Genomic features underlying evolutionary transitions of Apibacter to honeybee gut symbionts

Wenjun Zhang[1], Xue Zhang[1], Qinzhi Su[1], Min Tang[1], Hao Zheng[1], and Xin Zhou[2]

[1]China Agricultural University
[2]Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing 100193, People's Republic of China

May 15, 2020

Abstract

The symbiotic bacteria associated with honeybee gut have likely transformed from a free-living or parasitic lifestyle, through a close evolutionary association with the insect host. However, little is known about the genomic mechanism underlying bacterial transition to exclusive adaptation to the bee gut. Here we compared the genomes of bee gut symbionts Apibacter with their close relatives living in different lifestyles. We found that despite of general reduction in the Apibacter genome, genes involved in amino acid synthesis and monosaccharide detoxification were retained, which was likely beneficial to the host. Interestingly, the microaerobic Apibacter species have specifically preserved the NAR operon encoding for the nitrate respiration pathway which in contrast, is absent from the related non-free-living microaerobic pathogenic relatives. The NAR operon is also conserved in the cohabiting bee microbe Snodgrasella, but with a differed structure. This convergence implies a crucial role of respiration nitrate reduction for microaerophilic microbiomes to colonize bee gut epithelium. Genes involved in lipid, histidine and phenylacetate degradation are partially lost in Apibacter, possibly associated with the loss of pathogenicity. Antibiotic resistance genes were only sporadically distributed among Apibacter species, but condensed in their pathogenic relatives. Collectively, this study advanced our understanding of genomic transition underlying specialization in bee gut symbionts.

Introduction

Bacterial symbiotic association with insect host is ubiquitous in nature, which confers the host with traits to explore new ecological niches (Sudakaran, Kost, & Kaltenpoth, 2017). Symbiotic bacteria inhabit either intra- or extra-cellularly inside the insect, providing hosts with vital benefits, including nutrition supply, pathogen resistance, assistance in immunity development, among others (Engel & Moran, 2013; McCutcheon, Boyd, & Dale, 2019). Mutualistic symbionts may have various origins, including environmental bacteria or infective parasites (Sachs, Skophammer, & Regus, 2011; Sachs, Skophammer, Bansal, & Stajich, 2013). Despite of distinct evolutionary pathways, the convergent transition to a mutualistic symbiotic lifestyle often involves the loss of virulence factors, degeneration of superfluous functions, and preservation or obtaining of traits beneficial to the host (EWALD, 1987; McCutcheon & Moran, 2012). However, direct evidence showing taxonomic turn-over in insect symbionts are scarce. Alternatively, comparative genomic analysis of mutualistic symbionts with the free-living lineages that are phylogenetically closely related provides a feasible route to examine the evolutionary pattern and adaptive mechanism accompanying the evolutionary transition to mutualism (Boscaro et al., 2013; Zheng, Dietrich, & Brune, 2017).

The honeybees have simple but specific gut symbiotic bacteria from genera of *Gilliamella* , *Snodgrasella* , *Lactobacillus*and *Bifidobacterium* (Martinson, Moy, & Moran, 2012). They compose more than 95% of

the whole gut community, and the association with these five core bacteria can be dated back prior to the divergence of corbiculate bees (i.e., honeybee, bumblebee, stingless bee, and orchid bee) (Kwong et al., 2017). Although bee gut bacteria are not intracellular inhabitant and maternally transferred, they are inheritable among worker bees in the colony through social contacts (Powell, Martinson, Urban-Mead, & Moran, 2014). Along with the establishment of symbiotic association, these bacteria are subjected to genome reduction and evolutionary adaptation (Kwong, Engel, Koch, & Moran, 2014). This adaptive transition increases their fitness to the bee gut niche, but also limits the capacity to thrive in other environments (Ellegaard et al., 2019). As with other symbiotic microbes, honeybee gut bacteria share ancestry with bacterial species carrying varied lifestyles (Segers, Kešnerová, Kosoy, & Engel, 2017). Subsequently, genomic changes are expected to be remarkably different between the transitions from the common ancestry to gut symbionts and to other lifestyles (Tamarit et al., 2015). Thus, the honeybee-gut bacteria system provides a promising model to explore the genomic features underlying lifestyle transition of free-living bacteria to mutualistic gut symbionts.

Apibacter is a genus of bacteria that is prevalent in *Apis cerana* , *Apis dorsata* and bumblebee species (genus*Bombus*), however, they are only sporadically reported in*Apis mellifera* (Kwong et al., 2017; Kwong & Moran, 2016). Our knowledge on *Apibacter* is primarily based on genome sequences of only four strains till now (Kwong, Steele, & Moran, 2018). Phylogenetic analysis showed that *Apibacter* spp. form a monophyletic lineage, which is embedded in the *Chryseobacterium* clade (family*Flavobacteriaceae*) (Kwong & Moran, 2016). The members of the*Chryseobacterium* clade show a variety of lifestyles, including environmental free-living, opportunistic or obligate mammal and bird pathogens (McBride, 2014), of which the genus *Chryseobacterium* is the dominant member and consist of mostly environmental habitants. But strains of *Chryseobacterium gleum* and *Chryseobacterium indologenes* are common clinical isolates, which cause infections in immunocompromised patients (Calderán et al., 2011). Type strains of species from *Elizabethkingia* and *Empedobacter* are also opportunistic human pathogens that infect urinary and respiratory tract, causing sepsis, pneumonia and meningitis (Bhat, Priya, Krishnan, & Kanungo, 2016; Gupta, Zaman, Mohan, & Taneja, 2017; Lau et al., 2015; Zaman, Gupta, Kaur, Mohan, & Taneja, 2017). Isolates of genera*Bergeyella* and *Weeksella* infect mammal respiratory tracts and female genitourinary tract, respectively (Hugo, Bruun, & Jooste, 2006; Lin, Chen, & Liu, 2007; Slenker, Hess, Jungkind, & DeSimone, 2012). Members of *Riemerella* and *Ornithobacterium* are important bird pathogens, causing infections to respiratory tract and other organs (McBride, 2014). These bacteria can be transmitted horizontally by aerosol or vertically between host generations. Type species of genera *Bergeyella* , *Weeksella, Riemerella* and*Ornithobacterium* are considered non-free-living, showing a preference to micro-aerophilic conditions, which is different from most other strictly aerobic members of the *Chryseobacterium* clade (Mavromatis et al., 2011; Van Empel & Hafez, 1999). The common features of the *Chryseobacterium* clade include proteolytic activity, carbohydrate utilization and multiple antibiotic resistance, causing challenges to the treatment of infection. Protease and biofilm formation are suggested to be important for their pathogenesis (McBride, 2014). However, the genetic mechanisms underlying the pathogenicity remain uncovered.

In this study, we isolated and sequenced the genomes of 14*Apibacter* strains from the Asian honeybee *Apis cerana* . Molecular and physiology experiments were conducted to characterize its colonization positions and specificity inside the bee gut. Comparative genomic analyses of honeybee symbiotic *Apibacter* spp. with phylogenetically related free-living and pathogenic strains from the*Flavobacteriaceae* family revealed key genomic changes of*Apibacter* underlying adaptation to mutualistic symbiotic relationship with the honeybee.

Materials and Methods

Sample collection and *Apibacter*isolation

The worker bees of *Apis cerana* were collected from Sichuan, Jilin and Qinghai Provinces in China (Dataset S1). A total of six guts were dissected, homogenized and frozen in glycerol (25%, vol/vol) at –80°C. Frozen

stocks of homogenized guts were streaked out on the heart infusion agar (Oxoid) or Columbia agar (Oxoid) supplemented with 5% sheep blood (Solarbio). The plates were incubated at 35°C in 5 % CO_2 for 2–3 days. Colonies were screened by PCR using universal primers 27F and 1492R for the16S rRNA gene and Sanger sequencing.

Fluorescence *in situ* hybridization (FISH) microscopy

Ten adult worker bees of *A. cerana* were collected from a single colony in Beijing, China in May 2018. The whole guts of bees were dissected and frozen in RNAlater (Qiagen) at -80 ºC. The FISH protocol was adapted from Yuval et al. (Gottlieb et al., 2006). In brief, the guts preserved in RNAlater were fixed in the fixative solution (Ethanol: Chloroform: Acetic acid = 6:3:1) and kept overnight at the room temperature, followed by rinsing with 1 × PBS. The guts were then treated with 1 mg/mL proteinase K solution at 56 °C for 20 minutes, followed by rinsing with 1 × PBS. The guts were incubated with 6% H_2O_2 ethanol solution for 2 hours, followed by rinsing twice with 1 × PBS. Finally, the guts were permeated with 0.1% Triton for 2 hours, followed by rinsing for 2 or 3 times in 1 × PBS buffer.

Binding FISH probes targeting 16S rRNA genes were designed specifically for each of the genera *Apibacter* , *Snodgrasella* and*Gilliamella* , following Martinson et al. (Martinson et al., 2012) (Table S1). Probe hybridization was performed overnight at 37 °C, followed by rinsing in 1 × PBS for 3 times. Spectral imaging was used to visualize gut sections on a ZEISS LSM780 confocal microscope. Autofluorescence was assayed for each tissue type as the negative control. *Apibacter* were labeled with florescent together with either *S. alvi* or *G. apicola* .

Estimation of bacterial abundance using qPCR

Worker bees were collected from 2 colonies at the same apiary in Beijing, China in September 2018. Samples were stored at -80°C. Gut segments (i.e., midgut, ileum, and rectum) were dissected and separated. DNA was extracted following Powell et al. (Powell et al., 2014). Universal 16S primers 27F and 355R for bacteria were employed and*Apibacter* -specific primers Apiq9-F and Apiq9-R were designed in this study (Table S1). Standard curves were created and reactions were carried out on a LightCycler 480 (Roche Applied Science, Indianapolis, IN). Significance in differences within and between samples was determined using Mann-Whitney (Wilcoxon-rank) nonparametric U tests.

In vivo colonization experiments

Microbiota-depleted bees were obtained following Zheng et al. (Zheng, Powell, et al., 2017). Late stage pupae of both *A. cerana* or*A. mellifera* were removed from brood frames and incubated for 24-36 h in sterile plastic bins at 35 °C and 75% humidity. Newly emerged bees were kept in cup cages provided with sterilized sucrose syrup (0.5 M) and bee bread. For inoculation, bacteria strain of the*Apibacter* sp. B3706 was grown on the heart infusion agar (Oxoid), which was supplemented with 5% sheep blood from glycerol stocks, for two days. Cultivated strains were scraped and suspended in 1 × PBS to reach an OD_{600} of 1.0. Batches of 25 bees were placed in a 50 ml conical tube, and 50 µl sucrose syrup was added. The tube was rotated gently so that the syrup was coated on the surface of bees. The tube was rotated again after the addition of 50 µl of the*Apibacter* sp. B3706 suspension (˜2×10^6 cells per bee). The bacteria were inoculated into the bee guts as a result of auto- and allogrooming. Inoculated bees were reared in cup cages. Three replicate enclosures were set up for both*A. cerana* or *A. mellifera* with 20 bees in each cage. The inoculated bees were fed with sucrose syrup and sterilized pollen throughout the experiment. Colonization levels were determined at day 6 using qPCR as described previously.

Genome sequencing, assembly and annotation

Genomic DNA of *Apibacter* isolates were extracted using a bead-beating method as previously described (Powell et al., 2014). Two strains (*Apibacter* sp. B2966 and B3706) were sequenced with the PacBio RS (PacBio) at Nextomics Biosciences Co. Ltd., China, and assembled using OLC algorithm of the Celera (Chaisson & Tesler, 2012). Three strains (*Apibacter* sp. B3239, B3546 and B2912) were sequenced with a BGISEQ-500RS (BGI) at BGI-Qingdao, China. The other nine strains (*Apibacter* sp. B3813, B3883, B3887, B3889, B3912, B3913, B3918, B3924 and B3935) were sequenced with a HiSeq X-Ten (Illumina) at Novogene Co., Ltd, China. The twelve strains except for*Apibacter* sp. B2966 and B3706 were assembled with*SOAPdenovo-Trans* (version 2.04, -K 51 -m 91 –R for 150PE reads; -K 31 -m 63 –R for 100PE reads) (Xie et al., 2014), *SOAPdenovo*(only for 150PE reads) and *SPAdes* (version 3.13.0, -k 33,55,77,85) (Bankevich et al., 2012). The quality trimmed reads were mapped back to the assembled contigs using minimap2-2.9 (Li, 2018) to examine assembly quality. The bam file generated by *samtools*(version 1.8) (Li et al., 2009) and the assemblies were processed by*BamDeal*(*https://github.com/BGI-shenzhen/BamDeal*, version 0.19) to calculate and visualize sequencing coverage and GC contents of assembled contigs. Spurious contaminants, contigs with low depths, unnormal GC contents and those apparently differing from the cluster, were removed from the draft genome. Assembled genomes were then annotated using PROKKA 1.13.3 (Seemann, 2014).

Comparisons of genome structure, genome divergence, and gene contents

The contigs of each assembly were re-ordered according to the single circular genome of strain B3706 using the 'Contig Mover' tool of Mauve version 2.4.0 (Darling, Mau, Blattner, & Perna, 2004; Rissman et al., 2009). Pairwise average nucleotide identity (ANI) was calculated with JSpeciesWS (Richter, Rosselló-Móra, Oliver Glöckner, & Peplies, 2016) using the BLASTN algorithm (ANIb). Genome completeness was estimated with CheckM (Parks et al., 2015), which was available at KBase online (Arkin, Stevens, Cottingham, Maslov, & Perez, 2015) using recommended parameters. The genome structures were compared using the R-package genoPlotR (Guy, Kultima, Andersson, & Quackenbush, 2011).

Phylogenetic inference

Gene orthology was determined using OrthoMCL (Li et al., 2003) for all genomes used in this study. All steps of the OrthoMCL pipeline were executed as recommended in the manual and the mcl program was conducted using parameters '-abc -I 1.5'.

Protein sequences of the identified single-copy orthologous were aligned using Mafft-linsi (Katoh, Kuma, Miyata, & Toh, 2005). Alignment columns only containing gaps were removed and the alignments were concatenated. The phylogeny was reconstructed using RAxML v8.2.10 (Stamatakis, 2014) with the PROTGAMMAIJTT model and 100 bootstrap replicates.

Gene flux analysis

Gene gain and loss analyses and inferences for gene contents of LCAs (last common ancestors) were conducted using Count (Csurös, 2010). Standard methods used in previous works were employed in the present study (Segers et al., 2017). For each gene family, Wagner parsimony with a gene gain/loss penalty of 2 (Zaremba-Niedzwiedzka et al., 2013) was used to infer the most parsimonious ancestral states. Parameter choices followed a previous publication (Oyserman et al., 2016).

Analysis of functional gene contents

Gene contents were categorized based on COG and eggNOG (Huerta-Cepas et al., 2019) functions. Significantly differential genes and pathways were enriched by OrthoVenn (Yi, Colemanderr, Chen, & Gu, 2015). For gene families subset of interest, BLASTP(Altschul et al., 1997) was used to query against the NCBI's nr

database and TIGRFAM Hidden Markov Model (HMM) (Haft, Selengut, & White, 2003). The phylogenies of *narG*and *narH* were produced from protein sequences obtained by blastp against NCBI's nr database with default parameters. And the hits aligned using Mafft-linsi, trimmed with trimAL (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) for sites with over 50% gaps and phylogenetic trees were constructed using RAxML PROTGAMMALG with 100 bootstraps. The parameters were based on previous publication (Neuvonen et al., 2016).

Carbohydrate-active enzyme (CAZymes) gene families were identified for all analyzed genomes using the command-line version of dbCAN (Database for automated Carbohydrate-active enzyme Annotation) (Yin et al., 2012), following authors' instruction. The PULs (Polysaccharide utilization loci) were identified using the TIGRfam (Potter et al., 2018) and Pfam (Finn et al., 2014) models, following Terrapon et al. (Terrapon, Lombard, Gilbert, & Henrissat, 2015). Antibiotic resistance genes were identified by querying all the genomes against the Comprehensive Antibiotic Resistance Database (CARD) (McArthur et al., 2013).

Results

Spatial distribution of *Apibacter* in the gut of *A. cerana*

We characterized the colonization of *Apibacter* in the gut of*A. cerana* using fluorescence in situ hybridization (FISH) and qPCR. Honeybee symbionts *Snodgrassella* and *Gilliamella* are dominant in the ileum, with *Snodgrassella* colonizing the inner wall of ileum (Martinson et al., 2012). These two bacteria were used as reference coordinates to infer the location of *Apibacter* using species-specific probes (Fig. 1A-D). The spectral images showed that*Apibacter* co-resided with *Snodgrassella* in both ileum and midgut (Fig. 1A, C). The signals were stronger at the inner walls, indicating that *Apibacter* colonized the gut intima as did*Snodgrassella* . *Gilliamella* covered on the top of*Apibacter* and extended into the gut lumen (Fig. 1B, D).*Apibacter* could not be visualized clearly in the rectum, as the rectum was filled with pollen grains with auto-fluorescent under the excitation wavelength. We also quantified the absolute abundances of*Apibacter* in different gut compartments (midgut, ileum, rectum) using qPCR. It showed that the absolute abundances of *Apibacter*increased from midgut to rectum, with cell numbers ranged from 9.43×10^5 to 1.74×10^7 (Fig. 1E).

Host specialization is a common feature for many host-microbiome associations (Oh et al., 2010), which has been demonstrated by the honeybee symbiotic bacteria *Snodgrassella* (Kwong et al., 2014) and *Lactobacillus* (Ellegaard et al., 2019). To test if it is also the case for *Apibacter* , we performed the colonization specificity test. *Apibacter* sp. strain B3706 isolated from *A. cerana*was inoculated to the microbiota-free *A. cerana* and *A. mellifera* (see Methods). Six days after inoculation, the cell numbers in the gut were more than 10^7, which were much higher than in the inoculum ($< 10^6$), indicating that strain B3706 was able to colonize the guts of both *A. cerana*and *A. mellifera* . However, the colonization efficiency in*A. cerana* was significantly higher than that in *A. mellifera* (Fig. 1F), suggesting that strain B3706 was less adapted to the gut of *A. mellifera* .

Genome characters of *Apibacter* and phylogenetic inference

We isolated 14 strains of *Apibacter* from worker bees of *A. cerana* from Sichuan, Jilin, and Qinghai provinces in China (Dataset S1). Two genomes (strains B3706 and B2966) were completed into single circular chromosomes by sequencing on the PacBio platform. The assemblies of genomes sequenced with either Illumina or BGISEQ contained 12–49 contigs and with full completeness as evaluated by CheckM (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015) (Table S2).*Apibacter* strains isolated from *A. cerana* had genome sizes ranging from 2.26 to 2.35 Mb, similar to those from the bumble bees (2.33 Mbp; Brandt et al. 2016), but smaller than those from *A. dorsata* (2.63–2.76 Mbp; Kwong, Steele, and Moran 2018) (Table S2). Parasitic pathogenic species from genera *Riemerella* ,*Bergeyella* , *Weeksella* , and *Ornithobacterium* also have small genome sizes ranging from 2.16–2.44 Mb comparing to other type strains in the *Chryseobacterium* clade, suggesting that they are subjected to genome reduction as documented for other symbionts (Fig. S2 B, Table S2) (Pérez-Brocal, Latorre, & Moya, 2013). The GC contents of the honeybee symbiotic *Apibacter*

are lower than its non-symbiotic relatives, a common feature of symbionts attributed to the mutational bias and weak selections (Fig. S2 B) (McCutcheon et al., 2019; McCutcheon & Moran, 2012).

The average nucleotide identities (ANIs) between genomes obtained in this study and those from Malaysian *A. cerana* (*Apibacter*sp. wkB309, Kwong & Moran, 2016) are 95.82%–97.33%, which suggests that *Apibacter* carried by the mainland *A. cerana* have diverged from those from the Sundaland *A. cerana* , on the verge of speciation (Table S2). Six genomes among the 14 sequenced strains have almost identical ANIs (99.99%), the consensus of which was used in subsequent analyses. The genomic divergence in *Apibacter* isolates is more obvious between bee hosts, as the ANIs are only 85% and 74% to those isolated from bumble bees and *A. dorsata* , respectively. Despite of large sequence divergence, genome structures are mostly conserved across the *Apibacter* strains from *A. cerana* and*Bombus* , even though few gene rearrangements and inversions are observed in *Apibacter* genomes isolated from *A. dorsata*(Fig. S1). But the overall structural variation seems not substantial when *Apibacter* diverged with the hosts. Such a pattern is in congruent with the observations in *Bartonella apis* from*A. mellifera* , and in *Buchnera aphidicola* from aphids (Chong, Park, & Moran, 2019; Segers et al., 2017).

A maximum-likelihood phylogeny was inferred using a total of 681 orthologous single-copy genes present in all *Apibacter,* related taxa and one outgroup (*Flavobacterium aquatile* LMG 4008) chosen from the *Flavobacteriaceae* family (Fig. S2 A) (Li, Stoeckert, & Roos, 2003; Stamatakis, 2014). Consistent with the previous phylogenetic relationship (Kwong & Moran, 2016), *Apibacter* species formed a monophyletic group and the isolates from *A. cerana* clustered together (Fig. S2 A). The *Apibacter* clade was sister the lineage consisting of *Elizabethkingia* , *Riemerella* and*Chryseobacterium* genera, which is referred to as Clade C hereafter following Kwong and Moran (Kwong & Moran, 2016). These genera contain environmental non-pathogens, opportunistic pathogens, and parasitic pathogens (McBride, 2014). Finally, the *Apibacter* clade and Clade C together formed a sister relationship to Clade E (*Empedobacter* , *Weeksella* , and *Ornithobacterium*), which consists of free-living and parasitic strains associated with mammal and bird diseases (McBride, 2014).

Apibacter conferred a large number of genes loss but preserved specific host beneficial functions

To infer the genes loss patterns in *Apibacter* spp. during adaptive transition to a bee gut symbiont, we constructed the last common ancestor (LCA) for *Apibacter* spp. and close relatives using generalized parsimony and analyzed the gene flux. A total of 601 genes were lost at the node leading to the LCA of *Apibacter*branch (Fig. 2A). Of these, 498 were preserved in both Clade C and E, and 99 were only present in Clade C. Unique genes present in Clade C but lost in *Apibacter* might be associated with non-symbiotic lifestyles, e.g. environmental free-living. Among genes with known functions, those lost in *Apibacter* are enriched in the COG categories of inorganic ion transport, transcription, amino acid transport and metabolism, carbohydrate transport and metabolism, cell wall biogenesis and lipid transport and metabolism (8.7, 8.5, 8.1, 7.7, 7.7 and 6.5% of genes with COG annotation respectively, Dataset S2). The loss of cell wall biosynthesis is in line with the general pattern of genome reduction found in symbionts, which facilitates the exchange of molecules through the host-symbionts interface and allows further degeneration of metabolites in the symbionts (McCutcheon & Moran, 2012). The losses of amino acid membrane transport, carbohydrate and lipid metabolic processes could be explained as an adaptation to a nutrient-rich environment (Schmid et al., 2018).

It showed that a few symbiotic bacteria could provide major functions in pollen degradation in the bee gut, facilitating polysaccharide utilization for the host (Kešnerová et al., 2017), similar to*Bacteroidates* in human gut (Koropatkin, Cameron, & Martens, 2012). However, genes involved in carbohydrate metabolism are substantially lost in *Apibacter* . Compared with Clade C and E,*Apibacter* lost most of the carbohydrate active enzymes (CAZymes) genes and retained no more than two polysaccharide utilization loci (PUL). The remaining PULs in *Apibacter* are composed of only tandem *susCD* -like genes, and lack all surrounding CAZyme genes (Dataset S3). The protein sequences of these PUL residue structures are conserved across *Apibacter* isolates from *A. cerana* and*Bombus* (>87% similarity). To the contrast, the PULs in *Apibacter* genomes from *A. dorsata* are more diverse, with one from wkB301 homologous to those from *A. cerana* and

bumble bee (>70% similarity), one conserved among wkB301 and wkB180 from *A. dorsata* (>95% similarity), and one unique to wkB180 (isolate of bumble bee, Dataset S3). The variations in PUL residue structure among *Apibacter* strains suggests independent gene losses of CAZyme genes in the *A. dorsata*lineage.

In contrast to significant gene loss in *Apibacter* , specific genes beneficial to the host were preserved. For instance, the mannose-6-phosphate isomerase encoded by *manA* is responsible for degradation of the toxic mannose for the host (Zheng et al., 2016). This gene is lost in all strains of Clade C and E, but is preserved by all*Apibacter* strains, indicating that mannose detoxification is an important mutualistic trait in *Apibacter* (Dataset S5). It is known that symbiotic gut bacteria are capable of synthesizing amino acids for host and other co-occuring symbionts (Kwong et al., 2014; McCutcheon & Moran, 2012). Consistently, *Apibacter* have preserved all genes underlying amino acid synthesis, which are inherited from the LCA. On the contrary, the parasitic pathogens from genera*Bergeyella* , *Weeksella* , *Riemerella* and*Ornithobacterium* , have all conferred a genome reduction, while losing substantial amino acid synthesis genes (Dataset S4) (Rohmer, Hocquet, & Miller, 2011).

Respiratory nitrate reduction was enriched and conservative in*Apibacter*

A total of 1,349 gene families are shared by all three clades, representing core gene functions shared by bacteria of varied lifestyles (Fig. 2C). *Apibacter* LCA has 226 unique gene families, which include functions that are specific to the adaptation to the bee host. Clade C and Clade E have more gene families (467 gene families) in common compared to what is shared with the *Apibacter* group, in congruent with the non-symbiotic lifestyles shared between them. Gene ontology (GO) enrichment analysis was performed to explore overrepresented functions specific to the *Apibacter* group. Interestingly, the gene families belonging to the respiratory nitrate reduction (NAR) pathway are enriched in *Apibacter* (Dataset S5). Furthermore, the biosynthesis of the molybdenum cofactors (Moco), which are required for NAR nitrate reductase (Moreno-Vivián, Cabello, Martínez-Luque, Blasco, & Castillo, 1999; Stewart, 1988), are also enriched. These genes locate in the NAR operon and are harbored in all*Apibacter* isolates, with only one Moco synthesis gene (mosC) and one nitrate transporter missing in two *A. dorsata* isolates (Fig. 3). The NAR related genes are highly conserved in amino acid sequences among genomes isolated from the same bee hosts (>94% similarity). Isolates from *A. cerana* are more similar to the one from bumble bee than those from *A. dorsata* (Fig. 3). The function of the NAR pathway seems to be highly conserved, eventhough a key member gene is replaced by alternative gene. The *narI* encoding the membrane biheme b quinol-oxidizing γ subunit of the nitrate reductase is missing. Alternatively, a Rieske protein homolog was identified in the NAR operon. The Rieske protein is an iron-sulfur protein (2Fe-2S), with a function in transferring electrons from the quinone pool, which is equivalent to NarI (Schneider & Schmidt, 2005). Therefore, the Rieske protein homolog is expected to have replaced the function of NarI in transferring electrons to the NarGH complex (Arshad et al., 2015). The replacement of *narI* in the NAR operon structure was previously identified in halophilic archaea, which represents an ancient respiratory nitrate reductase (Cabello, Roldán, & Moreno-Vivián, 2004; Yoshimatsu, Iwasaki, & Fujiwara, 2002). Additionally, three copies of the *narK* nitrate transporter gene are identified in*Apibacter* (Fig. 3B) (Cole & Richardson, 2017), implying high efficiency of nitrate respiration in *Apibacter* . To infer the evolutionary origin of the nitrate reductase genes, the Maximum likelihood tree based on the *narG* and *narH* genes were constructed (Fig. S4). Interestingly, the *narG* and *narH* of*Apibacter* formed a monophyletic clade. The phylogenetic relationship of *narG* and *narH* is consistent with the phylogeny of the bacterial strains that harbor the genes. Their closely related genes are mostly from strains of genera *Flavobacterium* , implying that the nitrate reductase genes of *Apibacter* are vertically inherited.

In contrast, bacteria from Clade C and E lack intact NAR pathways, which is congruent with their aerobic nature. It is worth noting that parasitic pathogens from *Riemerella* and *Ornithobacterium*also lack the NAR operon (Mavromatis et al., 2011; Van Empel & Hafez, 1999), despite that they are microaerophilic, as with *Apibacter*(Fig. 3, Dataset S5). These results imply that the NAR pathway might be particularly beneficial to gut commensal bacteria, which prompted us to survey the NAR pathway in other honeybee gut bacteria. Interestingly, NAR pathway is also mostly conserved in *Snodgrassella* strains, but absent in the other four core

bee gut bacterial phylotypes (*Gilliamella* , *Bifidobacterium* , *Lactobacillus* Firm4 and Firm5). This difference further suggests that the NAR pathway might be generally required by microaerobic bacteria inhabiting gut epithelium (Dataset S5). However, the nitrate reductase of *Snodgrassella*showed obvious variations when compared with that of *Apibacter* : Four subunits of nitrate reductase are encoded by the *narGHJI*genes that are similar to *E. coli* (Dataset S5) (Moreno-Vivián et al., 1999), and genes involved in Moco synthesis are not located next to the NAR operon (Fig. 3B).

Apibacter lost ancestral gene families related to pathogenicity and antibiotic resistance

Some gene families shared by the common ancestor of *Apibacter* , Clade C and E are lost in the *Apibacter* group but are retained in Clade C and E. As both Clade C and E encompass important mammal and bird pathogens, the absence of these genes families among *Apibacter*group suggests that they might be superfluous or deleterious to the interactions with the host. The 467 gene families shared by Clade C and E are overrepresented in histidine metabolism, fatty acid degradation, phenylacetate catabolism and urease activity. A survey of the genes in these pathways showed that *Apibacter* spp. lost all genes related to histidine degradation, two key genes involved in long chain fatty acid beta-oxidation, and the genes responsible for the production of host toxic virulence using intermediates generated in the phenylacetate catabolism (Fig. 4) (Teufel et al., 2010). The loss of genes that are responsible for histidine degradation and long chain fatty acid beta-oxidation implies that these substrates are less accessible for*Apibacter* to use as energy sources. To the contrary, genes of these three pathways are prevalently distributed among Clade C and E (Dataset S5). All of these pathways are reported to be involved in host recognization, successful colonization, virulence factor regulation and production in pathogenic bacteria (Law et al., 2008; Moraes et al., 2014; Zarzycki-Siek et al., 2013; Zhang, Ritchie, & Rainey, 2014).

Antibiotic resistances are promiscuous for bacteria in the *Flavobacteriaceae* family, causing difficulties in the treatment of their infection (McBride, 2014). Referring to the CARD database, 13 antibiotic resistance genes which conferred resistance to beta-lactam, fluoroquinolones, tetracyclines and glycopeptides were identified in genomes of Clade C and E (Fig S3). These resistance genes are absent in the *Apibacter* group, except that a lincosamides resistance gene is identified in *A. adventories* wkB301. These results may be explained by the fact that*A. cerana* bee gut microbes are less exposed to antibiotics.

Discussion

Combining FISH and colonization experiments, we revealed the colonization specificity of *Apibacter* and its distribution in the bee gut. Comparative genomic analyses of 30 genomes from the*Flavobacteriaceae* family, including 14 newly sequenced*Apibacter* genomes from this study and publicly available genomes for the outgroups, we characterized gene signatures underlying lifestyle transition and adaptation to bee gut symbionts.

FISH visualization indicates that *Apibacter* coinhabit with*Snodgrassella* and colonize the epithelium of the bee gut. As core members of gut bacteria in *A. cerana* , both *Apibacter* and*Snodgrasella* are microaerophilic, sharing nutritional sources (Zheng, Powell, Steele, Dietrich, & Moran, 2017). We showed that*Apibacter* isolated from *A. cerana* were able to colonize*A. mellifera* , although at a significantly lower colonization rate. These results suggest that host incompatibility is probably not the constraining factor responsible for the rarity of *Apibacter*in *A. mellifera* . However, it is not yet possible to examine inter-host competition between *Apibacter* isolated from different honeybee species, because isolates from *A. mellifera* were not available to us.

Comparative genomic analysis revealed key gene functions potentially associated with the adaptation to bee gut niche. In a typical symbiotic system, benefits provision was considered crucial for the establishment of a mutualistic relationship (EWALD, 1987; Sachs et al., 2013). Our findings reveal that *Apibacter* are

indeed providing beneficial traits to the host. For example, genes involved in amino acid biosynthesis are preserved in *Apibacter* spp., at a background of overall genome reduction, which echoes those previously reported in other bee gut symbionts (Kwong et al., 2014). Furthermore, the*Apibacter* group retained the mannose catabolic gene, which was responsible for monosaccharide detoxification in the honeybee therefore broadening food choice for the host (Zheng et al., 2016).

Polysaccharides utilization is a prominent property carried by bee gut symbionts including *Gilliamella* , *Bifidobacterium* and*Lactobacillus* (Bonilla-Rosso & Engel, 2018; Engel, Martinson, & Moran, 2012; Kešnerová et al., 2017). However, relevant genes are substantially lost in the *Apibacter* group. Interestingly, the core bacterial species *Snodgrasella* that coinhabit with*Apibacter* at *A. cerana* gut epithelium also lack the capacity to utilize polysaccharides (Kwong et al., 2014). We speculate that polysaccharides might be limited in the niche that they share.

The gut lumen is mainly anaerobic, where the dominant symbiotic anaerobes inhabit. However, oxygen can diffuse from the intestinal epithelium cells and create a microaerobic environment for facultative anaerobes (He et al., 1999; Zheng, Powell, et al., 2017). A previous study found both cytochrome *bd* and cbb_3 in the *Apibacter* genome, which were presumably involved in microaerobic respiration (Kwong et al., 2018). In the present work, we identified additional anaerobic respiration NAR operon that was conserved within the *Apibacter* group and in the coinhabiting*Snodgrassella* , but absent from the other four core bee gut bacteria species. These observations suggest that the NAR pathway might be important for the microbiome to colonize intestinal epithelium. Such respiratory flexibility might enable *Apibacter* to survive altered oxygen tensions. This finding is congruent with the observation in mouse*E. coli* , where they require both microaerobic and anaerobic respirations for successful colonization (Jones et al., 2007). A further study proved that the NAR pathway played a key role in *E. coli*colonization of the mouse gut, because the NarG mutant showed colonization deficiency for both commensal bacteria and pathogenic*E. coli* (Jones et al., 2011). These results are in line with the observation that nitrate reduction could facilitate the growth of gut microaerobic bacteria at low oxygen conditions (Tiso & Schechter, 2015). Therefore, we conclude that the NAR operon is an important genetic signature for *Apibacter* adaptation to the bee gut.

Genes that are shared between the LCAs of Clade C and Clade E but absent from the *Apibacter* group, contain functions either deleterious to the mutualistic relationship with the host, or redundant for the symbiotic lifestyle. Histidine biosynthesis is one of the most energy consuming processes for bacteria, such that the degradation of histidine as carbon and nitrogen sources is strictly regulated (Bender, 2012). The histidine catabolism is limited in bee gut environments, as oxygen is required for the activation of the Hut operon (Goldberg & Hanau, 1980). Considering that the bee gut is mostly anoxic, the Hut pathway is highly likely to be malfunctioning in*Apibacter* and is susceptive to be lost. In addition, histidine degradation is important for pathogens to recognize eukaryotic hosts and to activate virulence factors (Zhang et al., 2014).

In conclusion, combining molecular and colonization experiments, for the first time, we visualized and quantified the distribution of*Apibacter* spp. inside the bee gut, and proved that*Apibacter* isolates of *A. cerana* could survive in *A. mellifera* . Genomic comparisons with relatives living on other lifestyles revealed that host beneficial traits and respiration nitrate reduction (NAR pathway) were key functions for adaptation to the bee gut environment.

Acknowledgements

The work was funded by the Program of Ministry of Science and Technology of China (2018FY100403) and National Natural Science Foundation of China (31772493).

References

Altschul, S., Madden, T. L., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. E., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein databases search programs. *Nucleic Acids Research* , *25* (17), 3389–3402. doi: 10.1093/nar/25.17.3389

Arkin, A., Cottingham, R., Henry, C., Harris, N., Stevens, R., Maslov, S., ... Yu, D.(2018). Kbase: the united states department of energy systems biology knowledgebase. *Nature Biotechnology, 36* (7), 566-569. doi: 10.1038/nbt.4163

Arshad, A., Speth, D. R., De Graaf, R. M., Op den Camp, H. J. M., Jetten, M. S. M., & Welte, C. U. (2015). A metagenomics-based metabolic model of nitrate-dependent anaerobic oxidation of methane by Methanoperedens-like archaea. *Frontiers in Microbiology* ,*6* (273), 1–14. doi: 10.3389/fmicb.2015.01423

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing.*Journal of Computational Biology* , *19* (5), 455–477. doi: 10.1089/cmb.2012.0021

Bender, R. A. (2012). Regulation of the Histidine Utilization (Hut) System in Bacteria. *Microbiology and Molecular Biology Reviews* ,*76* (3), 565–584. doi: 10.1128/mmbr.00014-12

Bhat, Ks., Priya, R., Krishnan, L., & Kanungo, R. (2016).*Elizabethkingia meningoseptica* bacteremia in a neonate: A case report and mini-review of the literature. *Journal of Current Research in Scientific Medicine* , *2* (1), 42. doi: 10.4103/2455-3069.184130

Bonilla-Rosso, G., & Engel, P. (2018). Functional roles and metabolic niches in the honey bee gut microbiota. *Current Opinion in Microbiology* , *43* , 69–76. doi: 10.1016/j.mib.2017.12.009

Boscaro, V., Felletti, M., Vannini, C., Ackerman, M. S., Chain, P. S. G., Malfatti, S., ... Petroni, G. (2013). *Polynucleobacter necessarius* , a model for genome reduction in both free-living and symbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* , *110* (46), 18590–18595. doi: 10.1073/pnas.1316687110

Cabello, P., Roldán, M. D., & Moreno-Vivián, C. (2004). Nitrate reduction and the nitrogen cycle in archaea. *Microbiology* ,*150* (11), 3527–3546. doi: 10.1099/mic.0.27303-0

Calderán, G., García, E., Rojas, P., García, E., Rosso, M., & Losada, A. (2011). *Chryseobacterium indologenes* infection in a newborn: A case report. *Journal of Medical Case Reports* , *5* , 10. doi: 10.1186/1752-1947-5-10

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* , *25* (15), 1972–1973. doi: 10.1093/bioinformatics/btp348

Chaisson, M. J., & Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinformatics* ,*238* (13), 1471-2105. doi: 10.1186/1471-2105-13-238

Chong, R. A., Park, H., & Moran, N. A. (2019). Genome Evolution of the Obligate Endosymbiont *Buchnera aphidicola* . *Molecular Biology and Evolution* , *36* (7), 1481–1489 doi: 10.1093/molbev/msz082

Cole, J. A., & Richardson, D. J. (2017). Respiration of Nitrate and Nitrite. *Ecosal Plus* , *3* (1). doi: 10.1128/ecosalplus.3.2.5

Csurös, M. (2010). Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* , *26* (15), 1910–1912. doi: 10.1093/bioinformatics/btq315

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* , *14* (7), 1394–1403. doi: 10.1101/gr.2289704

Ellegaard, K. M., Brochet, S., Bonilla-Rosso, G., Emery, O., Glover, N., Hadadi, N., ... Engel, P. (2019). Genomic changes underlying host specialization in the bee gut symbiont *Lactobacillus* Firm5.*Molecular Ecology* , *41* (0), 1–61. doi: 10.1111/mec.15075

Engel, P., Martinson, V. G., & Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee.*Proceedings of the National Academy of Sciences of the United States of America* , *109* (27), 11002–11007. doi: 10.1073/pnas.1202970109

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects - diversity in structure and function. *FEMS Microbiology Reviews* ,*37* (5), 699–735. doi: 10.1111/1574-6976.12025

EWALD, P. W. (1987). Transmission Modes and Evolution of the Parasitism-Mutualism Continuum. *Annals of the New York Academy of Sciences* , *503* (1), 295–306. doi: 10.1111/j.1749-6632.1987.tb40616.x

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., ... Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Research* , *42* (1), 222–230. doi: 10.1093/nar/gkt1223

Goldberg, R. B., & Hanau, R. (1980). Regulation of *Klebsiella pneumoniae* hut operons by oxygen. *Journal of Bacteriology* ,*141* (2), 745–750. doi: 10.1128/jb.141.2.745-750.1980

Gottlieb, Y., Ghanim, M., Chiel, E., Gerling, D., Portnoy, V., Steinberg, S., ... Zchori-Fein, E. (2006). Identification and localization of a *Rickettsia* sp. in Bemisia tabaci (Homoptera: Aleyrodidae). *Applied and Environmental Microbiology* ,*72* (5), 3646–3652. doi: 10.1128/AEM.72.5.3646-3652.2006

Gupta, P., Zaman, K., Mohan, B., & Taneja, N. (2017).*Elizabethkingia miricola* : A rare non-fermenter causing urinary tract infection. *World Journal of Clinical Cases* , *5* (5), 187-190. doi: 10.12998/wjcc.v5.i5.187

Guy, L., Kultima, J. R., Andersson, S. G. E., & Quackenbush, J. (2011). GenoPlotR: comparative gene and genome visualization in R.*Bioinformatics* , *27* (13), 2334–2335. doi: 10.1093/bioinformatics/btq413

Haft, D. H., Selengut, J. D., & White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Research* , *31* (1), 371–373. doi: 10.1093/nar/gkg128

He, G., Shankar, R. A., Chzhan, M., Samouilov, A., Kuppusamy, P., & Zweier, J. L. (1999). Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proceedings of the National Academy of Sciences of the United States of America* , *96* (8), 4586–4591. doi: 10.1073/pnas.96.8.4586

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., ... Bork, P. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* , *47* (1), 309–314. doi: 10.1093/nar/gky1085

Hugo, C. J., Bruun, B., & Jooste, P. J. (2006). The Genera*Bergeyella* and *Weeksella* . In: Dworkin M., Falkow S., Rosenberg E., Schleifer KH., Stackebrandt E (Eds.), *The Prokaryotes* (pp. 532–538). New York, NY: Springer.

Jones, S. A., Chowdhury, F. Z., Fabich, A. J., Anderson, A., Schreiner, D. M., House, A. L., ... Conway, T. (2007). Respiration of*Escherichia coli* in the mouse intestine. *Infection and Immunity* , *75* (10), 4891–4899. doi: 10.1128/IAI.00484-07

Jones, S. A., Gibson, T., Maltby, R. C., Chowdhury, F. Z., Stewart, V., Cohen, P. S., & Conway, T. (2011). Anaerobic Respiration of*Escherichia coli* in the Mouse Intestine. *Infection and Immunity* , *79* (10), 4218–4226. doi: 10.1128/iai.05395-11

Katoh, K., Kuma, K. ichi, Miyata, T., & Toh, H. (2005). Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Informatics. International Conference on Genome Informatics* ,*16* (1), 22–33. doi: 10.11234/gi1990.16.22

Kešnerová, L., Mars, R. A. T., Ellegaard, K. M., Troilo, M., Sauer, U., & Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biology* , *15* (12), e2003467. doi: 10.1371/journal.pbio.2003467

Koropatkin, N. M., Cameron, E. A., & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology* , *10* (5), 323–335. doi: 10.1038/nrmicro2746

Kwong, W. K., Engel, P., Koch, H., & Moran, N. A. (2014). Genomics and host specialization of honey bee and bumble bee gut symbionts.*Proceedings of the National Academy of Sciences* , *111* (31), 11509–11514. doi: 10.1073/pnas.1405838111

Kwong, W. K., Medina, L. A., Koch, H., Sing, K. W., Soh, E. J. Y., Ascher, J. S., ... Moran, N. A. (2017). Dynamic microbiome evolution in social bees. *Science Advances* , *3* (3), 1–16. doi: 10.1126/sciadv.1600513

Kwong, W. K., & Moran, N. A. (2016). *Apibacter adventoris* gen. nov., sp. nov., a member of the phylum Bacteroidetes isolated from honey bees. *International Journal of Systematic and Evolutionary Microbiology* , *66* (3), 1323–1329. doi: 10.1099/ijsem.0.000882

Kwong, W. K., Steele, M. I., & Moran, N. A. (2018). Genome Sequences of*Apibacter* spp., Gut Symbionts of Asian Honey Bees. *Genome Biology and Evolution* , *10* (4), 1174–1179. doi: 10.1093/gbe/evy076

Lau, S. K. P., Wu, A. K. L., Teng, J. L. L., Tse, H., Curreem, S. O. T., Tsui, S. K. W., ... Woo, P. C. Y. (2015). Evidence for*Elizabethkingia Anophelis* transmission from mother to infant, Hong Kong. *Emerging Infectious Diseases* , *21* (1), 232–241. doi: 10.3201/eid2102.140623

Law, R. J., Hamlin, J. N. R., Sivro, A., McCorrister, S. J., Cardama, G. A., & Cardona, S. T. (2008). A functional phenylacetic acid catabolic pathway is required for full pathogenicity of *Burkholderia cenocepacia* in the *Caenorhabditis elegans* host model.*Journal of Bacteriology* , *190* (21), 7209–7218. doi: 10.1128/JB.00481-08

Li, H. (2018). Minimap2: fast pairwise alignment for long DNA sequences.*Bioinformatics* , *34* (18), 3094–3100. doi: 0.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25* (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research* , *13* (9), 2178–2189. doi: 10.1101/gr.1224503

Lin, W. R., Chen, Y. S., & Liu, Y. C. (2007). Cellulitis and bacteremia caused by *Bergeyella zoohelcum* . *Journal of the Formosan Medical Association* , *106* (7), 573–576. doi: 10.1016/S0929-6646(07)60008-4

Martinson, V. G., Moy, J., & Moran, N. A. (2012). Establishment of characteristic gut bacteria during development of the honeybee worker.*Applied and Environmental Microbiology* , *78* (8), 2830–2840. doi: 10.1128/AEM.07810-11

Mavromatis, K., Lu, M., Misra, M., Lapidus, A., Nolan, M., Lucas, S., ... Kyrpides, N. C. (2011). Complete genome sequence of*Riemerella anatipestifer* type strain (ATCC 11845).*Standards in genomic sciences* , *4* (2), 144–153. doi:10.4056/sigs.1553862

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., ... Wright, G. D. (2013). The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy* , *57* (7), 3348–3357. doi: 10.1128/AAC.00419-13

McBride, M. J. (2014). The Family *Flavobacteriaceae* . In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F (Eds.),*The Prokaryotes* (pp. 643–676). Berlin, Heidelberg, Germany: Springer.

McCutcheon, J. P., Boyd, B. M., & Dale, C. (2019). The Life of an Insect Endosymbiont from the Cradle to the Grave. *Current Biology* , *29* (11), 485–495. doi: 10.1016/j.cub.2019.03.032

McCutcheon, J. P., & Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* , *10* (1), 13–26. doi: 10.1038/nrmicro2670

Moraes, P. M. R. O., Seyffert, N., Silva, W. M., Castro, T. L. P., Silva, R. F., Lima, D. D., ... Azevedo, V. (2014). Characterization of the Opp Peptide Transporter of *Corynebacterium pseudotuberculosis* and Its Role in Virulence and Pathogenicity.*BioMed Research International* , 1–7. doi: 10.1155/2014/489782

Moreno-Vivián, C., Cabello, P., Martínez-Luque, M., Blasco, R., & Castillo, F. (1999). Prokaryotic nitrate reduction: Molecular properties and functional distinction among bacterial nitrate reductases.*Journal of Bacteriology* , *181* (21), 6573–6584. doi: 10.1128/jb.181.21.6573-6584.1999

Neuvonen, M. M., Tamarit, D., Näslund, K., Liebig, J., Feldhaar, H., Moran, N. A., ... Andersson, S. G. E. (2016). The genome of*Rhizobiales* bacteria in predatory ants reveals urease gene functions but no genes for nitrogen fixation. *Scientific Reports* ,*6* , 39197. doi: 10.1038/srep39197

Oh, P. L., Benson, A. K., Peterson, D. A., Patil, P. B., Moriyama, E. N., Roos, S., & Walter, J. (2010). Diversification of the gut symbiont*Lactobacillus reuteri* as a result of host-driven evolution.*ISME Journal* , *4* (3), 377–387. doi: 10.1038/ismej.2009.123

Oyserman, B. O., Moya, F., Lawson, C. E., Garcia, A. L., Vogt, M., Heffernen, M., ... McMahon, K. D. (2016). Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating bacteria. *ISME Journal* ,*10* (12), 2931–2945. doi: 10.1038/ismej.2016.67

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* , *25* (7), 1043–1055. doi: 10.1101/gr.186072.114

Pérez-Brocal, V., Latorre, A., & Moya, A. (2013). Symbionts and Pathogens: What is the Difference? *Current Topics in Microbiology and Immunology* , *358* , 215–243. doi: 10.1007/82

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research* , *46* (1), 200–204. doi: 10.1093/nar/gky448

Powell, J. E., Martinson, V. G., Urban-Mead, K., & Moran, N. A. (2014). Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera* . *Applied and Environmental Microbiology* ,*80* (23), 7378–7387. doi: 10.1128/AEM.01861-14

Praet, J., Aerts, M., de Brandt, E., Meeus, I., Smagghe, G., & Vandamme, P. (2016). *Apibacter mensalis* sp. Nov.: A rare member of the bumblebee gut microbiota. *International Journal of Systematic and Evolutionary Microbiology* , *66* (4), 1645–1651. doi: 10.1099/ijsem.0.000921

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., & Peplies, J. (2016). JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* ,*32* (6), 929–931. doi: 10.1093/bioinformatics/btv681

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., & Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics (Oxford, England)* , *25* (16), 2071–2073. doi: 10.1093/bioinformatics/btp356

Rohmer, L., Hocquet, D., & Miller, S. I. (2011). Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis.*Trends in Microbiology* , *19* (7), 341–348 doi: 10.1016/j.tim.2011.04.003

Sachs, J. L., Skophammer, R. G., & Regus, J. U. (2011). Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences, 108* (Supplement_2), 10800–10807. doi: 10.1073/pnas.1100304108

Sachs, J. L., Skophammer, R. G., Bansal, N., & Stajich, J. E. (2013). Evolutionary origins and diversification of proteobacterial mutualists.*Proceedings of the Royal Society B: Biological Sciences* ,*281* (20132146). doi: 10.1098/rspb.2013.2146

Schmid, M., Muri, J., Melidis, D., Varadarajan, A. R., Somerville, V., Wicki, A., ... Ahrens, C. H. (2018). Comparative genomics of completely sequenced *Lactobacillus helveticus* genomes provides insights into strain-specific genes and resolves metagenomics data down to the strain level. *Frontiers in Microbiology* , *9* , 63. doi: 10.3389/fmicb.2018.00063

Schneider, D., & Schmidt, C. L. (2005). Multiple Rieske proteins in prokaryotes: Where and why? *Biochimica et Biophysica Acta - Bioenergetics* , *1710* (1), 1–12. doi: 10.1016/j.bbabio.2005.09.003

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation.*Bioinformatics* , *30* (14), 2068–2069. doi: 10.1093/bioinformatics/btu153

Segers, F. H., Kešnerová, L., Kosoy, M., & Engel, P. (2017). Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *The ISME Journal* ,*11* (5), 1232–1244. doi: 10.1038/ismej.2016.201

Slenker, A. K., Hess, B. D., Jungkind, D. L., & DeSimone, J. A. (2012). Fatal case of *Weeksella virosa* sepsis. *Journal of Clinical Microbiology* , *50* (12), 4166–4167. doi: 10.1128/JCM.01761-12

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* ,*30* (9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Stewart, V. (1988). Nitrate respiration in relation to facultative metabolism in enterobacteria. *Microbiological Reviews* ,*52* (2), 190–232. doi: 10.1128/mmbr.52.2.190-232.1988

Sudakaran, S., Kost, C., & Kaltenpoth, M. (2017). Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends in Microbiology* , *25* (5), 375–390. doi: 10.1016/j.tim.2017.02.014

Tamarit, D., Ellegaard, K. M., Wikander, J., Olofsson, T., Vásquez, A., & Andersson, S. G. E. (2015). Functionally structured genomes in*Lactobacillus kunkeei* colonizing the honey crop and food products of honeybees and stingless bees. *Genome Biology and Evolution* ,*7* (6), 1455–1473. doi: 10.1093/gbe/evv079

Terrapon, N., Lombard, V., Gilbert, H. J., & Henrissat, B. (2015). Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* , *31* (5), 647–655. doi: 10.1093/bioinformatics/btu716

Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., ... Fuchs, G. (2010). Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proceedings of the National Academy of Sciences* , *107* (32), 14390–14395. doi: 10.1073/pnas.1005399107

Tiso, M., & Schechter, A. N. (2015). Nitrate reduction to nitrite, nitric oxide and ammonia by gut bacteria under physiological conditions.*PLoS ONE* , *10* (3), e0119712. doi: 10.1371/journal.pone.0119712

Van Empel, P. C. M., & Hafez, H. M. (1999). *Ornithobacterium rhinotracheale* : A review. *Avian Pathology* , *28* (3), 217–227. doi: 10.1080/03079459994704

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., ... Wang, J. (2014). SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* , *30* (12), 1660–1666. doi: 10.1093/bioinformatics/btu077

Yi, W., Colemanderr, D., Chen, G., & Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* , *43* (1), 78–84. doi: 10.1093/nar/gkv487

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation.*Nucleic Acids Research* , *40* (1), 445-451. doi: 10.1093/nar/gks479

Yoshimatsu, K., Iwasaki, T., & Fujiwara, T. (2002). Sequence and electron paramagnetic resonance analyses of nitrate reductase NarGH from a denitrifying halophilic euryarchaeote *Haloarcula marismortui* .*FEBS Letters* , *516* (1–3), 145–150. doi: 10.1016/S0014-5793(02)02524-3

Zaman, K., Gupta, P., Kaur, V., Mohan, B., & Taneja, N. (2017).*Empedobacter Falsenii* : a Rare Non-Fermenter Causing Urinary Tract Infection in a Child with bladder cancer. *SOA: Clinical Medical Cases, Reports & Reviews* , *1* (1): 2-4.

Zaremba-Niedzwiedzka, K., Viklund, J., Zhao, W., Ast, J., Sczyrba, A., Woyke, T., ... Andersson, S. G. E. (2013). Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. *Genome Biology* , *14* (11), 130. doi: 10.1186/gb-2013-14-11-r130

Zarzycki-Siek, J., Norris, M. H., Kang, Y., Sun, Z., Bluhm, A. P., McMillan, I. A., & Hoang, T. T. (2013). Elucidating the*Pseudomonas aeruginosa* Fatty Acid Degradation Pathway: Identification of Additional Fatty Acyl-CoA Synthetase Homologues.*PLoS ONE* , *8* (5), e64554. doi: 10.1371/journal.pone.0064554

Zhang, X. X., Ritchie, S. R., & Rainey, P. B. (2014). Urocanate as a potential signaling molecule for bacterial recognition of eukaryotic hosts. *Cellular and Molecular Life Sciences* , *71* (4), 541–547. doi: 10.1007/s00018-013-1527-6

Zheng, H., Dietrich, C., & Brune, A. (2017). Genome analysis of*Endomicrobium proavitum* suggests loss and gain of relevant functions during the evolution of intracellular symbionts. *Applied and Environmental Microbiology* , *83* (17), 1–14. doi: 10.1128/AEM.00656-17

Zheng, H., Nishida, A., Kwong, W. K., Koch, H., Engel, P., Steele, M. I., & Moran, N. A. (2016). Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola* . *MBio* , *7* (6), e01326-16. doi: 10.1128/mBio.01326-16

Zheng, H., Powell, J. E., Steele, M. I., Dietrich, C., & Moran, N. A. (2017). Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proceedings of the National Academy of Sciences of the United States of America* , *114* (18), 4775–4780. doi: 10.1073/pnas.1701819114

Data Accessibility

The *Apibacter* genome sequences in this study are available under NCBI Bioproject accession PRJNA578212.

Supplementary Datasets:

Dataset S1. List of genome sequence deposition and strain collection sites.

Dataset S2. List of *Apibacter* specific gene families gain and loss and COG category abbreviations.

Dataset S3. Distribution of CAZy genes according to three groups and PULs similarities among *Apibacter* group.

Dataset S4. Distribution of amino acid biosynthesis genes distributions.

Dataset S5. List of sublineage specific gene families.

Author Contributions

X. Zhang conceived the idea and wrote the manuscript with support from W. Zhang. W. Zhang performed the experiment and bioinformatic analysis. X. Zhang and W. Zhang analyzed data. Q. Su and W. Zhang isolated the*Apibacter* strains used in this work. M. Tang conducted the assembly of the *Apibacter* genomes. X. Zhou supervised the findings of this work.

Figures

Hosted file

`image1.emf` available at https://authorea.com/users/322259/articles/451312-genomic-features-underlying-evolutionary-transitions-of-apibacter-to-honeybee-gut-symbionts

Figure 1 Characterization of the*Apibacter* colonization in honeybees gut. A - D. Localization of*Apibacter* spp. coordinating with *Snodgrassella* and*Gilliamella* within midgut and ileum of *A. cerana* worker bees. E. The colonization abundance of *Apibacter* spp. at different intestinal organs in the gut of *A. cerana* worker bees. (n=15 per treatment) F. Colonization of newly emerged germ free bees of*A. cerana* and *A. mellifera* with *Apibacter* isolated B3706 from *A. cerana* (n=9). The results of Mann–Whitney U tests (*P < 0.05) are shown.

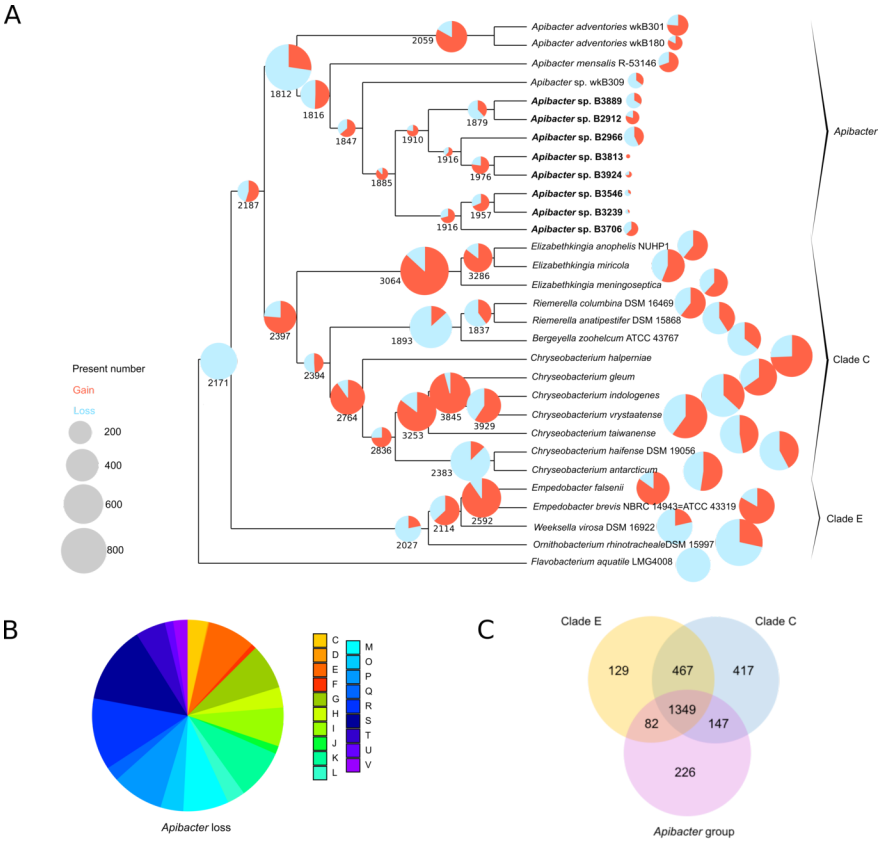

Figure 2 Genes flux analysis and functional classification of niche specifying genes. A. Loss and gain of gene families (blue, gain; red, loss) in gene content since the last common ancestor (numbers in black). The size of the pie chart reflects the amplitude of total gene flux (gain + loss). B. COG functional classification of genes lost in the *Apibacter* genome. See Dataset S2 for complete list of gene families with annotations and COG category abbreviations. C. Venn diagram shows gene family distribution among the three major groups:*Apibacter* group, Clade C and Clade E.

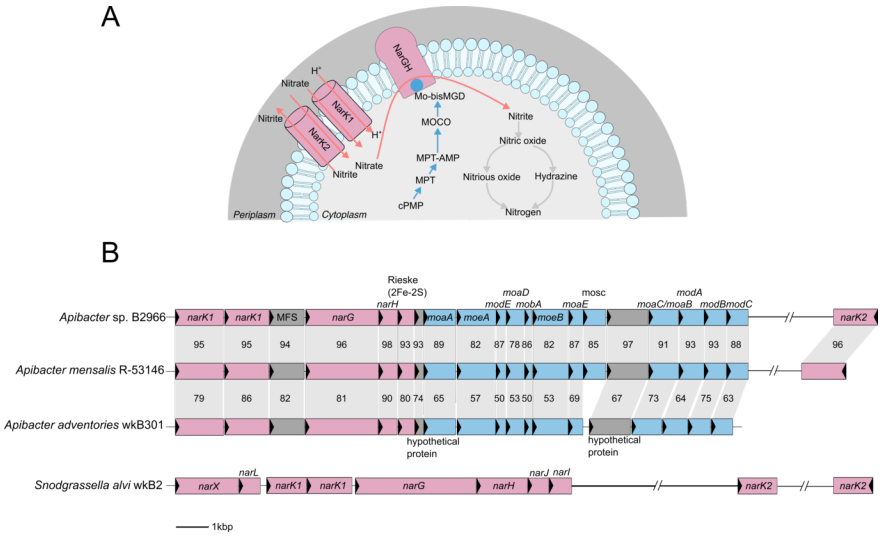

Figure 3 Respiratory nitrate reduction pathway is specifically conserved in the *Apibacter* group, which is also possessed by*Snodgrassella* . A. Schematic diagram shows that the respiration nitrate reduction pathway together with molybdate cofactor synthesis are conserved accoss the *Apibacter* group (pink arrows), and nitrite detoxification (grey arrows) is absent in the *Apibacter* group but possessed by genomes in Clade C/E (genes distribution are provided in Dataset S5). B. Genomic regions of type strains of *Apibacter* from*A. cerana* , bumble bee and *A. dorsata* encodes genes involved in respiratory nitrate reduction. Vertical grey blocks connect homologous genes among type strains, with numbers representing the percentage of sequence similarities. Genes in pink encodes nitrate reductase and transporter. Genes in blue encodes molybdate cofactor synthesis. Genes in grey are hypothetical genes or genes that are not directly related.

Hosted file

`image4.emf` available at https://authorea.com/users/322259/articles/451312-genomic-features-underlying-evolutionary-transitions-of-apibacter-to-honeybee-gut-symbionts

Figure 4 Metabolic pathways that are present in genomes of the Clade C and Clade E, but incomplete among genomes of *Apibacter*group. A. Histidine degradation pathway; B. Fatty acid beta-oxidation pathway; C. Phenylacetate oxidation pathway. Black arrow, genes mostly present in genomes from three groups; Red arrow, genes are absent among all genomes in *Apibacter* group, but possessed by genomes in Clade C and E (genes distributions are provided in Dataset S5). Substrates in red in phenylacetate degradation are documented to be toxic to host.