFrederick et al. (2017). We isolated seven samples from *Megachile rotundata* pollen provisions collected in Wellsville, UT, by plating a phosphate buffered saline serial dilution of pollen provisions on de Man Rogosa Sharpe (MRS, De Man et al. 1960) plates fortified with 2% or 20% fructose. After plating the pollen provisions, we purified single isolates by subculturing individual colonies three successive times. We extracted DNA from the pure cultures using the Qiagen Blood and Tissue kit protocol with lysozyme.

Genome Sequencing, Annotation, Assembly, and Reference Genome Sequence Access

To prepare sequencing libraries, we used the Illumina Tru-Seq DNA PCR-free kit, following the manufacturer's instructions. For libraries with low DNA concentrations, we used Illumina's universal primers to amplify libraries in a PCR programed for 98 °C for 3 min and 10 cycles of 98 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s with a final extension at 72 °C for 5 min. Once normalized, we sequenced the genome libraries with the Illumina MiSeq with 2x300 V3 reagents. After sequencing, we used the A5 pipeline (ver: 05222015) to assemble reads into contigs and scaffolds (Coil et al. 2015). Once assembled, we annotated our genomes using RAST (Aziz et al. 2008). We used CheckM on the Kbase browser-based tool to assess the completeness and contamination of the draft genomes (Parks et al. 2015; Arkin et al. 2018; McFrederick et al. 2018). After annotation, we aligned all *L. micheneri* clade genomes from McFrederick et al. (2018) and this study using SPINE and ClustAGE to identify accessory and core genome elements of the clade and each species with multiple isolates sequenced (Ozer et al. 2014; Ozer 2018). We uploaded our RAST annotations to the KEGG automatic annotation server (Kanehisa and Goto 2000; Moriya et al. 2007; Minoru Kanehisa et al. 2016). We accessed relative *Lactobacillus* genomes ($N = 22$) with the Pathosystems Resource Integration Center (PATRIC) database (Wattam et al. 2017). We used the Shapiro–Wilk test and Mann–Whitney–Wilcoxon test for pairwise comparisons of nonnormally distributed sample sets for GC, genome size and coding DNA sequence (CDS) in the R base software package (R Core Team 2018). Genomic feature graphs were created using the R package ggpubr (Kassambara 2017).

Ortholog Detection and Verification

We used OrthoMCL to detect orthologs in the 52 genomes using a recommended inflation value of 1.5 (Li et al. 2003). We used our 30 *L. micheneri* clade genomes along with 22 related lactobacilli. Using *Lactobacillus plantarum* as an outgroup, we reconstructed the phylogenetic history of the focal lactobacilli using the concatenated supermatrix of single copy orthologs ($N = 583$ orthologs) with RAxML (Stamatakis 2014) and the GTRGAMMA model for the entire supermatrix. We verified all genes by modeling protein structure with SWISS-model (Waterhouse et al. 2018) and BlastP searches against the UniProt database (UniProt Consortium 2017).

Ancestral Genome State Reconstruction

We analyzed the ortholog and singleton gene frequency table of the 52 genomes to reconstruct the ancestral genome state using

the program Count with the Wagner Parsimony assumption (Gene Gain Penalty = 2) (Csurös 2010). In addition to reconstructing the ancestral genome state of each clade of the core genome phylogeny (supplementary fig. 1A and B, Supplementary Material online), Count determined the genes that were gained or lost between each clade and branch tip. Using a custom Bash parser, we cross-referenced genes gained or lost between each clade and branch tip with the annotation.

Detecting Positive Selection in the *L. micheneri* Clade Using the Branch-Site Model

We searched ortholog sequences for nonsynonymous mutations which serve as signatures of positive selection using the PAML package codeML (Yang 2007). We searched for orthologs with a significantly better fit to models that include positive selection compared with the null, neutral selection. For each ortholog, we compared the likelihood ratio values of the two models to determine whether the resulting comparison yielded a significant *P* value. If the selection model significantly fit positive selection for the ortholog of interest, we identified amino acid sites under positive selection using the Bayesian empirical Bayes inference (Zhang et al. 2005). We used a phylogeny containing all *Lactobacillus* genomes in this study to search the ortholog sequences for sites under positive selection in the *L. micheneri clade* (supplementary fig. 1A and B, Supplementary Material online). Next, we searched for ortholog sequences with sites under positive selection between *L. micheneri* and *L. timberlakei* and the rest of the *L. micheneri* clade with *L. quenuiae* as the outgroup.

Allele Frequency and Population Structure Analyses

We retrieved all orthologs from *L. micheneri* and *L. timberlakei*, for which we have multiple isolates. We analyzed the allele frequency of orthologs of the *L. micheneri* clade species with DNAsp (Librado and Rozas 2009). We created a single nucleotide polymorphism (SNP) map in the VCF format using Bowtie2, Samtools, Bcftools, and Vcftools to analyze the population structure across the *L. micheneri* Hlig3 genome as a reference (Li et al. 2009; Danecek et al. 2011; Langmead and Salzberg 2012). Next, we inputted the SNP map into the R package "PopGenome" and used the sliding window method (1-kb windows and steps) to calculate F_{st} (Pfeifer et al. 2014). We excluded the genome of the type strain of *L. quenuiae* as we have isolated only one strain of *L. quenuiae* so far, which would not allow us to make accurate inferences in population structure. We calculated F_{st} for each contig containing variant calls between *L. micheneri* and *L. timberlakei* genomes. Based on prior usage and empirical determination (Zhang et al. 2015; Hardigan et al. 2017), we selected windows that have an F_{st} above the 95th percentile or 1,000-bp windows and steps with an F_{st} of at least 0.932 between *L. micheneri* and *L. timberlakei*.

Results

Draft Genome Quality and Genome Features of *L. micheneri* Clade Bacteria

Based on the CheckM results, we find that all the draft genomes have a completeness of 95% or greater. All but three genomes had no reported contamination, only HV_6, HV_10, and HV_23 had low contamination of 0.062%, 0.053%, and 1.5%, respectively. We tested all genomic feature data for normality using the Shapiro–Wilk test. The Shapiro–Wilk test indicated that all but two data sets were nonnormal distributions (table 1). *Lactobacillus micheneri* clade genomes ($N = 30$) were significantly more AT-biased ($P < 0.001$, Mann–Whitney–Wilcoxon test, fig. 1A and table 1) than the genomes of sister taxa *L. kunkeei*. *Lactobacillus micheneri* had a significantly larger number of CDSs per Mb than *L. kunkeei* genomes, because there was significant difference in the number of CDS ($P < 0.001$, Mann–Whitney–Wilcoxon test; fig. 1B and table 1A) and genome size ($P < 0.05$, Mann–Whitney–Wilcoxon test; fig. 1B and table 1A) between the *L. micheneri* clade and *L. kunkeei*. *Lactobacillus timberlakei* ($N = 9$) had significantly larger genomes and significantly higher counts of CDS than *L. micheneri* (Genomes: $P < 0.001$; CDS: $P < 0.001$, Mann–Whitney–Wilcoxon test; fig. 1C and D and table 1). Notable accessory genome features from the CLUSTAGE analysis in the *L. micheneri* clade include a polysaccharide biosynthesis gene cluster only in *L. timberlakei* (supplementary table 2, Supplementary Material online; accessory genome element 15 and 51). Other features include one chitinase that is annotated as a fibronectin/fibrinogen-binding protein in only *L. micheneri* HV_63, HV_64, HV_67, and *L. quenuiae* type HV_6 (supplementary table 2, Supplementary Material online; accessory genome element 99.1, 25.10), and a pectate lyase copy present only in *L. micheneri* and *L. quenuiae* (supplementary table 2, Supplementary Material online; accessory genome element 97.7).

Ortholog Detection and Core Ortholog Phylogeny of Distant and Close Relatives of *L. micheneri*

OrthoMCL detected 969 single copy orthologs shared across the 30 genomes from the *L. micheneri* clade. In addition, 1,118 and 1,092 orthologs were shared within the *L. micheneri* and *L. timberlakei* species, respectively. When comparing the *L. micheneri* clade with close relatives such as *L. kunkeei* and distant relatives such as the *L. plantarum* clade we detected 583 orthologs. We aligned and concatenated the amino acid sequence of these 583 orthologs to create a supermatrix for input to RAxML to construct an ortholog phylogeny of 11 *Lactobacillus* species (supplementary fig. 1A and B, Supplementary Material online). Of the 583 orthologs, 32 orthologs did not pass alignment for the branch-site analysis, leading to only 551 orthologs analyzed. The resulting phylogeny showed 100% bootstrap support of the three species as

Table 1

Statistical Analyses of Genomic Features: GC% (*A*), CDS (*B*), and Genome Size (*C*) between *L. kunkeei* and the *L. micheneri* Clade; *L. micheneri* and *L. timberlakei*

| Compared Species/Groups	Comparison Mean	Shapiro–Wilk P Value	Mann–Whitney W	P Value
(*A*) GC%				
L. kunkeei (*N* = 15) and *L. micheneri* clade (*N* = 30)	36.8%, 30.5%	0.0002771, 0.0004398	480	2.238e-08
L. micheneri species (*N* = 20) and *L. timberlakei* (*N* = 9)	30.5%, 30.4%	0.0001767, 0.008338	135	0.02629
(*B*) CDS				
L. kunkeei and *L. micheneri* clade	1,407, 1,540	0.4868, 0.0007096	29.5	1.276e-06
L. micheneri species and *L. timberlakei*	1,502, 1,619	0.04671, 0.006994	11	3.834e-05
(*C*) Genome size				
L. kunkeei and *L. micheneri* clade	1,536,381, 1,500,683	0.1275, 7.966e-05	341	0.02045
L. micheneri species and *L. timberlakei*	1,468,371, 1,567,777	0.0005775, 0.02314	18.5	8.158e-04
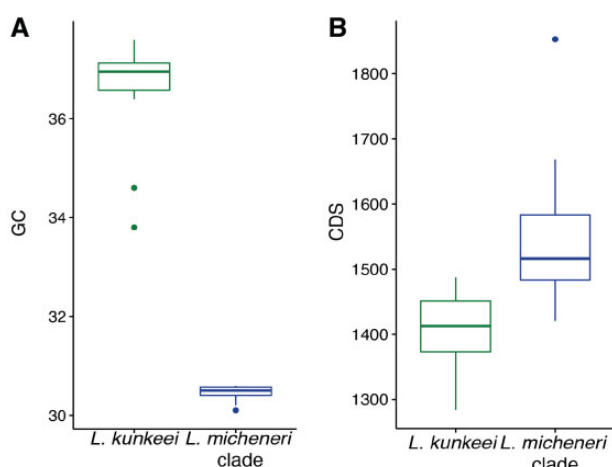
Fig. 1.—*Lactobacillus micheneri* clade bacteria significantly differ from *L. kunkeei* in genomic GC% (*A*) and CDS (*B*). Means and *P* values are in table 1.

monophyletic groups (supplementary fig. 1*A* and *B*, Supplementary Material online).

Gene Gain and Loss Reconstruction

We used the phylogeny created from the ortholog analyses as input to reconstruct gene gain and loss (supplementary fig. 1*A* and *B*, Supplementary Material online). The gain or loss of genes may represent an important adaptation to the bee or flower host environments for *L. micheneri* via a loss or gain of possible function. The common ancestor of the *L. micheneri* clade gained 115 genes and lost 50 genes. The 115 genes gained included pectate lyase, which has been biochemically characterized as a bioactive extracellular enzyme in *Lactobacillus* (Sakellaris et al. 1989). Other examples include two copies of beta-hexosaminidase identified as a dispersin B (dspB) (Donelli et al. 2007), a biphenyl degradation gene (Ishigooka et al. 1986), and a cobalt–zinc–cadmium resistance gene involved in heavy metal transport (Zhai et al. 2017) (fig. 2 and supplementary table 3, Supplementary Material online). Compared with the common ancestor to the *L. micheneri* clade, *L. quenuiae* gained 158 genes, including a

gene annotated by UniProt and SWISS-model as a fibronectin/fibrinogen-binding protein, and lost 31 genes (fig. 2 and supplementary table 3, Supplementary Material online). The common ancestor of *L. timberlakei* gained 31 genes, including the YoeB/YefM toxin anti-toxin system and genes involved in the persistence of cells in the presence of antibiotics and the absence of nutrients (Wang and Wood 2011) (fig. 2 and supplementary table 3, Supplementary Material online).

Metabolism and Biosynthetic Pathway Reconstruction of the *L. micheneri* Clade

Using the KEGG annotations, we analyzed the biosynthetic and metabolic pathways of the type strains of *L. micheneri*, *L. timberlakei*, and *L. quenuiae* described in McFrederick et al. (2018) and outgroups. We found several consistent patterns across the *L. micheneri* clade. 1) There was little to no hexose metabolism outside of glucose, sucrose, and fructose as maltose and sucrose phosphorylases are absent (table 2 and supplementary fig. 2*A*, Supplementary Material online). 2) Although the pathway to produce tyrosine or phenylalanine is incomplete, the *L. timberlakei* type strain can convert 3-dehydroquinate to 4-hydroxyphenyl pyruvate, an intermediate that is a potential lactic acid fermentation substrate (supplementary fig. 2*B*, Supplementary Material online). However, the other two *L. micheneri* clade type strains can only convert chorismate to prephenate (supplementary fig. 2*B*, Supplementary Material online). 3) None of the type strains has the full pathway to synthesize thiamine from purine or biosynthesize biotin (supplementary fig. 2*C* and *D*, Supplementary Material online).

Population Structure Analyses of *L. micheneri* and *L. timberlakei*

During our population structure analysis, we removed four *L. timberlakei* genomes (HV_04, HV_09, HV_12, and HV_27) from the overall data set. Although these genomes were assembled with a small number of contigs (15–29) and acceptable coverage (15.31–31.36), they had the fewest number of error-corrected reads (112,820–272,407) and percentage of
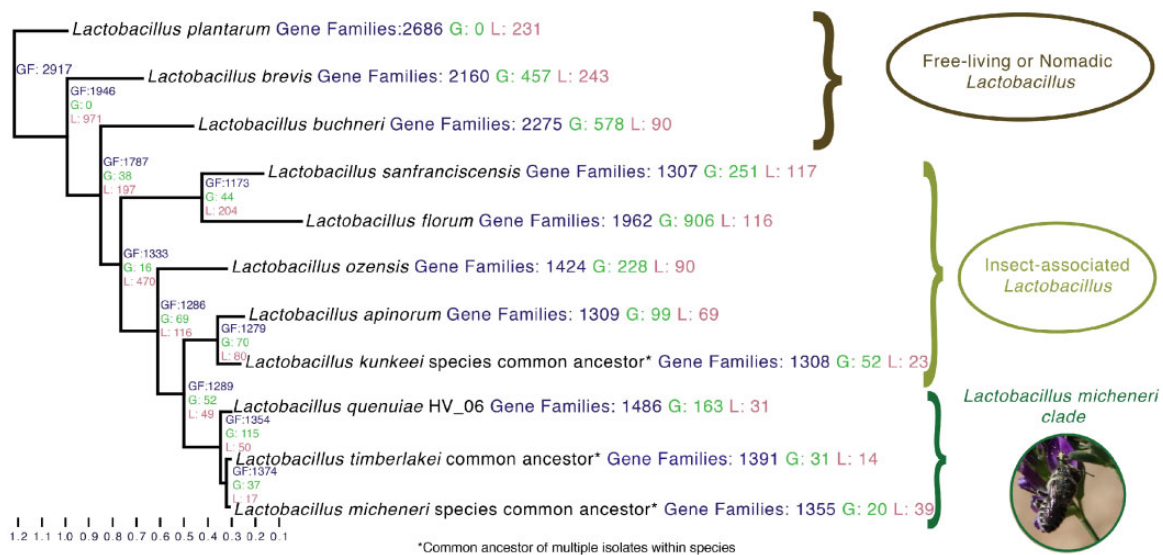
Fig. 2.—Genome reduction from the common ancestor of *Lactobacillus sanfranciscensis* and *Lactobacillus micheneri* is conserved in the present-day species. Gene gain and loss was calculated with the Wagner Parsimony assumption in the Count program (Gene Gain Penalty = 2). Numbers that are labeled GF (blue) are the total counts of gene families present in the genome or ancestral state. Numbers labeled G (green) and L (red) are the counts of gene families gained and lost in the genome or ancestral state, respectively.

Table 2

Carbohydrate Metabolism Absent (White) or Present (Blue) Summary Based on KEGG and RAST Annotations of *Lactobacillus* Type Strains in Study

	Amylose, and Mannitol	Xylose	Arabinose, Lactose, Cellobiose, and Galactose	Mannose	Trehalose	Dextrin	Maltose	Sucrose, Glucose, and Fructose	Sorbitol	Beta-Hexosanimidase	Pectate Lyase
L. plantarum	■		■					■	■		■ a
L. buchneri			■		■ b	■ b		■	■ b		
L. brevis			■		■ b	■ b		■	■ b		
L. sanfranciscensis				■		■ b		■	■ b		
L. florum								■	■ b	■	■
L. ozensis								■	■ b	■	■
L. apinorum								■	■ b	■	■
L. kunkeei					■		■ a	■	■ b	■	■
L. quenuiae								■	■ b	■	■
L. timberlakei								■	■ b	■	■
L. micheneri								■	■ b	■	■

[a]Found in other genomes of species in study.
[b]Not by KEGG but by RAST annotation.

nucleotides (44.29–63.82%) passing quality filtering via the A5 pipeline. Because our population structure analyses use raw reads instead of assemblies from error-corrected reads, the uncorrected raw reads could lead to a missing variant call at a variant site. Including these genomes could, therefore, lead to inaccurate F_{st} readings. With the remaining genomes, we calculated F_{st} of *L. timberlakei* and *L. micheneri* to detect diverging genes between two species. F_{st} is the fixation index between a subpopulation and the total population or two populations, at a single SNP or a nucleotide range or window. An F_{st} value of 1 means that the SNP is fixed in one population and not in another. An F_{st} value of 0 indicates no differentiation between the two populations. Genes with F_{st} values above the 95th percentile are of interest as their respective function may be under selective pressure resulting in divergence between *L. micheneri* and *L. timberlakei*. We detected eleven 1-kbp windows where $F_{st} = 1$, which include the coding sequence of seven protein-coding genes and one promoter site (fig. 3 and supplementary table 4, Supplementary Material online). The protein-coding genes where $F_{st} = 1$ include a kup K$^+$ uptake system, a gene involved in osmoregulation, and pH tolerance (Trchounian and Kobayashi 1999).
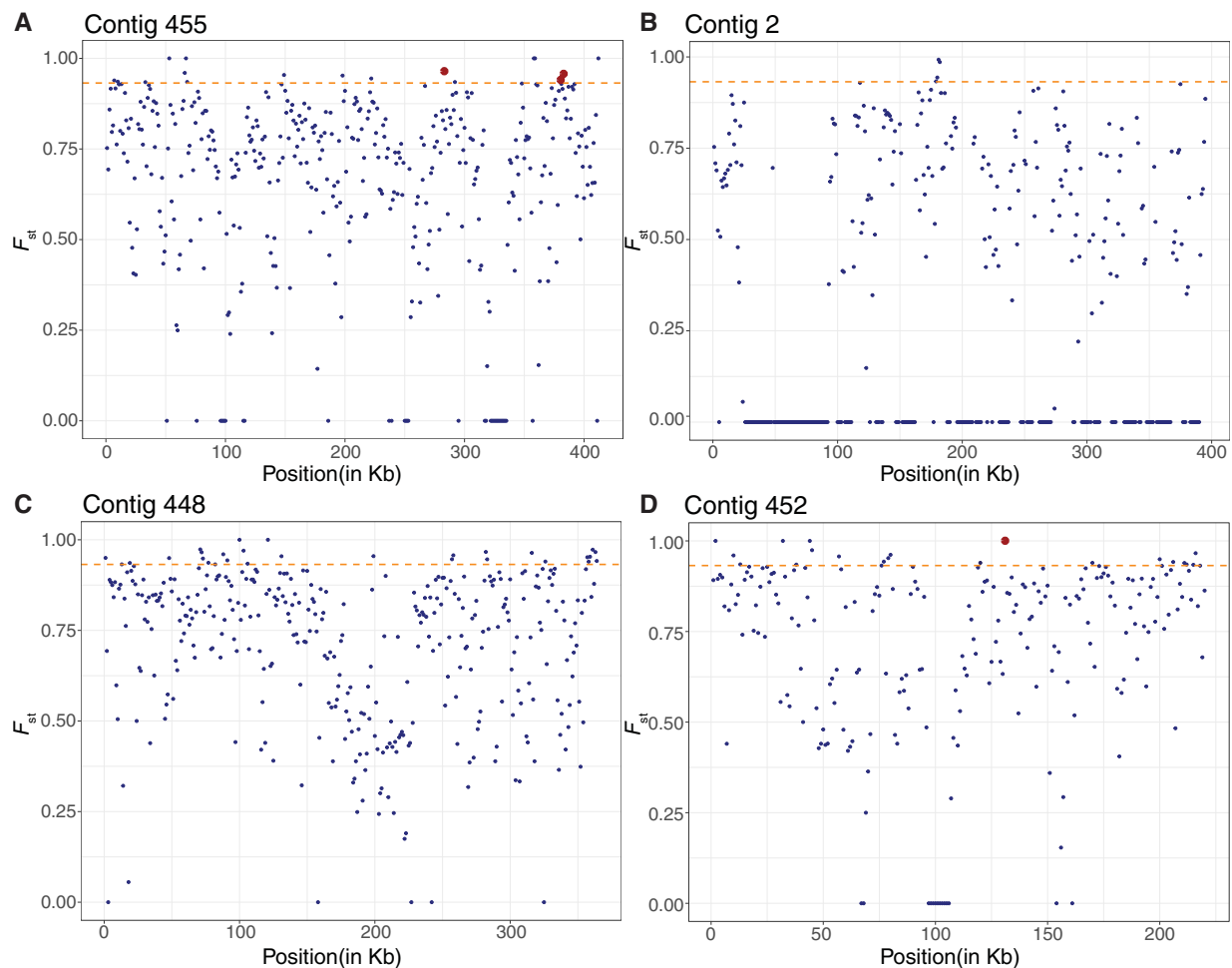
Fig. 3.—Population structure analyses in 1-kb windows and 1-kb steps between *Lactobacillus micheneri* and *Lactobacillus timberlakei* mapped to four *Lactobacillus micheneri* Hlig3 contigs reveal strong population structure. The orange dashed line represents the threshold for the 1-kb windows with the 95th percentile F_{st} (0.932). Red dots represent gene regions discussed in text, listed from left to right as: LtaS, cshA, MurR, and kup. Regions where F_{st} above the 95th percentile is listed in supplementary table 4, Supplementary Material online.

We retrieved 86 genes from the 70 windows that were in the 95th percentile of F_{st} values. These genes include a cold-shock DEAD-box protein A (cshA) a temperature sensitive regulator (Hunger et al. 2006), lipoteichoic acid synthase type IIb (ltaS), and UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopi-melate–D-alanyl-D-alanine ligase (murF) (supplementary table 4, Supplementary Material online).

Lactobacillus micheneri Clade Species Ortholog Allele Frequency Analyses

We calculated Tajima's *D* for each ortholog within *L. miche-neri* and *L. timberlakei* (Librado and Rozas 2009). A significant positive Tajima's *D* score suggests that the respective ortholog is undergoing balancing selection, where more than expected alleles are maintained in the population. Alternatively, genome-wide significant positive Tajima's *D* scores suggest that the population has suddenly contracted or is splitting.

A negative Tajima's *D* score suggests that the population is either recovering from a bottleneck (genome-wide) or that the respective gene has recently undergone a selective sweep, where fewer than expected alleles are maintained. The 30 *L. timberlakei* orthologs with a significantly negative Tajima's *D* score include glutamine transport ATP-binding protein (glnQ), a transporter involved in osmotolerance and pH resistance (Cotter and Hill 2003) and DEAD-box ATP-dependent RNA helicase (cshB), a temperature sensitive regulator (Hunger et al. 2006) (supplementary table 5, Supplementary Material online). The 125 orthologs with a significant positive Tajima's *D* score in *L. timberlakei* include the glycine betaine ABC trans-port system (opuAC) involved in osmotolerance and pH resis-tance (Papadimitriou et al. 2016) (supplementary table 5, Supplementary Material online). The 473 orthologs of *L. micheneri* with a significant negative Tajima's *D* score includes betaine transport protein (betT) involved in glycine transport for osmotolerance and pH resistance. There were orthologs

with significant positive Tajima's *D* scores in both *L. timberlakei* and *L. micheneri* (supplementary table 5, Supplementary Material online). These genes include the glycine betaine ABC transport system genes (opuAA\AB) involved in osmotolerance and pH resistance (Papadimitriou et al. 2016) (supplementary table 5, Supplementary Material online).

Positive Selection Detection Using the Branch-Site Model

Across the *L. micheneri* clade, 108 genes out of 551 tested orthologs showed signatures of positive selection (supplementary table 6, Supplementary Material online). Fixation of nonsynonymous mutations is rare as amino acid sequences are conserved under purifying selection. However, amino acid sequence changes can result from selection pressure arising from adapting to novel or changing environments. Genes in the *L. micheneri* clade exhibiting signatures of positive selection may, therefore, be important in adaptation to the wild bee and/or flower niche. Genes undergoing positive selection in only *L. micheneri* or *L. timberlakei* may also be important for niche differentiation between these two species. The DEAD-box ATP-dependent RNA helicase (cshA), a gene involved in temperature sensitive gene regulation (Hunger et al. 2006), had the lowest *P* value from the positive selection analysis (supplementary table 6, Supplementary Material online). The other 107 genes included 2,3-butanediol dehydrogenase, a protein involved in producing a volatile important for resisting gut acidification in honey bee core lactobacilli (Lee et al. 2015) and fibronectin/fibrinogen-binding protein (supplementary table 6, Supplementary Material online). When searching for orthologs containing sites under positive selection in the genomes of *L. micheneri* and *L. timberlakei* we found 8 and 12 orthologs, respectively (supplementary table 6, Supplementary Material online). In *L. micheneri*, the eight genes include the cadmium, zinc, and mercury transporting ATPase (copB) (Solioz et al. 2010) (supplementary table 6, Supplementary Material online). In *L. timberlakei*, the genes included heavy metal transporter manganese transport protein mntH (Makui et al. 2000) and a heme-dependent catalase that is orthologous to a heme-dependent catalase involved in reactive oxidative stress in *L. plantarum* (Knauf et al. 1992) (supplementary table 6, Supplementary Material online).

Discussion

The genomes of the *L. micheneri* clade show signatures of selection and novel functions that underlie the ability of these bacteria to thrive in nutrient-rich environments. Given the small size of their genomes, the ability of these bacteria to persist on flowers and proliferate in pollen and nectar masses inside bee nests and inside the bee gut is remarkable. Although all these environments are rich in plant metabolites, the bee gut immune system exerts further pressure. It is

therefore perhaps not surprising that their genomes reveal adaptations to harsh, nutrient-rich environments as well as animal host-specific adaptations (supplementary table 7, Supplementary Material online). Compared with the honey bee-associated lactobacilli, the *L. micheneri* genomes are more reminiscent of pathogens or endosymbionts with more compact and more AT-biased genomes with less carbohydrate metabolism capabilities (fig. 1 and table 1) (Ellegaard et al. 2015; Zheng et al. 2016).

Although ours is the first comparative genomics study of wild bee-associated lactobacilli, previous work on honey bee core microbes and *L. kunkeei* allows interesting contrasts to our results. For example, Tamarit et al. (2015) found that *L. kunkeei* has short generation times. Ellegard et al. (2015) found that the honey bee gut lactobacilli genomes are rich in carbohydrate metabolism and transport genes. *Gilliamella* can utilize a variety of monosaccharides that are toxic to honey bees (Zheng et al. 2016) and *Gilliamella* and *Snodgrasella* may work together to provide nutrients to their host (Kwong 2014).

Honey bees and wild bees share common ancestry but differ in their life histories. Although honey bees are active year-round, wild bees often have short periods of adult activity and spend unsuitable seasons in a quiescent state (Michener 2007). The phenology of wild bees and flowers may therefore cause a bottleneck which reduces gene flow occurring in the *L. micheneri* clade. Although not an adaptation, this may be why the *L. micheneri* clade genomes are more AT-rich than *L. kunkeei* and core honey bee gut lactobacilli (Ellegaard et al. 2015). In addition to AT-biases, the *L. micheneri* clade have more compact genomes than core honey bee gut lactobacilli (Ellegaard et al. 2015). Additionally, compared with within-colony, social transmission in honey bees (Kwong and Moran 2016) horizontal transmission occurring at flowers may be more important for maintenance of *L. micheneri* in wild bee communities which include both solitary and primitively eusocial host taxa (McFrederick et al. 2017).

These similarities and differences in host biology may explain why we find both convergence and novelty in symbiont functions by honey bee microbes and the *L. micheneri* clade. *Gilliamella* sp. has pectate lyase genes which allow them to digest pollen intines and increase honey bee worker weight (Zheng et al. 2017). The presence of pectate lyase in the *L. micheneri* clade may mean that these bacteria are important for wild bee nutrition, as nearly all female bees digest pollen to build protein reserves necessary for egg production (Cane 2016; Cane et al. 2017). *Gilliamella* sp., however, has more enzymes for digesting nectar and pollen saccharides compared with the *L. micheneri* clade (Engel et al. 2012; Zheng et al. 2016) and the *L. micheneri* clade is depauperate in regards to carbohydrate metabolism genes.

Inhibition of pathogenic organisms observed in *L. kunkeei* (Vásquez and Olofsson 2009; Forsgren et al. 2010) may be

possible in the *L. micheneri* clade. dspB may be used to inhibit unwanted biofilm-forming bacteria or pathogens in synthetic environments (Donelli et al. 2007) and *L. micheneri* clade bacteria may use dspB to regulate the dispersal of their biofilms in response to environmental stressors such as reactive oxygen species (ROS) produced by competitors or hosts (Stacy et al. 2014). Fungi may be inhibited by the *L. micheneri* clade by D- and L-lactic acid production and can be enhanced by hydroxyphenyllactate and phenyllactate in *L. timberlakei*. These compounds are all inhibitory to fungi and inhibition has been documented in *Lactobacillus sanfranciscensis* (Lavermicocca et al. 2000) and *L. kunkeei* (Vojvodic et al. 2012). The *L. micheneri* clade has gained, lost, and refined functions that are likely behind its ability to thrive in the new wild bee guts and pollen provisions.

Host Binding and Adherence Factors

In silico prediction and annotation of the genes involved in host binding and adherence factors provides evidence supporting the hypothesis that the *L. micheneri* clade is adapted to the bee gut. The *L. micheneri* clade and their outgroups have genes predicted to be fibronectin/fibrinogen-binding proteins that target animal specific proteins, suggesting that they are more likely to have a role in animal association (Henderson et al. 2011). Specifically, in insects, fibrinogen is known to be present in the peritrophic matrix of the midgut (Hu et al. 2012), whereas both fibrinogen and fibronectin transcripts are expressed in the hind gut (Neira Oviedo et al. 2008). The fibronectin binding gene of the *L. micheneri* clade is under positive selection, potentially improving *L. micheneri* clade interaction with bees compared with the outgroups. Copies of fibronectin/fibrinogen-binding protein are also variable in the *L. micheneri* clade, which may improve their ability to establish in certain strains of animal hosts (Buck et al. 2005). The production of polysaccharides or exopolysaccharides in *L. timberlakei* may allow more efficient host immune evasion and adhesion in bees guts compared with *L. quenuiae* or *L. micheneri* (Leigh and Coplin 1992; D'Haeze and Holsters 2004; Lee et al. 2016).

Metabolic Capabilities Correlate to Nectar and Pollen-Rich Niches

Our data suggest that the *L. micheneri* clade thrive best in pollen and nectar-rich environments such as wild bee pollen provisions or wild bee guts. We previously found *L. micheneri* to be abundant in these environments (McFrederick et al. 2017). However, our findings do not exclude the possibility that these adaptations are important to their survival in flowers, where these lactobacilli are also found (McFrederick et al. 2017). The *L. micheneri* clade appears to be streamlined to utilize nectar carbohydrates, as they can only digest sucrose, glucose, and fructose, the three most predominant sugars of nectar (Nicolson 1998). The presence of pectate

lyase genes is also remarkable, as the closest related lactobacilli to contain these genes is *L. plantarum* (Sakellaris et al. 1989).

Adaptation to pH, Osmotolerance, and Temperature in Bee Guts and Flowers

Although *L. micheneri* and *L. timberlakei* are very closely related (McFrederick et al. 2018), we find divergence in genes involved in osmotolerance, pH, and temperature tolerance. Such genes are also involved in response to rapid osmolarity or pH changes. Divergence in these functions suggests that these bacteria have adapted to wild bee guts and pollen provisions. These adaptations can be especially important as different sections of the bee gut have different osmotic potentials and pH (Peng and Dobson 1997). In addition to different osmotic potentials or pHs in bee guts, there are variable osmolarities in different flower nectars (Nicolson 1998). These genes may, therefore, underlie the divergence between *L. micheneri* and *L. timberlakei*.

Temperature-sensitive transcriptional regulators under positive selection may be important as wild bees and flowers do not thermoregulate their nests as honey bees and bumble bees can, leading to temperatures dropping during nights or cold days (Graham and Patterson 1982; Liu et al. 2005). Differences in host biology also correlate with the optimum growth temperature of core honey bee lactobacilli, which is the optimum maintained temperature in honey bee hives (35 °C, Jones et al. 2004; Olofsson et al. 2014), as opposed to 32 °C for the *L. micheneri* clade (McFrederick et al. 2018). Further studies on these genes and their roles in niche differentiation in the *L. micheneri* clade are needed to test these hypotheses.

Detoxification and Oxidative Stress Response

The *L. micheneri* clade appears to be adept at detoxification, as we found signatures of toxin tolerance or mitigation in their genomes. Toxins such as heavy metals and biphenyl can bioaccumulate in flowers and are therefore likely to occur in pollen provisions and bee guts in areas with contaminated soils (Hladun et al. 2015; McFrederick et al. 2017). Although catalase genes are extremely uncommon in lactobacilli, they may be important for the *L. micheneri* clade because both flowers and bees produce ROS that repel microbes (O'Brien et al. 2012; Buchon et al. 2014). Additionally, the signatures of positive selection in *L. timberlakei* in catalase suggest a competitive advantage when exposed to ROS compared with *L. micheneri* or *L. quenuiae*. Finally, *L. timberlakei* may be more competitive than *L. micheneri* or *L. quenuiae* under antibiotic or nutrient-limiting stress due to the additional toxin anti-toxin genes in *L. timberlakei* (Wang and Wood 2011).

Conclusion

Although experimental verification is needed, the genomic data presented here suggest several possible mechanisms by which *L. micheneri* clade bacteria may benefit their bee hosts. Given the largely absent metabolic and biosynthetic capability of the *L. micheneri* clade, it is unlikely that *L. micheneri* clade wild bee associates are biosynthetic mutualists. This is not surprising as all the necessary nutrients are largely present in pollen and nectar (Nicolson 1998; Roulston and Cane 2000). None the less, based on genomic in silico predictions, the *L. micheneri* clade bacteria have the potential to aid pollen digestion, detoxification, thrive in acidic environments, inhibit potential pathogens, and establish in wild bees. However, as our findings are limited to only genomic data, it is possible that *L. micheneri* clade bacteria are either commensal or detrimental to their hosts. If commensal, *L. micheneri* may exploit the nutrient-rich bee host environment, and bees may act as vectors for *L. micheneri* to spread to more flowers and bees. Finally, if parasitic, *L. micheneri* may exploit the nutrients in nectar and pollen at the expense of the bee hosts. Regardless, these reduced genomes offer strong clues as to how these bacteria persist and thrive in harsh but nutrient-rich niches.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Anderson KE, et al. 2013. Microbial ecology of the hive and pollination landscape: bacterial associates from floral nectar, the alimentary tract and stored food of honey bees (*Apis mellifera*). PLoS One 8(12):e83125.

Arkin AP, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 36(7):566–569.

Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9(1):75.

Buchon N, Silverman N, Cherry S. 2014. Immunity in *Drosophila melanogaster*—from microbial recognition to whole-organism physiology. Nat Rev Immunol. 14(12):796–810.

Buck LB, Eric A, Tina S, Klaenhammer TR. 2005. Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. Appl Environ Microbiol. 71(12):8344–8351.

Burkle LA, Marlin JC, Knight TM. 2013. Plant–pollinator interactions over 120 years: loss of species, co-occurrence, and function. Science 339(6127):1611–1615.

Cane JH. 2016. Adult pollen diet essential for egg maturation by a solitary Osmia bee. J Insect Physiol. 95:105–109.

Cane JH, Dobson HEM, Boyer B. 2017. Timing and size of daily pollen meals eaten by adult females of a solitary bee (*Nomia melanderi*) (Apiformes: Halictidae). Apidologie 48(1):17–30.

Chiou T-Y, et al. 2018. *Lactobacillus kosoi* sp. nov., a fructophilic species isolated from Kôso, a Japanese sugar-vegetable fermented beverage. Antonie van Leeuwenhoek 111(7):1149–1156.

Coil D, Jospin G, Darling AE. 2015. A5-MiSeq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31(4):587–589.

Cotter PD, Hill C. 2003. Surviving the acid test: responses of gram-positive bacteria to low pH. Microbiol Mol Biol Rev. 67(3):429–453, table of contents.

Csűrös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26(15):1910–1912.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.

De Man JCD, Rogosa D, Sharpe ME. 1960. A medium for the cultivation of lactobacilli. J Appl Bacteriol. 23(1):130–135.

D'Haeze W, Holsters M. 2004. Surface polysaccharides enable bacteria to evade plant immunity. Trends Microbiol. 12(12):555–561.

Donelli G, et al. 2007. Synergistic activity of dispersin B and cefamandole nafate in inhibition of staphylococcal biofilm growth on polyurethanes. Antimicrob Agents Chemother. 51(8):2733–2740.

Duar RM, et al. 2017. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. FEMS Microbiol Rev. 41(Suppl 1):S27–S48.

Ellegaard KM, et al. 2015. Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. BMC Genomics. 16:284.

Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A. 109(27):11002–11007.

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

Forsgren E, Olofsson TC, Vásquez A, Fries I. 2010. Novel lactic acid bacteria inhibiting *Paenibacillus* larvae in honey bee larvae. Apidologie 41(1):99–108.

Graham D, Patterson BD. 1982. Responses of plants to low, nonfreezing temperatures: proteins, metabolism, and acclimation. Annu Rev Plant Physiol. 33(1):347–372.

Gyorfy Z, et al. 2015. Engineered ribosomal RNA operon copy-number 100 variants of E. coli reveal the evolutionary trade-offs shaping rRNA operon number. Nucleic Acids Res. 43(3):1783–1794.

Hardigan MA, et al. 2017. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc Natl Acad Sci U S A. 114(46):E9999–E10008.

Henderson B, Nair S, Pallas J, Williams MA. 2011. Fibronectin: a multi-domain host adhesin targeted by bacterial fibronectin-binding proteins. FEMS Microbiol Rev. 35(1):147–200.

Hladun KR, Parker DR, Trumble JT. 2015. Cadmium, copper, and lead accumulation and bioconcentration in the vegetative and reproductive organs of *Raphanus sativus*: implications for plant performance and pollination. J Chem Ecol. 41(4):386–395.

Hu X, et al. 2012. Proteomic analysis of peritrophic membrane (PM) from the midgut of fifth-instar larvae, *Bombyx mori*. Mol Biol Rep. 39(4):3427–3434.

Hunger K, Beckering CL, Wiegeshoff F, Graumann PL, Marahiel MA. 2006. Cold-induced putative DEAD box RNA helicases CshA and CshB are essential for cold adaptation and interact with cold shock protein B in *Bacillus subtilis*. J Bacteriol. 188(1):240–248.

Ishigooka H, Yoshida Y, Omori T, Minoda Y. 1986. Enzymatic dioxygenation of biphenyl-2,3-diol and 3-isopropylcatechol. Agric Biol Chem. 50(4):1045–1046.

Jones JC, Myerscough MR, Graham S, Oldroyd BP. 2004. Honey bee nest thermoregulation: diversity promotes stability. Science 305(5682):402–404.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28(1):27–30.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1):D457–D462.

Kassambara A. 2017. Ggpubr: 'ggplot2' based publication ready plots. R package version 0.1.5. https://CRAN.R-project.org/package=ggpubr.

Kawasaki S, et al. 2011. *Lactobacillus ozensis* sp. nov., isolated from mountain flowers. Int J Syst Evol Microbiol. 61(10):2435–2438.

Knauf HJ, Vogel RF, Hammes WP. 1992. Cloning, sequence, and phenotypic expression of KatA, which encodes the catalase of *Lactobacillus sake* LTH677. Appl Environ Microbiol 58(3):832–839.

Kwong WK, Engel P, Koch H, Moran NA. 2014. Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc Natl Acad Sci U S A. 111(31):11509–11514.

Kwong WK, Mancenido AL, Moran NA. 2014. Genome sequences of *Lactobacillus* sp. strains WkB8 and WkB10, members of the Firm-5 clade, from honey bee guts. Genome Announc. 2(6):e01176-14.

Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. Nat Rev Microbiol. 14(6):374–384.

Kwong WK, et al. 2017. Dynamic microbiome evolution in social bees. Sci Adv. 3(3):e1600513.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods. 9(4):357–359.

Lavermicocca P, et al. 2000. Purification and characterization of novel antifungal compounds from the sourdough *Lactobacillus plantarum* strain 21B. Appl Environ Microbiol. 66(9):4084–4090.

Lee FJ, Rusch DB, Stewart FJ, Mattila HR, Newton ILG. 2015. Saccharide breakdown and fermentation by the honey bee gut microbiome. Environ Microbiol. 17(3):796–815.

Lee I-C, et al. 2016. Strain-specific features of extracellular polysaccharides and their impact on *Lactobacillus plantarum*–host interactions. Appl Environ Microbiol. 82(13):3959–3970.

Leigh JA, Coplin DL. 1992. Exopolysaccharides in plant-bacterial interactions. Annu Rev Microbiol. 46:307–346.

Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li J, et al. 2015. Two gut community enterotypes recur in diverse bumblebee species. Curr Biol. 25(15):R652–R653.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178–2189.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451–1452.

Liu Z, et al. 2005. Patterns of diversifying selection in the phytotoxin-like Scr74 gene family of *Phytophthora infestans*. Mol Biol Evol. 22(3):659–672.

Makui H, et al. 2000. Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter. Mol Microbiol. 35(5):1065–1078.

Martinson VG, et al. 2011. A simple and distinctive microbiota associated with honey bees and bumble bees: The microbiota of honey bees and bumble bees. Mol Ecol. 20(3):619–628.

McFrederick QS, Vuong HQ, Rothman JA. 2018. *Lactobacillus micheneri* sp. nov., *Lactobacillus timberlakei* sp. nov. and *Lactobacillus quenuiae*

sp. nov., lactic acid bacteria isolated from wild bees and flowers. Int J Syst Evol Microbiol. 68(6):1879–1884.

McFrederick QS, Wcislo WT, Hout MC, Mueller UG. 2014. Host species and developmental stage, but not host social structure, affects bacterial community structure in socially polymorphic bees. FEMS Microbiol Ecol. 88(2):398–406.

McFrederick QS, et al. 2013. Specificity between lactobacilli and hymenopteran hosts Is the exception rather than the rule. Appl Environ Microbiol. 79(6):1803–1812.

McFrederick QS, et al. 2017. Flowers and wild megachilid bees share microbes. Microb Ecol. 73(1):188–200.

Michener CD. 2007. The bees of the world. 2nd ed. Baltimore (MD): Johns Hopkins.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35(Web Server issue):W182–W185.

Neira Oviedo M, Van Ekeris L, Corena-Mcleod MDP, Linser PJ. 2008. A microarray-based analysis of transcriptional compartmentalization in the alimentary canal of *Anopheles gambiae* (Diptera: Culicidae) larvae. Insect Mol. Biol. 17(1):61–72.

Nicolson SW. 1998. The importance of osmosis in nectar secretion and its consumption by insects. Am Zool. 38(3):418–425.

O'Brien JA, Daudi A, Butt VS, Paul Bolwell G. 2012. Reactive oxygen species and their role in plant defence and cell wall metabolism. Planta 236(3):765–779.

Olofsson TC, Alsterfjord M, Nilson B, Butler E, Vásquez A. 2014. *Lactobacillus apinorum* sp. nov., *Lactobacillus mellifer* sp. nov., *Lactobacillus mellis* sp. nov., *Lactobacillus melliventris* sp. nov., *Lactobacillus kimbladii* sp. nov., *Lactobacillus helsingborgensis* sp. nov. and *Lactobacillus kullabergensis* sp. nov., isolated from the honey stomach of the honeybee *Apis mellifera*. Int J Syst Evol Microbiol. 64(Pt 9):3109–3119.

Ozer EA. 2018. ClustAGE: a tool for clustering and distribution analysis of bacterial accessory genomic elements. BMC Bioinformatics 19(1):150.

Ozer EA, Allen JP, Hauser AR. 2014. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools spine and AGEnt. BMC Genomics. 15:737.

Papadimitriou K, et al. 2016. Stress physiology of lactic acid bacteria. Microbiol Mol Biol Rev. 80(3):837–890.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25(7):1043–1055.

Peng Y-S, Dobson HEM. 1997. Digestion of pollen components by larvae of the flower-specialist bee *Chelostoma florisomne* (Hymenoptera: Megachilidae). J Insect Physiol. 43(1):89–100.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 31(7):1929–1936.

Potts SG, et al. 2010. Declines of managed honey bees and beekeepers in Europe. J Apic Res. 49(1):15–22.

Qixiao Z, et al. 2017. Identification of key proteins and pathways in cadmium tolerance of Lactobacillus plantarum strains by proteomic analysis. Sci Rep. 7(1):1182.

R Core Team 2018. R: A language and environment for statisticalcomputing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roulston TH, Cane JH. 2000. Pollen nutritional content and digestibility for animals. Plant Syst Evol 222 (1/4):187–209.

Rothman JA, Andrikopoulos C, Cox-Foster D, McFrederick QS. Forthcoming 2018. Floral and foliar source affect the bee nest microbial community. Microb. Ecol. https://doi.org/10.1007/s00248-018-1300-3.

Sakellaris G, Nikolaropoulos S, Evangelopoulos AE. 1989. Purification and characterization of an extracellular polygalacturonase from *Lactobacillus plantarum* strain BA 11. J Appl Bacteriol. 67(1):77–85.

Solioz M, Abicht HK, Mermod M, Mancini S. 2010. Response of gram-positive bacteria to copper stress. J Biol Inorg Chem. 15(1):3–14.

Stacy A, et al. 2014. Bacterial fight-and-flight responses enhance virulence in a polymicrobial infection. Proc Natl Acad Sci U S A. 111(21):7819–7824.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312.

Tamarit D, et al. 2015. Functionally structured genomes in Lactobacillus kunkeei colonizing the honey crop and food products of honeybees and stingless bees. Genome Biol Evol. 7(6):1455–1473.

Trchounian A, Kobayashi H. 1999. Kup is the major K+ uptake system in *Escherichia coli* upon hyper-osmotic stress at a low PH. FEBS Lett. 447(2-3):144–148.

UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45(D1):D158–D169.

Vásquez A, Olofsson TC. 2009. The lactic acid bacteria involved in the production of bee pollen and bee bread. J Apic Res. 48(3):189–195.

Vogel RF, et al. 2011. Genomic analysis reveals *Lactobacillus sanfrancis-censis* as stable element in traditional sourdoughs. Microb Cell Fact. 10(Suppl 1):S6.

Vojvodic S, Boomsma JJ, Eilenberg J, Jensen AB. 2012. Virulence of mixed fungal infections in honey bee brood. Front Zool. 9(1):5.

Vojvodic S, Rehan SM, Anderson KE. 2013. Microbial gut diversity of Africanized and European honey bee larval instars. PLoS One 8(8):e72106.

Wang X, Wood TK. 2011. Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. Appl Environ Microbiol. 77(16):5577–5583.

Wattam AR, et al. 2017. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res. 45(D1):D535–42.

Waterhouse A, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1):W296–W303.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Zhai Q, et al. 2017. Identification of key proteins and pathways in cadmium tolerance of Lactobacillus plantarum strains by proteomic analysis. Sci Rep. 7(1):1182.

Zhang C, et al. 2015. Differential natural selection of human zinc transporter genes between African and non-African populations. Sci Rep. 5:9658.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22(12):2472–2479.

Zheng H, et al. 2016. Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola*. MBio 7(6):e01326-16.

Zheng H, Powell JE, Steele MI, Dietrich C, Moran NA. 2017. Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. Proc Natl Acad Sci U S A. 114(18):4775–4780.

Associate editor: Esther Angert