



# Beyond 16S rRNA Community Profiling: Intra-Species Diversity in the Gut Microbiota

Kirsten M. Ellegaard\* and Philipp Engel\*

Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

## OPEN ACCESS

### Edited by:

Robert Brucker,  
Rowland Institute at Harvard, USA

### Reviewed by:

Benoit Chassaing,  
Georgia State University, USA  
Adam C. N. Wong,  
Boston Children's Hospital and  
Harvard Medical School, USA  
Mathieu Groussin,  
Massachusetts Institute  
of Technology, USA

### \*Correspondence:

Philipp Engel  
philipp.engel@unil.ch  
Kirsten M. Ellegaard  
kirsten.ellegaard@unil.ch

### Specialty section:

This article was submitted to  
Microbial Symbioses,  
a section of the journal  
Frontiers in Microbiology

**Received:** 29 June 2016

**Accepted:** 05 September 2016

**Published:** 21 September 2016

### Citation:

Ellegaard KM and Engel P (2016)  
Beyond 16S rRNA Community  
Profiling: Intra-Species Diversity  
in the Gut Microbiota.  
Front. Microbiol. 7:1475.  
doi: 10.3389/fmicb.2016.01475

Interactions with microbes affect many aspects of animal biology, including immune system development, nutrition and health. In vertebrates, the gut microbiota is dominated by a small subset of phyla, but the species composition within these phyla is typically not conserved. Moreover, several recent studies have shown that bacterial species in the gut are composed of a multitude of strains, which frequently co-exist in their host, and may be host-specific. However, since the study of intra-species diversity is challenging, particularly in the setting of complex, host-associated microbial communities, our current understanding of the distribution, evolution and functional relevance of intra-species diversity in the gut is scarce. In order to unravel how genomic diversity translates into phenotypic diversity, community analyses going beyond 16S rRNA profiling, in combination with experimental approaches, are needed. Recently, the honeybee has emerged as a promising model for studying gut bacterial communities, particularly in terms of strain-level diversity. Unlike most other invertebrates, the honeybee gut is colonized by a remarkably consistent and specific core microbiota, which is dominated by only eight bacterial species. As for the vertebrate gut microbiota, these species are composed of highly diverse strains suggesting that similar evolutionary forces shape gut community structures in vertebrates and social insects. In this review, we outline current knowledge on the evolution and functional relevance of strain diversity within the gut microbiota, including recent insights gained from mammals and other animals such as the honeybee. We discuss methodological approaches and propose possible future avenues for studying strain diversity in complex bacterial communities.

**Keywords:** strain diversity, sub-species diversity, community analysis, gut microbiota evolution, metagenomics, honeybee

## INTRODUCTION

All animals have evolved in a bacterial world, and interactions with microbes are known to affect many aspects of animal biology (Sekirov et al., 2010; McFall-Ngai et al., 2013). One of the major arenas for such interactions is the animal gut, which is typically colonized by a large number of diverse microbes (Ley et al., 2008b). Several important functions have by now been attributed to the gut microbiota, including host immune system development, nutritional supplementation, and pathogen colonization resistance (Round and Mazmanian, 2009; Sekirov et al., 2010; Lee and Hase, 2014).

The mammalian gut typically hosts hundreds of bacterial species, where the taxonomic composition varies both within and between host species (Ley et al., 2008a; Muegge et al., 2011; Delsuc et al., 2014; Sanders et al., 2015; Bik et al., 2016). Based on network analysis of gut taxonomic profiles in combination with host-specific traits, several parameters have been shown to play a role in shaping the composition of the mammalian gut microbiota, including diet, gut morphology and host habitat (Ley et al., 2008a; Muegge et al., 2011; Delsuc et al., 2014; Sanders et al., 2015; Bik et al., 2016). For example, herbivores and carnivores can be clearly distinguished by network analysis, consistent with the notion that the gut microbiota provides complementary diet-dependent metabolic functions to their hosts, e.g., the ability of herbivores to digest cellulose (Ley et al., 2008b; Douglas, 2014). However, examples of co-diversification of specific bacterial lineages with their mammalian hosts have also been found at shorter evolutionary time-scales (Falush et al., 2003; Moodley and Linz, 2009; Moeller et al., 2016), suggesting that host-adaptation also plays an important role in the evolution of the gut microbiota.

Despite variability in species composition, the mammalian gut microbiota is dominated by relatively few bacterial phyla, including Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, and Verrucomicrobia (Ley et al., 2008a). Interestingly, thanks to an increasing number of 16S rRNA profiling studies, it is becoming clear that the same phyla recur in other metazoa, albeit with different relative abundances. For example, the Proteobacteria are abundant in both birds and fish (Sullam et al., 2012; Waite and Taylor, 2014), whereas they tend to have a lower abundance in mammals. Likewise, based on currently available data, both Proteobacteria and Firmicutes are common colonizers of the insect gut (Colman et al., 2012; Engel and Moran, 2013; Douglas, 2015). In contrast, free-living bacterial communities from other habitats, such as soil and water, have been shown to have a much broader representation of phyla (Ley et al., 2008b). Therefore, it appears that adaptation to life in the gut has occurred predominantly in a small number of bacterial phyla, which have diversified within the gut environment (Ley et al., 2006, 2008b).

Overall, 16S rRNA profiling studies have provided invaluable insights into the taxonomic composition of the gut microbiota, which in turn has facilitated the inference of broad evolutionary patterns. However, species diversity may only present the tip of the iceberg of microbial complexity in the gut. Recent studies applying methods that go beyond the typical 16S rRNA profiling approach have provided compelling evidence that the bacterial species of the gut microbiota are composed of a multitude of strains, which are likely to influence gut microbiota function (Faith et al., 2013; Schloissnig et al., 2013; Greenblum et al., 2015; Zhu et al., 2015) (**Figure 1**). Analysis of intra-species diversity is therefore needed to shed light on how horizontal gene transfer, competition and selection shape the evolution of gut-colonizing bacteria. In this review, we will outline current knowledge on intra-species diversity within the gut microbiota, discuss methodological approaches and propose possible future research avenues in this exciting new field.

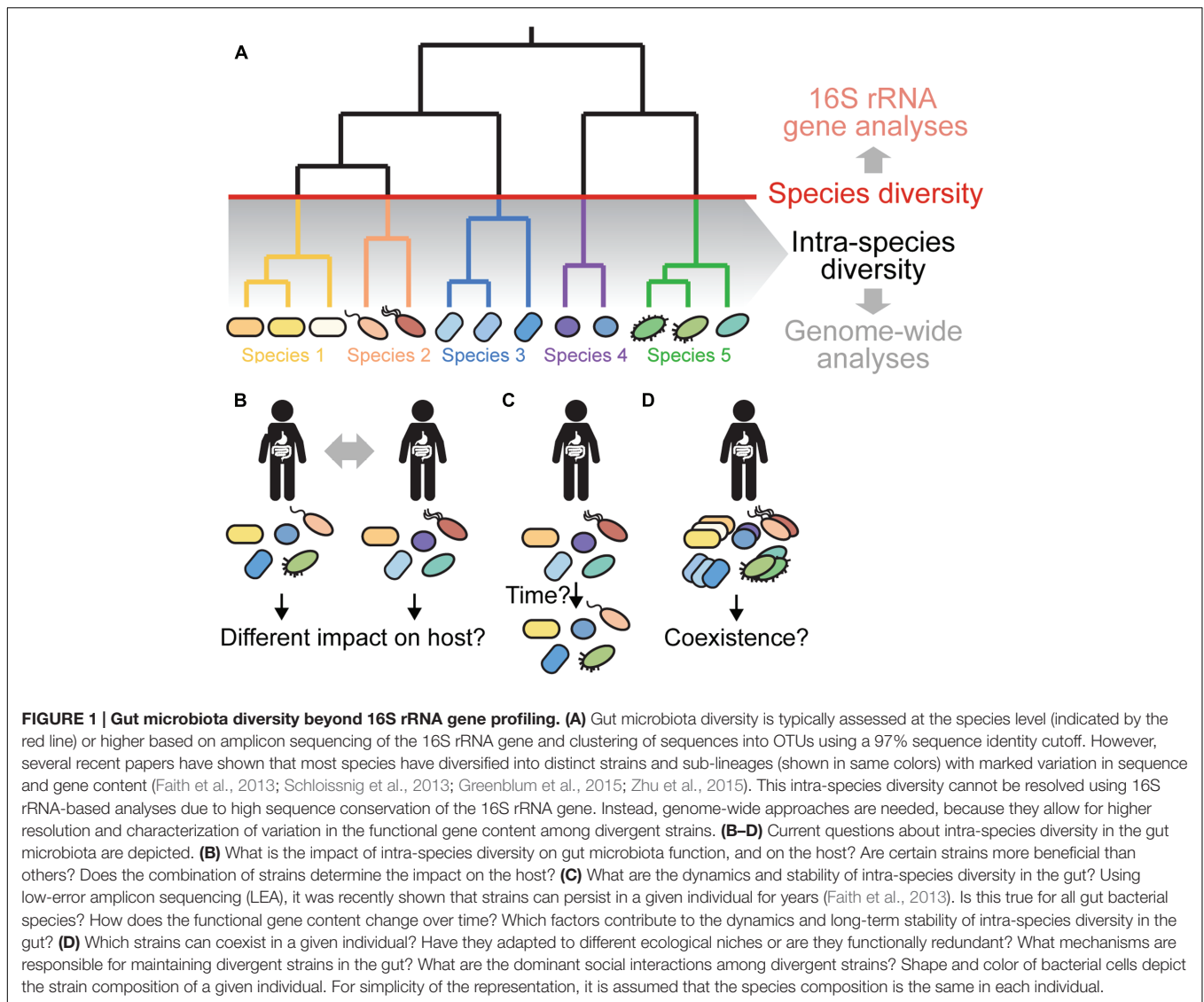
## WHAT IS INTRA-SPECIES DIVERSITY?

The bacterial species concept has been debated for decades (Wayne, 1988; Stackebrandt and Goebel, 1994; Sapp and Fox, 2013), and continues to be a contentious issue (Cohan, 2002; Caro-Quintero and Konstantinidis, 2012; Doolittle, 2012). Currently, there is no agreement as to what a bacterial species is, and it has even been proposed that the very concept is meaningless (Booth et al., 2016).

As in all organisms, bacterial evolution is shaped by genetic variation arising from mutations (Eyre-Walker and Keightley, 2007; Koskella and Vos, 2015). However, unlike multi-cellular eukaryotes, bacterial evolution is also strongly influenced by horizontal gene transfer (Ochman et al., 2000; Treangen and Rocha, 2011; Koskella and Vos, 2015). Although horizontal gene transfers happen more frequently between closely related bacteria, they also occur among distantly related species, and as such represent a major challenge for delineating bacterial species (Majewski, 2001; Fraser et al., 2007; Smillie et al., 2011). Indeed, bacterial strain diversity typically encompasses substantial gene content variation. To describe this bipartite nature of bacterial genomes, Tettelin et al. (2005) introduced the term “pan-genome” for the complete collection of genes that can be encountered in a bacterial species, while the genes shared by all members of a species was termed the “core genome.” In this early paper the comparative genome analysis revealed a core genome corresponding to approximately 80% of the total genome for each of the six sequenced strains of *Streptococcus agalactiae*. However, much higher levels of gene content variability have been found in later studies (reviewed by Land et al., 2015), perhaps in part as a consequence of the inflexible current species definitions.

Variability in gene content can be caused by either gene loss or horizontal gene transfer, where a large fraction of the variably present genes in a species can typically be attributed to phages and genes of unknown function (Ochman et al., 2000). Nevertheless, horizontally acquired (or lost) genes can also contribute to ecological adaptation, and they are likely to be major drivers of niche differentiation (and eventually speciation) among closely related bacteria (Smillie et al., 2011; Shapiro et al., 2012; Cordero and Polz, 2014; Biller et al., 2015; Bendall et al., 2016).

Variability in gene content among bacteria from the same species group was initially inferred from comparative genome analyses of cultured strains, typically isolated from samples collected at different time-points or from different locations (Rocap et al., 2003; Tettelin et al., 2005). However, during the last decade, studies of diversity directly from environmental samples have started to emerge, thereby providing novel insights into the occurrence of *co-existing* intra-species diversity in natural populations. Several studies have shown that natural communities harbor populations of bacteria, which form discrete clusters based on sequence analysis (Acinas et al., 2004; Venter et al., 2004; Fraser et al., 2007; Caro-Quintero and Konstantinidis, 2012). These studies are therefore consistent with the idea that discrete bacterial populations (akin to species) exist, but cannot be simply defined according to a universal sequence divergence



cut-off. However, since a single environmental sample only provides a “snap-shot” of a population in time, the stability of the observed clusters cannot be known. Indeed, clusters would be expected to form in any population where lineages emerge and disappear (Fraser et al., 2007; Doolittle and Zhaxybayeva, 2009). This problem was recently addressed in a metagenomic study, where the authors followed bacterial populations in a freshwater lake over nine consecutive years (Bendall et al., 2016). Notably, evidence of both accumulation and purging of SNPs within populations was found, underscoring the importance of sampling at multiple time-points.

Still, as of yet, studies following natural bacterial communities over time-frames long enough to evaluate the coherence of bacterial populations are rare, and it is therefore not known whether similar patterns hold true in other bacterial populations or environments. For example, different populations might be expected to display different patterns of coherence depending on their propensity for recombination, which is known to differ

by several orders of magnitude among species (Ochman et al., 2000; Fraser et al., 2007). Additionally, different habitats can be expected to support different levels of intra-species diversity and to be subject to distinct selection pressures.

In conclusion, while there is evidence from multiple studies to support that bacterial diversity is organized into discrete phenotypic and genetic clusters, there is no agreement as to how and whether such clusters correspond to bacterial species (Cohan, 2002). Meanwhile, however, scientists interested in grouping and analyzing the distribution of bacterial diversity from environmental samples seem to have arrived at a pragmatic solution; in the absence of a functional species concept, taxonomic profiling of bacterial communities is typically done by clustering 16S rRNA sequences into groups of 97% identity in the gene alignment (**Box 1**). These groups are subsequently referred to as “OTUs” (operational taxonomic units), thereby circumventing the need for the term “species.” Nevertheless, the 97% cut-off was originally proposed as a proxy for bacterial

**BOX 1 | 16S rRNA Profiling and Intra-Species Diversity.**

Culture-independent taxonomic profiling of unknown bacterial communities is most commonly done by PCR amplification of the 16S rRNA gene, and has provided invaluable information on the diversity and complexity of the bacterial kingdom. However, for the purpose of studying intra-species diversity, the method suffers from the following limitations:

**PCR Artifacts**

Several primer-sets have been designed to target the 16S rRNA gene across bacterial phyla. However, while the goal is to target all bacteria, different primer-sets are known to generate different results, due to differences in annealing and amplification efficiency (Kuczynski et al., 2012). Additionally, PCR amplification can generate chimeric sequences and artificial SNPs due to polymerase errors. Therefore, an important initial step of any 16S rRNA pipeline is a stringent quality control, including trimming and filtering.

**Clustering and Analysis**

Theoretically, any unique short 16S rRNA sequence within a sample can be assumed to correspond to at least one distinct evolutionary lineage. However, in practice, the distinction between methodological artifacts (i.e., PCR and sequencing errors) and true diversity becomes non-trivial when the informative number of SNPs is small (Kunin et al., 2010). 16S rRNA sequences are therefore usually clustered into groups, which are referred to as OTUs. While any similarity threshold can in principle be applied, the most common cut-off is 97% identity in the 16S rRNA alignment, which is widely taken as a proxy for bacterial species (Stackebrandt and Goebel, 1994). Several clustering algorithms are available, and are known to generate different numbers of OTUs, particularly if higher similarity thresholds are applied (Schmidt et al., 2014). The common practice of clustering 16S rRNA sequences at 97% identity or higher therefore helps to reduce the inflation of OTUs caused by methodological artifacts, but at the expense of not resolving intra-species diversity.

species, as it was found to correspond well to the 70% cut-off in DNA-DNA hybridization analysis (which was considered a molecular gold-standard for species delineation at the time; Stackebrandt and Goebel, 1994). Moreover, OTUs clustered at this level are generally analyzed as distinct functional units (akin to species), and several tools are available for estimating functional profiles from OTUs defined at this threshold (Langille et al., 2013; Asshauer et al., 2015). For simplicity, and due to the widespread application of the 97% sequence identity cut-off in community profiling studies, we will therefore refer to such groups as “species” or “species group” throughout this review, although we recognize the arbitrary nature of the classification. Furthermore, we will refer to all sequences falling within such groups as “strains.”

## INTRA-SPECIES DIVERSITY IN THE GUT MICROBIOTA

Since standard 16S rRNA profiling provides little resolution for inferring intra-species diversity (**Box 1**), our current knowledge on the extent and distribution of intra-species diversity in the gut microbiota is scarce. However, pioneering studies have been conducted on the human gut microbiota, which have provided exciting new insights. Pilot studies have also been done on the honeybee, likewise unraveling the presence of extensive intra-species diversity. Thus, in the following, we will describe what

is known about intra-species diversity in the gut microbiota, focusing primarily on these two model organisms.

## The Human Gut Microbiota

Culture-independent investigation of diversity in the human gut microbiota started with Sanger-sequencing of the 16S rRNA gene (Wilson and Blichington, 1996; Suau et al., 1999), which provides relatively long sequence fragments. Therefore, although the highly conserved 16S rRNA gene provides low resolution power for distinguishing closely related bacterial strains (Stackebrandt and Goebel, 1994), the presence of intra-species diversity was already noted in early studies, as the number of unique long 16S rRNA sequences was found to be several orders of magnitude higher than the number of inferred species groups (Eckburg et al., 2005; Ley et al., 2006; Dethlefsen et al., 2007). However, due to the cost and labor associated with Sanger-sequencing, the overall sequencing depth was initially shallow, and it was evident that only a small fraction of the total diversity had been sequenced (Eckburg et al., 2005). With the advent of next-generation sequencing technologies, it has become possible to perform large surveys on hundreds of individuals, and to sample the gut microbiota deeply (Human Microbiome Project C, 2012). Unfortunately, the shift from Sanger-sequencing to next-generation sequencing technologies has also resulted in the generation of shorter read fragments, thereby reducing the information content available of intra-species diversity analysis even further (Stackebrandt and Goebel, 1994), while also increasing the impact of PCR/sequencing artifacts (Kunin et al., 2010).

Nevertheless, studies employing short-read shotgun metagenomics have shown that the functional profile of the gut microbiota can change significantly even when the taxonomic profile display minor changes, thereby providing indirect evidence for the functional relevance of intra-species diversity (Morgan et al., 2012; Qin et al., 2012). Indeed, if the functional profile changes in the absence of major changes in the species profile, replacement or addition of functionally distinct strains within pre-existing species groups can be presumed to have occurred. Considering the large accessory gene content typical of bacterial species (as defined in metagenomic samples based on the 97% cut-off in 16S rRNA identity), it is perhaps not entirely surprising that functional changes can occur in the absence of changes in the taxonomic profile.

In 2013, two large-scale studies were published, in which intra-species diversity was quantified in a more direct manner, providing a first estimate of the scope of intra-species diversity occurring in the human gut microbiota (Faith et al., 2013; Schloissnig et al., 2013). Firstly, these studies showed that the majority of species in the gut microbiota harbor multiple strains, also within individual hosts. Secondly, by following host individuals over time, both studies found that strains persisted within their hosts for years, indicating that strains can co-exist in the gut in a stable manner. Finally, strains were found to be largely host-specific, with similar strains being shared among closely related individuals.

These studies raised several important questions about the gut microbiota (**Figure 1**): what is the functional relevance of

intra-species diversity in the gut? And how are host-specific profiles generated? On one hand, the host could potentially select strains with specific functional profiles. In fact, since the gut microbiota strain profile tends to be more similar among related individuals, it is even possible that strains could adapt to specific host backgrounds. On the other hand, host-specific strain profiles could also result from colonization bottlenecks early in life (Schloissnig et al., 2013), particularly if strains occupy similar niches in the gut.

Given the potentially large impact of early-life colonization events, several studies have by now been conducted on the infant gut microbiota. Overall, infants are known to have a more variable and less complex gut microbiota species composition than adults, with a higher abundance of bifidobacteria (Bäckhed et al., 2015). Nevertheless, based on comparisons between the gut microbiota of newborns and their mothers, it is clear that the maternal microbiota provides an important source of first-colonizers (Bäckhed et al., 2015). Whether strains acquired early in life also persist into adulthood or alternatively could be continuously exchanged among family members is not known. Interestingly, despite an overall low species complexity, the infant gut microbiota has also been shown to harbor extensive intra-species diversity (Sharon et al., 2013; Luo et al., 2015). Thus, Luo et al. (2015) found an average of 4.88 strains per subject in infants sampled during the first 3 years of life (as estimated based on SNP analysis within species groups).

In conclusion, several recent studies have reported on the presence of extensive intra-species diversity in the human gut microbiota. The functional relevance of this diversity is still unclear, but intra-species diversity has been associated with large differences in gene content, where variably present functions include transport, signaling and carbohydrate metabolism (Greenblum et al., 2015; Zhu et al., 2015).

With intra-species diversity being a common feature of the bacteria colonizing the gut, the molecular and experimental characterization of individual strains represents an enormous challenge. Moreover, most methods of strain characterization require culturing, which has widely been assumed to be an unrealistic prospect for the majority of gut bacteria. However, the “unculturability” of the gut microbiota has recently been challenged by two studies, where the authors in some cases were able to culture more than 90% of bacterial species present at more than 0.1% abundance (as estimated by 16S rRNA profiling; Browne et al., 2016; Lau et al., 2016). Thus, a wider range of tools for strain characterization of the gut microbiota have become available, which will be of great importance for elucidating the functional relevance of intra-species diversity in the human gut microbiota.

## The Honeybee Gut Microbiota

While the majority of studies on the gut microbiota have been conducted in mammalian model animals, invertebrates are becoming increasingly popular for functional studies, not least due to their experimental amenability. As for mammals, the invertebrate gut microbiota has been shown to play a role in nutrition and interactions with pathogens, indicating that these

are general functional attributes of the gut microbiota of the metazoa (Dishaw et al., 2014; Douglas, 2015).

However, in contrast to most mammals, invertebrates typically host gut bacterial communities with much lower species complexity (Colman et al., 2012; Engel and Moran, 2013). With some notable exceptions, such as termites (Colman et al., 2012; Brune and Dietrich, 2015), the number of species in the invertebrate gut is generally less than 50 species per host (Colman et al., 2012). Thus, the sequencing of metagenomic samples derived from the invertebrate gut can provide a deeper coverage of all microbiota members, and thereby greatly facilitate the study of intra-species diversity.

For example, the honeybee (*Apis mellifera*) has been shown to harbor 8–10 bacterial species groups in the gut, which usually constitute more than 98% of the total bacterial community (Martinson et al., 2011; Ahn et al., 2012; Moran et al., 2012; Sabree et al., 2012; Corby-Harris et al., 2014). However, similarly to the human gut microbiota, the core gut microbiota of the honeybee has been shown to include an impressive amount of intra-species diversity (Engel et al., 2012, 2014; Ellegaard et al., 2015). Thus, in the first metagenomic study employing shotgun metagenomics on a single honeybee colony, *de novo* genome assembly resulted in multiple sequence variants of phylogenetic marker genes within species groups (Engel et al., 2012). Moreover, the length of the assembled contigs for each species was found to be several orders of magnitude higher than the expected genome sizes, indicating the presence of multiple co-existing strains within species groups. Consistently, sequencing of genomes from cultured strains of lactobacilli and bifidobacteria, isolated from a single apiary, revealed that around one third of the gene content in each sequenced strain was variably present within species groups (Ellegaard et al., 2015). Likewise, a surprisingly high extent of sequence divergence and gene content variation was identified by single-cell genome sequencing in two proteobacterial species of the honeybee gut microbiota, *Gilliamella apicola* and *Snodgrassella alvi* (Engel et al., 2014). Although most of the variable gene content is of unknown function, a strong enrichment of genes related to carbohydrate metabolism and transport was observed in all three studies, suggesting that the presence of intra-species diversity allow for higher metabolic flexibility in the community.

So far, comparative metagenomic studies have not been conducted on the honeybee gut microbiota. But, a gene phylogeny incorporating sequences from genome and metagenome data from two independent studies (Engel et al., 2012; Ellegaard et al., 2015) confirmed the presence of four distinct sub-lineages within the *Lactobacillus apis* species group at both sampling locations. These data therefore suggest that the intra-species diversity of the honeybee gut microbiota is structured around distinct evolutionary lineages, which may have different functional roles.

The mechanisms involved in generating and maintaining intra-species diversity in the honeybee gut microbiota are not known, and the distribution of intra-species diversity among individuals of a given honeybee colony has not yet been studied in detail. However, based on full-length 16S rRNA gene sequences,

Moran et al. (2012) could show that colony-specific strains seem to exist. Moreover, in another study from the same group, deep-sequencing of the protein-encoding gene *minD* revealed that most individual honeybees host more than one strain of the gut symbiont *Snodgrassella alvi* (Powell et al., 2016). Multiple functional and spatial niches are potentially available within honeybee colonies to support specialization and co-existence. For example, honeybees go through distinct life-stages and perform different tasks within the colony, which is also associated with distinct diets (Seeley, 1985). Thus, a recent study found significant changes in strain abundance for five out of 25 *Lactobacillus* strains between honeybees sampled at the age of 3 and 7 days old (based on differential abundance of unique 16S rRNA sequences; Anderson et al., 2016).

Whether the patterns of intra-species diversity observed in the honeybee also hold true for other invertebrate species is not yet known. However, based on diversity within 16S rRNA profiling data, the gut microbiota of the *Cephalotes varians* ant was also found to harbor extensive intra-species diversity, which in this case was found to distribute geographically (Hu et al., 2014). Likewise, early Sanger-sequencing of the 16S rRNA gene in termites revealed extensive intra-species diversity in the termite gut microbiota (Warnecke et al., 2007).

In conclusion, intra-species diversity has been described in the gut microbiota of both mammals and insects, and several studies have been published which indicate that intra-species diversity is likely to be functionally relevant. Future studies will show to what extent this is also the case for other animals and thus may be a general, recurrent pattern in gut microbiota evolution. In the following section, we will discuss possible mechanisms that may be involved in shaping and maintaining intra-species diversity in the gut microbiota.

## WHAT DRIVES STRAIN DIVERSIFICATION, COEXISTENCE, AND COMMUNITY MAINTENANCE

A long-standing tenet in ecology, with roots dating back at least a century, hold that “complete competitors cannot coexist” (also known as “the competitive exclusion principle”; Hardin, 1960). Since strains of the same species are expected to have functionally overlapping gene repertoires, the co-existence of such lineages in natural communities is intriguing. Can closely related bacterial strains co-exist over long time-spans without out-competing each other, and if so, how?

### Micro-Niche Specialization

The simplest explanation for stable co-existence within the gut would arguably be that bacterial strains belonging to the same species group do in fact occupy distinct niches, despite being closely related. Indeed, several studies have shown that strains of the same species-group can be ecologically distinct within a shared habitat (Hunt et al., 2008; Gonzaga et al., 2012; Kashtan et al., 2014). Within the animal gut, such “micro-niche” specialization could be facilitated both by spatial heterogeneity and functional differentiation (Figure 2A).

The animal gut harbors a wide range of microbial habitats that vary in physiological parameters such as pH, oxygen and concentration of antimicrobial compounds (Donaldson et al., 2016). Thus, different bacterial families dominate the community of the small intestine and colon in humans (Donaldson et al., 2016), and the ileum and hindgut in honeybees (Martinson et al., 2012). However, marked local differences have also been demonstrated, i.e., between the outer mucus layer, inter-fold regions, colonic crypts and the luminal content in mice (Nava et al., 2011; Pedron et al., 2012; Li H. et al., 2015). Moreover, spatial structure in itself (i.e., patchy growth) is also expected to reduce the strength of interactions between populations and to dampen competition (Mitri and Foster, 2013; Coyte et al., 2015).

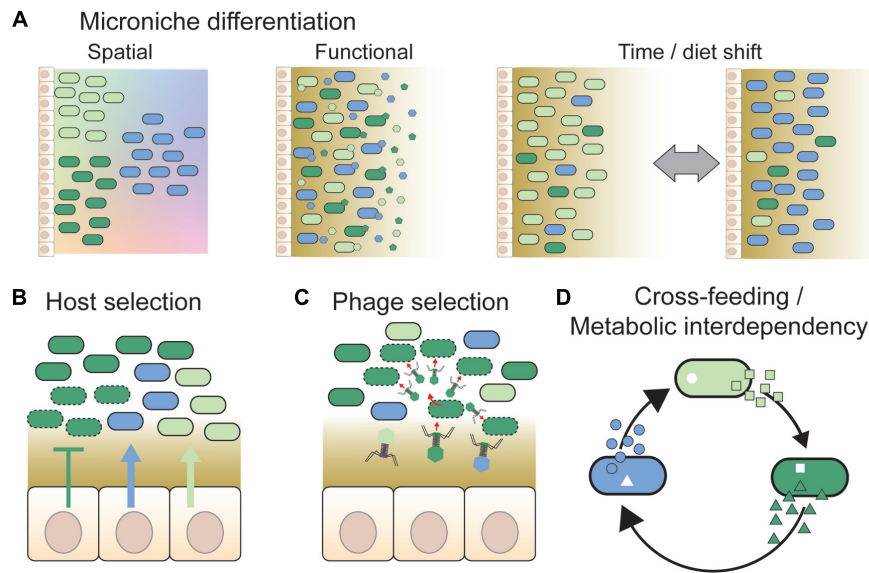
Functional differentiation, such as specialization toward different host- or diet-derived nutrients, may also facilitate co-existence between co-localized strains in the gut (Figure 2A). For example, strain diversity within *Bacteroides* spp. is associated with PULs (polysaccharide utilization loci; Xu et al., 2007; Zhu et al., 2015), which are used to breakdown distinct carbohydrates (Xu et al., 2007; Martens et al., 2014). Thus, although experimental studies on sugar utilization in *Bacteroides* spp. have so far been limited to cross-species comparisons (Rakoff-Nahoum et al., 2014, 2016), it is possible that such interactions also occur at the strain-level.

Finally, temporal variation in nutrient availability and composition may contribute to generating additional micro-niches within the gut (Figure 2A). Assuming that members of the gut microbiota can tolerate unfavorable conditions, at least for some time, a variable diet could potentially support a large number of specialized strains. For example, a large proportion of the human gut microbiota was recently shown to have spore-forming capabilities (Browne et al., 2016), which might ensure survival and resilience to perturbations. Moreover, temporal fluctuations in nutrient availability may also promote co-existence, by preventing any individual members of the gut microbiota from being at a competitive advantage over time-frames long enough to out-compete other community members (Gravel et al., 2011).

Since most of the gene content varying between bacterial strains is typically of unknown function, it is challenging to determine the extent of functional differentiation among strains of the same species group based on genomic data. Furthermore, it is unclear how much functional overlap can be tolerated under a scenario of stable co-existence. However, some studies suggest that subtle resource partitioning or functional differences may be sufficient for long-term co-existence (Tannock et al., 2012; Raghavan and Groisman, 2015).

### Host Selection

Unlike free-living bacterial communities, gut bacteria have to adapt to a biotic environment composed of a host, which may respond differentially to the presence of distinct bacteria. To what extent the host is able to control the composition of its gut microbiota is not known, but there is clear evidence for an ongoing cross-talk between the host immune response and the gut microbiota (Rooks and Garrett, 2016; Figure 2B). For example, both mammals and insects secrete antimicrobial peptides in



**FIGURE 2 | Overview of different mechanisms hypothesized to facilitate strain co-existence. (A)** Divergent strains of a given species may be adapted to distinct micro-niches. These could be spatial niches, e.g., along nutrient or physicochemical gradients (shown by color gradients in the background) or according to variation in surface properties in different gut regions (not shown). Alternatively, strains may be adapted to functional niches, such as different types of dietary or host-derived nutrients. In this case, strains need not be spatially separated from each other. Temporal variation in environmental conditions, such as changing dietary habits or intake of drugs, may also facilitate co-existence by expanding the number of available niches and changing the fitness of strains continuously. This is also expected to generate temporal variation in strain abundance and diversity. **(B)** The host may select specific strains by the release of nutrients (arrows) or antimicrobial peptides (T bars) at the epithelial surface (Schluter and Foster, 2012), or by providing specific binding sites for bacteria (McLoughlin et al., 2016). Dashed outlines indicate killed or growth-arrested cells due to the action of antimicrobial peptides. **(C)** Phages have previously been proposed to play an important role in maintaining strain diversity in natural bacterial populations, via “kill-the-winner dynamics” (Rodríguez-Valera et al., 2009). Likewise, they may prevent massive expansions of strains in the gut by targeting abundant strains preferentially, thereby facilitating the maintenance of strain diversity (Barr et al., 2013). **(D)** Co-existence of bacterial strains may be facilitated by cross-feeding interactions, i.e., one strain produces a metabolite that is beneficial for another strain (Zelezniak et al., 2015). Such interactions may result in metabolic interdependencies, if the obtained metabolites are indispensable for bacterial growth and cannot be acquired or produced by other means (Morris et al., 2012). In conclusion, several mechanisms have been proposed to facilitate strain co-existence; notably, these mechanisms are not mutually exclusive, and may act together.

the gut, which affect bacteria differentially (Dishaw et al., 2014; Bunker et al., 2015; Donaldson et al., 2016; Planer et al., 2016).

From a host perspective, it is crucial to maintain a beneficial microbiota, and thereby exclude pathogenic strains. Moreover, a microbiota in which beneficial functions are present in diverse genomic backgrounds might increase productivity and insure robustness against disturbance (Yachi and Loreau, 1999). On the bacterial side, theory predicts that competition with other members of the gut microbiota will render costly extracellular functions susceptible to “cheaters” (Mitri and Foster, 2013). Therefore, host selection could potentially be important for the maintenance of beneficial functions of the gut microbiota.

Although the taxonomic profile of the gut microbiota varies both within and between host species, some studies have shown that gut bacteria colonize their native host better than other hosts (Kwong et al., 2014a; Seedorf et al., 2014). The mechanisms underlying this specificity are not clear, but it seems likely that bacterial outer-surface structures are involved. For example, intra-species diversity in adherence and aggregation factors correlates with host specificity in *L. reuteri* (Frese et al., 2011). Interestingly, while bacteria are expected to adapt their outer-surface structures to be able to colonize and attach to their host, it was recently proposed that the host might also exploit such

structures to control the microbiota composition (McLoughlin et al., 2016). Specifically, since mammalian mucins are heavily glycosylated, and many bacteria have the ability to attach to glycans, the host could potentially promote and ensure the survival of beneficial bacteria via controlled glycan secretions (McLoughlin et al., 2016). Similarly, since several gut bacteria feed on epithelial-derived nutrients, it has also been proposed that the host may be able to promote the growth of specific groups of bacteria by providing selective nutrients (Schluter and Foster, 2012).

Based on these ideas, Donaldson et al. (2016) proposed that through “localized, immune-facilitated and adherence-dependent nutrient selection, the host maintains the stability of a diverse community of microbial symbionts.” However, whether such host selection can extend as far as to the maintenance of intra-species strain diversity in the gut remains speculative.

## Phage Selection

In an interesting twist, it has also been proposed that the maintenance of strain diversity at the epithelial gut surface is maintained by a third player, namely bacteriophages (Barr et al., 2013) (Figure 2C). Phages have previously been proposed to play an important role in maintaining strain diversity in

natural bacterial populations, via “kill-the-winner” dynamics (Rodriguez-Valera et al., 2009; Koskella and Brockhurst, 2014; Thingstad et al., 2014; Takeuchi et al., 2015). According to this hypothesis, bacterial strains increasing in abundance are also more likely to be targeted by phages, thereby causing negative frequency-dependent selection. Although phages remain highly understudied, they do appear to be present in high abundance within the gut, just as in most other habitats (Ogilvie and Jones, 2015). Notably, in the study by Barr et al. (2013), a particularly high phage-to-bacteria ratio was found in the mucus layer of diverse animals (using epifluorescence microscopy). Moreover, bacterial attachment to mucus-producing cells was significantly reduced *in vitro*, when the mucus-producing cells were pre-treated with the mucus-adherent phage T4. Phage predation could therefore be of importance both for maintaining the mucosal barrier against invading pathogens, and for promoting and maintaining diversity within the gut microbiota.

Consistent with a phage predation model, Sharon et al. (2013) did indeed find patterns of co-variation in phage and strain abundance in the relatively simple gut microbiota of a premature infant. On the other hand, another study, likewise following strain abundance in the infant human gut microbiota over time, found that different species displayed different abundance dynamics, with some species maintaining a constant hierarchy of strain abundance over time, whereas other species displayed more variable strain abundance profiles (Luo et al., 2015). Furthermore, the strain abundance variation in one species group was accompanied by functional differences in sugar utilization genes, suggesting that dietary changes were involved, rather than phages.

As of yet, large-scale studies of co-variance in phage-strain abundance have not been conducted, since both quantification of strains and phages are technically challenging (Ogilvie and Jones, 2015). Notably, the presence of phages in the gut in itself need not result in “kill-the-winner” dynamics, since phages can also display a temperate life-cycle in which insertion into host genomes is preferred over host lysis (Reyes et al., 2010; Ogilvie and Jones, 2015).

## Inter-Dependence and Cross-Feeding Interactions

According to classic ecological principles, co-operation can be expected to be common among bacteria of the same genotype, whereas competition should prevail among different genotypes, i.e., strains of the same species-group (Mitri and Foster, 2013). Nevertheless, according to “the Black Queen Hypothesis,” recently proposed by (Morris et al., 2012), the evolution of co-operation between members in bacterial communities can be expected to occur if bacterial functions are “leaky,” resulting in the availability of “common goods.” Firstly, based on genome streamlining theory (Giovannoni et al., 2014), bacteria are expected to lose gene functions that are not strictly needed, due to a fitness advantage associated with avoiding the cost of maintaining the function. For example, as a proof of principle, it was shown that deletion of complementary metabolic genes for amino acid biosynthesis in two strains of

*Escherichia coli* resulted in increased Darwinian fitness of the strains while grown in co-culture (Pande et al., 2014). Therefore, bacterial species may become dependent on other members of their community for providing specific functions. Moreover, if reciprocal losses occur, a network of inter-dependencies may evolve, preventing individual members from out-competing others (Figure 2D).

Since the presence of such dependencies is difficult to predict solely from genomic data, there is currently little evidence for their existence. However, while the impact of “black queen dynamics” on the evolution and maintenance of the gut microbiota is not known, cross-feeding interactions, whereby the waste-product of one bacterium becomes a nutrient source for another, are likely to be common in the gut (Fischbach and Sonnenburg, 2011; Zelezniak et al., 2015). Indeed, cross-feeding interactions have been demonstrated *in vitro* among members of the *Bacteroidetes* genus (Rakoff-Nahoum et al., 2014, 2016). Similarly, bacterial strains may benefit from each other's presence without being strictly dependent on each other.

## Concluding Remarks

Several different mechanisms have been proposed to be involved in the maintenance of diversity within the gut microbiota (Figures 2A–D), but their relative importance for shaping and maintaining intra-species diversity is largely unknown. Is the gut simply an environment, which supports a high degree of micro-niche differentiation? Or do bacterial strains depend on or benefit from each other's presence? And to what extent is the host involved in controlling and manipulating these highly complex communities? In order to answer these questions, we need to be able to quantify and follow the abundance of bacterial strains over time and in response to experimental manipulation, directly from bacterial communities. In the following sections, we will describe current methods employed to study intra-species diversity and dynamics directly from metagenomic samples.

## CULTURE-INDEPENDENT METHODS FOR INVESTIGATING INTRA-SPECIES DIVERSITY

### PCR-Based Community Profiling

Taxonomic profiling of the gut microbiota is most commonly done based on PCR amplification of the 16S rRNA gene. However, for the study of intra-species diversity, the 16S rRNA profiling method provides limited resolution, and is susceptible to biases introduced by sequencing and PCR artifacts (Kunin et al., 2010; Box 1). To address this problem, Faith et al. (2013) developed a method termed “LEA-seq” (low-error amplicon sequencing). In this method, template sequences are tagged with barcodes in an initial linear PCR amplification step, followed by regular PCR amplification. After sequencing the final PCR products, the sequences are aligned in groups according to their barcodes, from which consensus-sequences can be derived and true variants can be inferred with much higher confidence. However, the read correction approach comes at the cost of a



strongly decreased sampling depth, since each target has to be sequenced multiple times to obtain reliable consensus sequences.

Given the low information content of the 16S rRNA gene, primers targeting alternative genes have also been developed for studies on intra-species diversity. Although such primers can only target a subset of any community, they can provide a high level of resolution, and have been successfully applied to demonstrate the presence of intra-species diversity and co-diversification on shorter evolutionary time-scales in several recent papers (Lee et al., 2014; Caro-Quintero and Ochman, 2015; Moeller et al., 2016; Powell et al., 2016). However, regardless of the gene target, all PCR profiling methods are expected to induce skewed community profiles, due to differences in annealing and amplification efficiency of primers relative to their targets. Comparisons between samples can therefore only be done when the same primers and PCR conditions are applied. Moreover, in the case of the 16S rRNA gene, copy numbers differ widely between species, contributing further to skewed abundance profiles (Kembel et al., 2012).

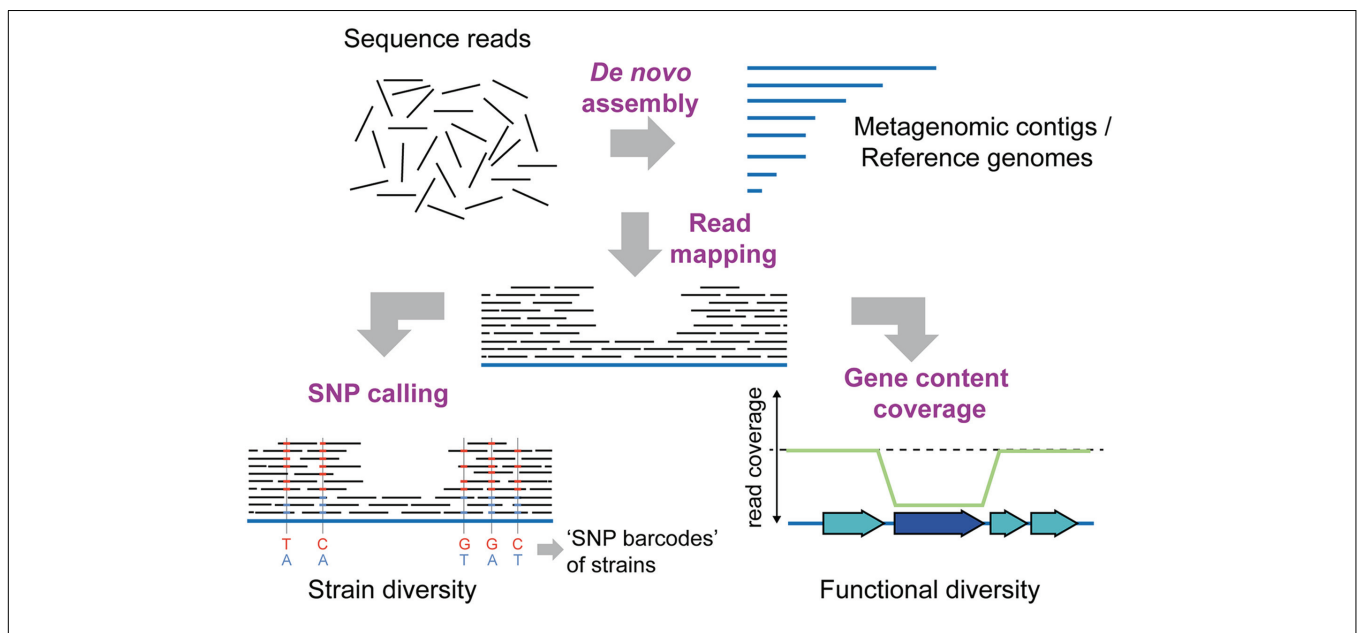
Aside from PCR-associated artifacts, the targeting of single genes for diversity and evolutionary analyses is also affected by the variable rates of gene evolution in different bacterial species. For example, a comparison of the 16S rRNA identity and ANI (average nucleotide identity) scores between species showed that the core genome and the 16S rRNA gene evolve at different relative rates in different species (Kim et al., 2014). Therefore,

even the 16S rRNA gene does not accurately reflect the evolution of bacterial communities. Finally, PCR-based profiling does not provide direct functional information pertaining to the diversity of a community.

## SNP-Based Analyses

Using shotgun metagenomics, sequencing data is obtained from complete genomes within a sample. Thus, although shotgun metagenomics has mostly been used to derive functional profiles of bacterial communities, this approach also has the potential to provide a high level of resolution for intra-species diversity analysis.

Due to the targeting of complete genomes, shotgun metagenomics require a much higher sequencing effort than 16S rRNA profiling, in order to obtain a representative sample. Accurate high-throughput short-read sequencing technologies are therefore usually chosen. By mapping metagenomic reads, SNPs can be called on genes in reference genomes or on *de novo* assembled metagenomic contigs (Figure 3). Thereby, the extent of intra-species diversity can be estimated, and genes under positive or diversifying selection may be identified (Schloissnig et al., 2013; Bendall et al., 2016). For example, Schloissnig et al. (2013) were able to identify 10.3 million SNPs in 101 prevalent bacterial species, across 207 individuals, with bacterial species displaying highly variable levels of intra-species diversity. Moreover, by following 43 individuals over time, the authors



**FIGURE 3 | Shotgun metagenomics to determine intra-species diversity in the gut microbiota.** Short-read sequencing technologies (100–150 bp), which provide high throughput and base-calling accuracy, are usually preferred for shotgun metagenomics. To infer intra-species diversity, sequence reads can be mapped to a database of reference sequences. Reference sequences can be genomes of previously sequenced cultured isolates or contigs obtained from *de novo* assembly of the metagenomic sequence reads themselves. Two distinct types of analyses are possible after read mapping: SNP (single nucleotide polymorphism) calling and gene content coverage analyses. SNP calling allows for determining the percentage of variable sites in reference sequences, and can provide insights into the extent of strain diversity for the different species in the community (in this case 5 sites, SNP variants indicated in blue and red colors). Based on the coverage of SNPs, strain-specific “bar-codes” may also be inferred (Luo et al., 2015). Thereby, the number of strains present in a sample can be estimated, and the abundance of strains can be followed over time in different samples. Alternatively, analysis of gene content coverage allows for determining gene sets and functions that vary in abundance in strains in the community. Genes not present in every strain in the community will have a lower read coverage than the remaining reference genome (dark blue gene). Similarly, genes present in multiple copies within the community will have a higher coverage than the rest of the reference genome (not shown).

could show that strain profiles were host-specific and persisted over a sampling period of 1 year.

While a global analysis of SNP diversity within metagenomic samples is of interest on its own, a higher goal would be to infer the presence of specific strains or lineages, and follow their abundance over time. A large number of genomes have by now been sequenced for many species, particularly from the human gut microbiota, which should facilitate this type of analysis. Unfortunately, the correct mapping of short reads among closely related strains is at best difficult. Tools have therefore recently been developed to improve the mapping accuracy in databases containing genomes from closely related bacteria, using statistical or probabilistic models to re-assign ambiguously mapped reads to their most likely origin genome (Hong et al., 2014; Ahn et al., 2015). In this manner, it is possible to follow the abundance and spread of known strains (or strains recently diverged from known strains), for example during a disease outbreak. Still, the efficiency of these tools is highly dependent on the availability of suitable reference genomes, since only strains with closely related sequenced genomes can be followed.

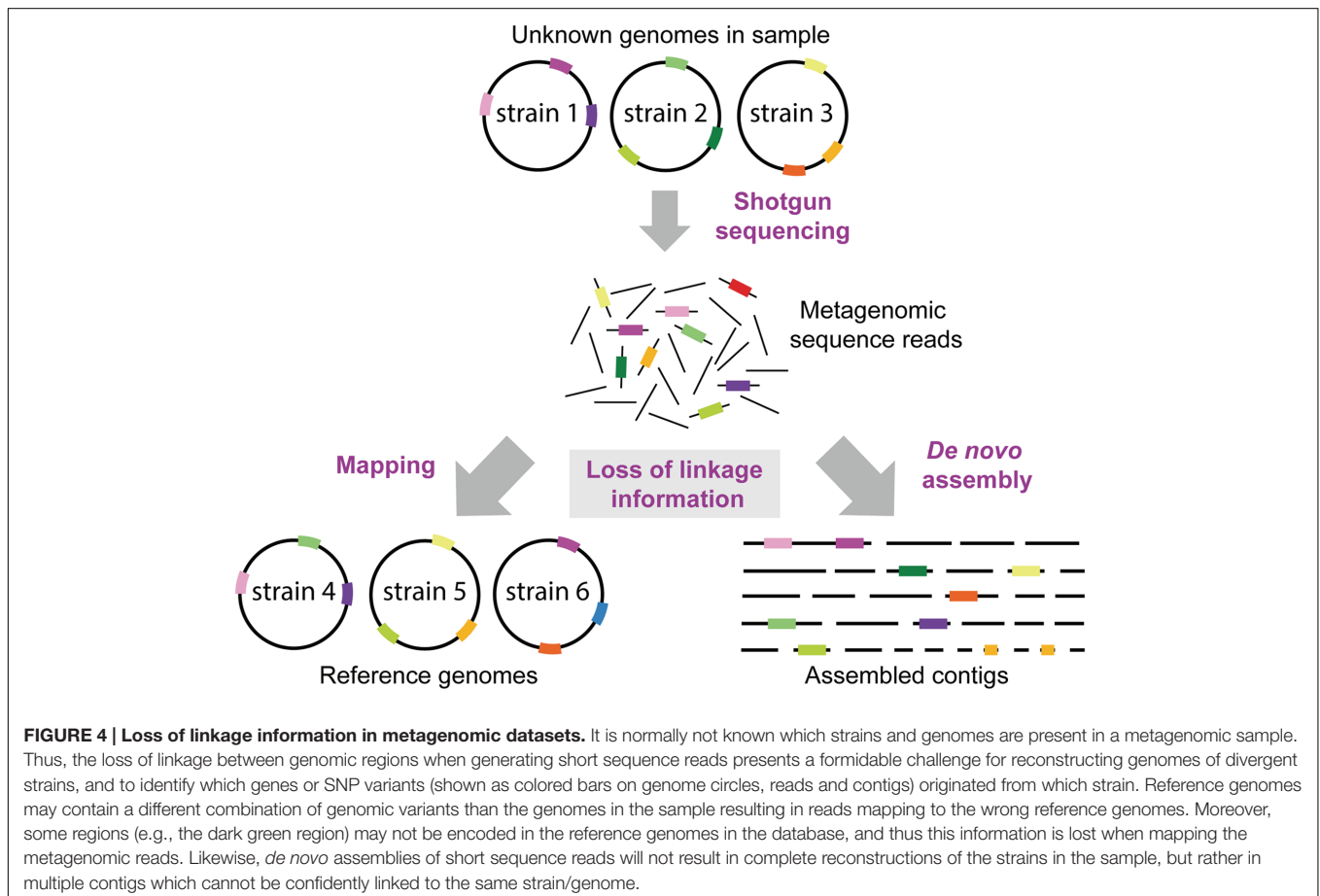
To reduce this dependency, Luo et al. (2015) recently published the tool “Constrains,” which attempts to quantify the most likely number of strains present in a sample. In this method, species with sufficiently high coverage for strain inference (set

to 10x coverage) are initially identified using Metaphlan (Segata et al., 2012). Next, a custom database of representative marker genes is generated for each species, against which all reads are mapped. After mapping, the reference gene sequences are removed, and a table of coverage by base-position is created for all variable positions. Thus, only a single reference genome per species is necessary (provided that reads from divergent strains can still be confidently mapped to the genome). Since SNPs originating from the same strain are expected to have similar coverage within samples, model “barcodes” of SNPs with similar coverage can be generated and compared across samples, to estimate the most likely strain-specific SNP-barcode (Figure 3). Moreover, when samples are available for multiple time-points, the abundance of strains (i.e., barcodes) can be followed over time.

In conclusion, SNP-based analysis of shotgun metagenomic data provides a powerful framework for quantifying intra-species diversity, selection and abundance dynamics. However, since all current SNP-based methods are based on core genes or marker genes, the functional profiles of quantified strains are not known, complicating the interpretation of abundance fluctuations.

## Gene-Content Analyses

Given the lack of functional information provided by PCR profiling and SNP calling methods, several recent methods have



**BOX 2 | Novel Sequencing Technologies For Metagenomics.**

Since a major obstacle for intra-species diversity analysis of metagenomic samples is the lack of linkage between SNPs/genes and their corresponding bacterial genomes (**Figure 4**), the development of sequencing technologies that can generate longer reads has the potential to greatly advance the field. Several novel technologies are currently under development, some of which have already become commercially available, with demonstrated improvements on applications such as *de novo* genome assembly (Chin et al., 2013; Loman et al., 2015). However, as of yet, short-read sequencing technologies are still preferred for metagenomic studies due to superior base-calling accuracy, high throughput and cost-efficiency.

Base-calling accuracy is fundamentally important for metagenomic studies, where multiple closely related bacterial strains are sequenced from the same sample. This requirement is currently not met by long-read sequencing technologies such as Single Molecule Real-Time (SMRT) sequencing (Chin et al., 2013) and Nanopore sequencing (Loman et al., 2015). In contrast, TruSeq synthetic long reads (TSLR) sequencing, recently introduced by Illumina, can generate long reads (8–10 kb) with very high base-calling accuracy (McCoy et al., 2014), making this technology interesting for metagenomic applications. Pilot metagenomic studies using TSLR reads have already yielded promising results, although taxonomic representation biases have also been noted (Sharon et al., 2015; Kuleshov et al., 2016).

Finally, metagenomic analysis targeting complete genomes also require very high throughput. Even with Illumina HiSeq sequencing, which is currently the most cost-effective and large-scale sequencing platform available, insufficient coverage is frequently a problem for strain-level analysis. For example, in a re-analysis of publicly available metagenomic data from the Human Microbiome Project, intra-species gene content variation was investigated for only 11 bacterial species, since all other species did not have sufficiently high coverage for reliable strain quantification (Zhu et al., 2015).

been developed to study intra-species gene content variation, based on shotgun metagenomics data.

In order to determine the gene content within a sample, the sequences can either be assembled *de novo* or mapped against a reference database (**Figure 3**). *De novo* assembly is desirable, since it allows for the discovery of novel genes or genomes in a sample. However, the assembly of short reads from complex bacterial communities is highly challenging. Firstly, the use of short reads is a general problem for genome assembly, since any repetitive sequences will hinder the assembly. This problem is exacerbated in samples from complex bacterial communities, where different species will have different optimal assembly parameters. Moreover, sequences from related strains may assemble into single or multiple contigs, depending on the conservation of the genomic region. Therefore, although several assemblers have by now been developed specifically for metagenomic data (Boisvert et al., 2012; Peng et al., 2012; Afiahayati and Sakakibara, 2015; Li D. et al., 2015; Nurk et al., 2016), metagenomic *de novo* assembly typically results in “a bag of genes” which are largely unlinked (**Figure 4**).

If a representative database of previously sequenced genomes is available, reference genome mapping can be a viable alternative. Relative to *de novo* assembly, read mapping represents a much less complex computational problem and several mapping tools are available that can efficiently handle even very large datasets (Hattem et al., 2013). However, due to the frequency of horizontal gene transfer in bacterial populations, reads mapping to the same genome in a reference database

cannot be assumed to be linked within the sample. In fact, horizontal gene transfer is expected to be particularly frequent among closely related bacteria (Majewski, 2001). Thus, the genetic linkage within a metagenomic sample is lost in both *de novo* assembly and reference genome mapping (**Figure 4**).

Nevertheless, the presence/absence of genes in a metagenomic sample relative to a reference genome database can readily be determined, based on mapped read coverage (Zhu et al., 2015). Alternatively, gene content diversity within samples can be estimated by comparing mapped read coverage on genes to conserved single-copy genes, such as the universally conserved ribosome-related genes (Greenblum et al., 2015) (**Figure 3**). Similarly, when a suitable reference genome database is available, the pan genome can be estimated and genes with higher or lower coverage relative to the core genes identified (Scholz et al., 2016). Finally, the widely used tool “Metaphlan,” originally designed to characterize the species composition of metagenomic samples based on mapped read coverage of clade-specific marker genes (Segata et al., 2012), was recently updated to detect and track strains in addition to species (Truong et al., 2015).

In conclusion, all current methods for inferring intra-species gene content variation in metagenomic samples are based on mapped read coverage on genes, and therefore depend on the availability of a suitable reference database or a representative *de novo* metagenome assembly. In practice, it is challenging to evaluate the accuracy of these methods in real-life metagenomic samples; for example, mapped read coverage can depend on the location of genes within their host genomes, since growing bacterial populations will have a higher gene coverage if they are located near the origin of replication relative to the terminus (Korem et al., 2015). Finally, while gene content variation analysis does provide functional information at the community level, the identified variably present functions remain unlinked to their origin bacterial strains within the community (**Figure 4, Box 2**).

**BOX 3 | Key Questions Related To Intra-Species Diversity In The Gut Microbiota.**

- Evolution
  - How do strains diversify in the gut?
  - How do strains adapt to specific host environments?
  - What is the impact of selection pressure exerted by the host versus other bacterial community members?
- Function
  - Do specific strains contribute distinct complementary functions in the gut?
  - Does the functional contribution of specific strains depend on host-specific factors or the presence of other community members?
  - Does a high level of intra-species diversity benefit the host, i.e., by increased productivity or by providing robustness against disturbances?
- Community assembly and interactions
  - How are communities with high levels of intra-species diversity established and maintained?
  - Do strains of the same species group occupy distinct niches in the gut?
  - Do strains of the same species group compete or collaborate?
  - Which mechanisms facilitate stable co-existence within the gut?

## FUNCTIONAL RELEVANCE OF INTRA-SPECIES DIVERSITY – FUTURE AVENUES OF RESEARCH

Based on several recent studies, it is now clear that intra-species diversity is omnipresent in the gut microbiota, also within host individuals (Faith et al., 2013; Schloissnig et al., 2013; Greenblum et al., 2015; Zhu et al., 2015; Kuleshov et al., 2016). However, little is known about how such diversity is generated or the mechanisms involved in obtaining, assembling and stabilizing diverse communities. Moreover, although several studies have shown that intra-species diversity also encompasses substantial gene content variation, the functional relevance of this diversity is not clear. A major problem in this context is the lack of accurate functional annotations, which is particularly common for genes that are variably present within species. Indeed, a major fraction of such genes typically cannot even be assigned a broad general function (Nayfach et al., 2015; Zhu et al., 2015; Oh et al., 2016). Therefore, in order to move the field forward from quantifying intra-species diversity to identifying causes and consequences of such diversity, we believe it will be crucial to link patterns of intra-species diversity to experimental and phenotypic data (Franzosa et al., 2015). To this end, experimentally amenable model systems with overall lower complexity of microbial diversity in the gut, such as the honeybee, can be a valuable addition to established vertebrate model species.

Although the honeybee is a relatively new model species for gut microbiota research, several tools have already been established (Kwong and Moran, 2016). Firstly, all members of the honeybee gut microbiota can now be cultured (e.g., Engel et al., 2013; Kwong and Moran, 2013; Olofsson et al., 2014; Kesnerova et al., 2016), and extensive frozen stock libraries are under development in several labs. Secondly, protocols for generating microbiota-free bees and re-colonizing bees have been established and applied successfully in a number of studies (Kwong et al., 2014a; Powell et al., 2014; Engel et al., 2015a). Thus, strain interactions can be queried both *in vitro* and *in vivo*. Thirdly, a large number of genomes isolated from the honeybee gut have by now been sequenced and analyzed, paving the way for future bioinformatic studies (Bottacini et al., 2012; Kwong et al., 2014a,b; Milani et al., 2014; Ellegaard et al., 2015; Engel et al.,

2015b). Finally, in comparison to other established invertebrate models, such as *Drosophila*, the honeybee differs in being a social insect. Thus, the impact of social interactions on gut microbiota colonization and maintenance can be tested, either by restricting possibilities for social interactions experimentally (Powell et al., 2014) or by tracking interactions directly.

A combination of shotgun metagenomics and experimental data from animal models has the potential to provide novel insights into many of the currently outstanding questions concerning intra-species diversity in the gut microbiota (Box 3). In addition to providing distinct technical advantages, the use of diverse animal models will be of particular importance for understanding differences and commonalities of gut microbiota evolution across the metazoa. Likewise, a better understanding of fundamental mechanisms involved in maintaining and establishing diverse gut bacterial communities will be indispensable for establishing safe and effective probiotics to cure disease. For example, a recent study found that the success of colonization after fecal microbiota transplantation differed between hosts when using the same donor sample (Li et al., 2016). Moreover, bacterial strains from the donor sample were more successfully established in the new host when strains of the same species were already present (Li et al., 2016), highlighting once again the need to understand interactions at the strain level, also from an applied perspective. With the rapid development of sequencing technologies, computational tools and large-scale culture-based methods, exciting insights into the intricate relationships within natural bacterial communities will undoubtedly soon change and challenge the current views on our bacterial world.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

Financial support was provided by the SNSF to PE (grant no. 31003A\_160345) and the University of Lausanne.

## REFERENCES

- Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L., et al. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430, 551–554. doi: 10.1038/nature02649
- Afahayati, K. S., and Sakakibara, Y. (2015). MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Res.* 22, 69–77. doi: 10.1093/dnares/dsu041
- Ahn, J. H., Hong, I. P., Bok, J. I., Kim, B. Y., Song, J., and Weon, H. Y. (2012). Pyrosequencing analysis of the bacterial communities in the guts of honey bees *Apis cerana* and *Apis mellifera* in Korea. *J. Microbiol.* 50, 735–745. doi: 10.1007/s12275-012-2188-0
- Ahn, T. H., Chai, J., and Pan, C. (2015). Sigma: strain-level inference of genomes from metagenomic analysis for biosurveillance. *Bioinformatics* 31, 170–177. doi: 10.1093/bioinformatics/btu641
- Anderson, K. E., Rodrigues, P. A., Mott, B. M., Maes, P., and Corby-Harris, V. (2016). Ecological succession in the honey bee gut: shift in *Lactobacillus* strain dominance during early adult development. *Microb. Ecol.* 71, 1008–1019. doi: 10.1007/s00248-015-0716-2
- Asshauer, K. P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31, 2882–2884. doi: 10.1093/bioinformatics/btv287
- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* 17, 690–703. doi: 10.1016/j.chom.2015.05.012
- Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10771–10776. doi: 10.1073/pnas.1305923110
- Bendall, M. L., Stevens, S. L., Chan, L. K., Malfatti, S., Schwientek, P., Tremblay, J., et al. (2016). Genome-wide selective sweeps and gene-specific sweeps in

- natural bacterial populations. *ISME J.* 10, 1589–601. doi: 10.1038/ismej.2015.241
- Bik, E. M., Costello, E. K., Switzer, A. D., Callahan, B. J., Holmes, S. P., Wells, R. S., et al. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nat. Commun.* 7:10516. doi: 10.1038/ncomms10516
- Billler, S. J., Berube, P. M., Lindell, D., and Chisholm, S. W. (2015). *Prochlorococcus*: the structure and function of collective diversity. *Nat. Rev. Microbiol.* 13, 13–27. doi: 10.1038/nrmicro3378
- Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., and Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol.* 13:R122. doi: 10.1186/gb-2012-13-12-r122
- Booth, A., Mariscal, C., and Doolittle, W. F. (2016). The modern synthesis in the light of microbial genomics. *Annu. Rev. Microbiol.* doi: 10.1146/annurev-micro-102215-095456 [Epub ahead of print].
- Bottacini, F., Milani, C., Turrone, F., Sanchez, B., Foroni, E., Duranti, S., et al. (2012). *Bifidobacterium asteroides* PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS ONE* 7:e44229. doi: 10.1371/journal.pone.0044229
- Browne, H. P., Forster, S. C., Anonyer, B. O., Kumar, N., Neville, B. A., Stares, M. D., et al. (2016). Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546. doi: 10.1038/nature17645
- Brune, A., and Dietrich, C. (2015). The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annu. Rev. Microbiol.* 69, 145–166. doi: 10.1146/annurev-micro-092412-155715
- Bunker, J. J., Flynn, T. M., Koval, J. C., Shaw, D. G., Meisel, M., McDonald, B. D., et al. (2015). Innate and adaptive humoral responses coat distinct commensal bacteria with immunoglobulin A. *Immunity* 43, 541–553. doi: 10.1016/j.immuni.2015.08.007
- Caro-Quintero, A., and Konstantinidis, K. T. (2012). Bacterial species may exist, metagenomics reveal. *Environ. Microbiol.* 14, 347–355. doi: 10.1111/j.1462-2920.2011.02668.x
- Caro-Quintero, A., and Ochman, H. (2015). Assessing the unseen bacterial diversity in microbial communities. *Genome Biol. Evol.* 7, 3416–3425. doi: 10.1093/gbe/evv234
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Cohan, F. M. (2002). What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487. doi: 10.1146/annurev.micro.56.012302.160634
- Colman, D. R., Toolson, E. C., and Takacs-Vesbach, C. D. (2012). Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* 21, 5124–5137. doi: 10.1111/j.1365-294X.2012.05752.x
- Corby-Harris, V., Maes, P., and Anderson, K. E. (2014). The bacterial communities associated with honey bee (*Apis mellifera*) foragers. *PLoS ONE* 9:e95056. doi: 10.1371/journal.pone.0095056
- Cordero, O. X., and Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12, 263–273. doi: 10.1038/nrmicro3218
- Coyte, K. Z., Schluter, J., and Foster, K. R. (2015). The ecology of the microbiome: networks, competition, and stability. *Science* 350, 663–666. doi: 10.1126/science.aad2602
- Delsuc, F., Metcalf, J. L., Wegener Parfrey, L., Song, S. J., Gonzalez, A., and Knight, R. (2014). Convergence of gut microbiomes in myrmecophilous mammals. *Mol. Ecol.* 23, 1301–1317. doi: 10.1111/mec.12501
- Dethlefsen, L., McFall-Ngai, M., and Relman, D. A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449, 811–818. doi: 10.1038/nature06245
- Dishaw, L. J., Cannon, J. P., Litman, G. W., and Parker, W. (2014). Immune-directed support of rich microbial communities in the gut has ancient roots. *Dev. Comp. Immunol.* 47, 36–51. doi: 10.1016/j.dci.2014.06.011
- Donaldson, G. P., Lee, S. M., and Mazmanian, S. K. (2016). Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* 14, 20–32. doi: 10.1038/nrmicro3552
- Doolittle, W. F. (2012). Population genomics: how bacterial species form and why they don’t exist. *Curr. Biol.* 22, R451–R453. doi: 10.1016/j.cub.2012.04.034
- Doolittle, W. F., and Zhaxybayeva, O. (2009). On the origin of prokaryotic species. *Genome Res.* 19, 744–756. doi: 10.1101/gr.086645.108
- Douglas, A. E. (2014). Symbiosis as a general principle in eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* 6:a016113. doi: 10.1101/cshperspect.a016113
- Douglas, A. E. (2015). Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* 60, 17–34. doi: 10.1146/annurev-ento-010814-020822
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638. doi: 10.1126/science.1110591
- Ellegaard, K. M., Tamarit, D., Javelind, E., Olofsson, T. C., Andersson, S. G. E., and Vasquez, A. (2015). Extensive intra-phylogroup diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* 16:284. doi: 10.1186/S12864-015-1476-6
- Engel, P., Bartlett, K. D., and Moran, N. A. (2015a). The Bacterium *Frischella perrara* causes scab formation in the gut of its Honeybee host. *MBio* 6, e193–e115. doi: 10.1128/mBio.00193-15
- Engel, P., Kwong, W. K., and Moran, N. A. (2013). *Frischella perrara* gen. nov., sp. nov., a gammaproteobacterium isolated from the gut of the honeybee, *Apis mellifera*. *Int. J. Syst. Evol. Microbiol.* 63(Pt. 10), 3646–3651. doi: 10.1099/ijs.0.049569-0
- Engel, P., Martinson, V. G., and Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11002–11007. doi: 10.1073/pnas.1202970109
- Engel, P., and Moran, N. A. (2013). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiol. Rev.* 37, 699–735. doi: 10.1111/1574-6976.12025
- Engel, P., Stepanauskas, R., and Moran, N. A. (2014). Hidden diversity in honey bee gut symbionts detected by single-cell genomics. *PLoS Genet.* 10:e1004596. doi: 10.1371/journal.pgen.1004596
- Engel, P., Vizcaino, M. L., and Crawford, J. M. (2015b). Gut symbionts from distinct hosts exhibit genotoxic activity via divergent colibactin biosynthesis pathways. *Appl. Environ. Microbiol.* 81, 1502–1512. doi: 10.1128/AEM.03283-14
- Eyre-Walker, A., and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8, 610–618. doi: 10.1038/nrg2146
- Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., et al. (2013). The long-term stability of the human gut microbiota. *Science* 341:1237439. doi: 10.1126/science.1237439
- Falush, D., Wirth, T., Linz, B., Pritchard, J. K., Stephens, M., Kidd, M., et al. (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science* 299, 1582–1585. doi: 10.1126/science.1080857
- Fischbach, M. A., and Sonnenburg, J. L. (2011). Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* 10, 336–347. doi: 10.1016/j.chom.2011.10.002
- Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C., et al. (2015). Sequencing and beyond: integrating molecular ‘omics’ for microbial community profiling. *Nat. Rev. Microbiol.* 13, 360–372. doi: 10.1038/nrmicro3451
- Fraser, C., Hanage, W. P., and Spratt, B. G. (2007). Recombination and the nature of bacterial speciation. *Science* 315, 476–480. doi: 10.1126/science.1127573
- Frese, S. A., Benson, A. K., Tannock, G. W., Loach, D. M., Kim, J., Zhang, M., et al. (2011). The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genet.* 7:e1001314. doi: 10.1371/journal.pgen.1001314
- Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565. doi: 10.1038/ismej.2014.60
- Gonzaga, A., Martin-Cuadrado, A. B., Lopez-Perez, M., Megumi Mizuno, C., Garcia-Heredia, I., Kimes, N. E., et al. (2012). Polyclonality of concurrent natural populations of *Alteromonas macleodii*. *Genome Biol. Evol.* 4, 1360–1374. doi: 10.1093/gbe/evs112
- Gravel, D., Guichard, F., and Hochberg, M. E. (2011). Species coexistence in a variable world. *Ecol. Lett.* 14, 828–839. doi: 10.1111/j.1461-0248.2011.01643.x
- Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. *Cell* 160, 583–594. doi: 10.1016/j.cell.2014.12.038
- Hardin, G. (1960). The competitive exclusion principle. *Science* 131, 1292–1297. doi: 10.1126/science.131.3409.1292
- Hatem, A., Bozdogan, D., Toland, A. E., and Catalyurek, U. V. (2013). Benchmarking short sequence mapping tools. *BMC Bioinformatics* 14:184. doi: 10.1186/1471-2105-14-184

- Hong, C., Manimaran, S., Shen, Y., Perez-Rogers, J. F., Byrd, A. L., Castro-Nallar, E., et al. (2014). PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* 2:33. doi: 10.1186/2049-2618-2-33
- Hu, Y., Lukasik, P., Moreau, C. S., and Russell, J. A. (2014). Correlates of gut community composition across an ant species (*Cephalotes varians*) elucidate causes and consequences of symbiotic variability. *Mol. Ecol.* 23, 1284–1300. doi: 10.1111/mec.12607
- Human Microbiome Project C (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234
- Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., and Polz, M. F. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320, 1081–1085. doi: 10.1126/science.1157890
- Kashan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344, 416–420. doi: 10.1126/science.1248575
- Kembel, S. W., Wu, M., Eisen, J. A., and Green, J. L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput. Biol.* 8:e1002743. doi: 10.1371/journal.pcbi.1002743
- Kesnerova, L., Moritz, R., and Engel, P. (2016). *Bartonella apis* sp. nov., a honey bee gut symbiont of the class Alphaproteobacteria. *Int. J. Syst. Evol. Microbiol.* 66, 414–421. doi: 10.1099/ijsem.0.000736
- Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64(Pt. 2), 346–351. doi: 10.1099/ijms.0.059774-0
- Korem, T., Zeevi, D., Suez, J., Weinberger, A., Avnit-Sagi, T., Pompan-Lotan, M., et al. (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* 349, 1101–1106. doi: 10.1126/science.aac4812
- Koskella, B., and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* 38, 916–931. doi: 10.1111/1574-6976.12072
- Koskella, B., and Vos, M. (2015). Adaptation in natural microbial populations. *Annu. Rev. Ecol. Syst.* 46, 503–522. doi: 10.1146/annurev-ecolsys-112414-054458
- Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., et al. (2012). Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* 13, 47–58. doi: 10.1038/nrg3129
- Kuleshov, V., Jiang, C., Zhou, W., Jahanbani, F., Batzoglou, S., and Snyder, M. (2016). Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome. *Nat. Biotechnol.* 34, 64–69. doi: 10.1038/nbt.3416
- Kunin, V., Engelbrekton, A., Ochman, H., and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118–123. doi: 10.1111/j.1462-2920.2009.02051.x
- Kwong, W. K., Engel, P., Koch, H., and Moran, N. A. (2014a). Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11509–11514. doi: 10.1073/pnas.1405838111
- Kwong, W. K., Mancenido, A. L., and Moran, N. A. (2014b). Genome sequences of *Lactobacillus* sp. strains wkB8 and wkB10, members of the Firm-5 Clade, from honey bee guts. *Genome Announc.* 2, e01176-14. doi: 10.1128/genomeA.01176-14
- Kwong, W. K., and Moran, N. A. (2013). Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of *Snodgrassella alvi* gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and *Gilliamella apicola* gen. nov., sp. nov., a member of *Orbaceae* fam. nov., *Orbales* ord. nov., a sister taxon to the order 'Enterobacteriales' of the Gammaproteobacteria. *Int. J. Syst. Evol. Microbiol.* 63(Pt. 6), 2008–2018. doi: 10.1099/ijms.0.044875-0
- Kwong, W. K., and Moran, N. A. (2016). Gut microbial communities of social bees. *Nat. Rev. Microbiol.* 14, 374–384. doi: 10.1038/nrmicro.2016.43
- Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15, 141–161. doi: 10.1007/s10142-015-0433-4
- Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676
- Lau, J. T., Whelan, F. J., Herath, I., Lee, C. H., Collins, S. M., Bercik, P., et al. (2016). Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med.* 8:72. doi: 10.1186/s13073-016-0327-7
- Lee, S., Cantarel, B., Henrissat, B., Gevers, D., Birren, B. W., Huttenhower, C., et al. (2014). Gene-targeted metagenomic analysis of glucan-branching enzyme gene profiles among human and animal fecal microbiota. *ISME J.* 8, 493–503. doi: 10.1038/ismej.2013.167
- Lee, W. J., and Hase, K. (2014). Gut microbiota-generated metabolites in animal health and disease. *Nat. Chem. Biol.* 10, 416–424. doi: 10.1038/nchembio.1535
- Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., et al. (2008a). Evolution of mammals and their gut microbes. *Science* 320, 1647–1651. doi: 10.1126/science.1155725
- Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., and Gordon, J. I. (2008b). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* 6, 776–788. doi: 10.1038/nrmicro1978
- Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848. doi: 10.1016/j.cell.2006.02.017
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Li, H., Limenitakis, J. P., Fuhrer, T., Geuking, M. B., Lawson, M. A., Wyss, M., et al. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* 6:8292. doi: 10.1038/ncomms9292
- Li, S. S., Zhu, A., Benes, V., Costea, P. I., Hercog, R., Hildebrand, F., et al. (2016). Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science* 352, 586–589. doi: 10.1126/science.aad8852
- Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444
- Luo, C., Knight, R., Siljander, H., Knip, M., Xavier, R. J., and Gevers, D. (2015). ConStrains identifies microbial strains in metagenomic datasets. *Nat. Biotechnol.* 33, 1045–1052. doi: 10.1038/nbt.3319
- Majewski, J. (2001). Sexual isolation in bacteria. *FEMS Microbiol. Lett.* 199, 161–169. doi: 10.1111/j.1574-6968.2001.tb10668.x
- Martens, E. C., Kelly, A. G., Tauzin, A. S., and Brumer, H. (2014). The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J. Mol. Biol.* 426, 3851–3865. doi: 10.1016/j.jmb.2014.06.022
- Martinson, V. G., Danforth, B. N., Minckley, R. L., Rueppell, O., Tingek, S., and Moran, N. A. (2011). A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol. Ecol.* 20, 619–628. doi: 10.1111/j.1365-294X.2010.04959.x
- Martinson, V. G., Moy, J., and Moran, N. A. (2012). Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl. Environ. Microbiol.* 78, 2830–2840. doi: 10.1128/AEM.07810-11
- McCoy, R. C., Taylor, R. W., Blauwkamp, T. A., Kelley, J. L., Kertesz, M., Pushkarev, D., et al. (2014). Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS ONE* 9:e106689. doi: 10.1371/journal.pone.0106689
- McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H. V., Domazet-Loso, T., Douglas, A. E., et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3229–3236. doi: 10.1073/pnas.1218525110
- McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L., and Foster, K. R. (2016). Host selection of microbiota via differential adhesion. *Cell Host Microbe* 19, 550–559. doi: 10.1016/j.chom.2016.02.021
- Milani, C., Lugli, G. A., Duranti, S., Turrone, F., Bottacini, F., Mangifesta, M., et al. (2014). Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Appl. Environ. Microbiol.* 80, 6290–6302. doi: 10.1128/AEM.02308-14
- Mitri, S., and Foster, K. R. (2013). The genotypic view of social interactions in microbial communities. *Annu. Rev. Genet.* 47, 247–273. doi: 10.1146/annurev-genet-111212-133307

- Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., et al. (2016). Cospeciation of gut microbiota with hominids. *Science* 353, 380–382. doi: 10.1126/science.aaf3951
- Moodley, Y., and Linz, B. (2009). *Helicobacter pylori* sequences reflect past human migrations. *Genome Dyn.* 6, 62–74. doi: 10.1159/000235763
- Moran, N. A., Hansen, A. K., Powell, J. E., and Sabree, Z. L. (2012). Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS ONE* 7:e36393. doi: 10.1371/journal.pone.0036393
- Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13:R79. doi: 10.1186/gb-2012-13-9-r79
- Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3, e00036-12. doi: 10.1128/mBio.00036-12
- Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., Gonzalez, A., Fontana, L., et al. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332, 970–974. doi: 10.1126/science.1198719
- Nava, G. M., Friedrichsen, H. J., and Stappenbeck, T. S. (2011). Spatial organization of intestinal microbiota in the mouse ascending colon. *ISME J.* 5, 627–638. doi: 10.1038/ismej.2010.161
- Nayfach, S., Fischbach, M. A., and Pollard, K. S. (2015). MetaQuery: a web server for rapid annotation and quantitative analysis of specific genes in the human gut microbiome. *Bioinformatics* 31, 3368–3370. doi: 10.1093/bioinformatics/btv382
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. (2016). *MetaSPADES: A New Versatile de Novo Metagenomics Assembler*. *Arxiv e-Prints*. Ithaca, NY: Cornell University Library.
- Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500
- Ogilvie, L. A., and Jones, B. V. (2015). The human gut virome: a multifaceted majority. *Front. Microbiol.* 6:918. doi: 10.3389/fmicb.2015.00918
- Oh, J., Byrd, A. L., Park, M., Program, N. C. S., Kong, H. H., and Segre, J. A. (2016). Temporal stability of the human skin microbiome. *Cell* 165, 854–866. doi: 10.1016/j.cell.2016.04.008
- Olofsson, T. C., Alsterfjord, M., Nilson, B., Butler, E., and Vasquez, 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. doi: 10.1099/ijs.0.059600-0
- Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L. F., et al. (2014). Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* 8, 953–962. doi: 10.1038/ismej.2013.211
- Pedron, T., Mulet, C., Dauga, C., Frangeul, L., Chervaux, C., Grompone, G., et al. (2012). A crypt-specific core microbiota resides in the mouse colon. *MBio* 3:e00116-12. doi: 10.1128/mBio.00116-12
- Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Planer, J. D., Peng, Y., Kau, A. L., Blanton, L. V., Ndao, I. M., Tarr, P. I., et al. (2016). Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* 534, 263–266. doi: 10.1038/nature17940
- Powell, J. E., Martinson, V. G., Urban-Mead, K., and Moran, N. A. (2014). Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera*. *Appl. Environ. Microbiol.* 80, 7378–7387. doi: 10.1128/AEM.01861-14
- Powell, J. E., Ratnayake, N., and Moran, N. A. (2016). Strain diversity and host specificity in a specialized gut symbiont of honey bees and bumble bees. *Mol. Ecol.* doi: 10.1111/mec.13787 [Epub ahead of print].
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi: 10.1038/nature11450
- Raghavan, V., and Groisman, E. A. (2015). Species-specific dynamic responses of gut bacteria to a mammalian glycan. *J. Bacteriol.* 197, 1538–1548. doi: 10.1128/JB.00010-15
- Rakoff-Nahoum, S., Coyne, M. J., and Comstock, L. E. (2014). An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* 24, 40–49. doi: 10.1016/j.cub.2013.10.077
- Rakoff-Nahoum, S., Foster, K. R., and Comstock, L. E. (2016). The evolution of cooperation within the gut microbiota. *Nature* 533, 255–259. doi: 10.1038/nature17626
- Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Heath, A. C., Rohwer, F., et al. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338. doi: 10.1038/nature09199
- Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., et al. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042–1047. doi: 10.1038/nature01947
- Rodriguez-Valera, F., Martin-Cuadrado, A. B., Rodriguez-Brito, B., Pasic, L., Thingstad, T. F., Rohwer, F., et al. (2009). Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* 7, 828–836. doi: 10.1038/nrmicro2235
- Rooks, M. G., and Garrett, W. S. (2016). Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* 16, 341–352. doi: 10.1038/nri.2016.42
- Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323. doi: 10.1038/nri2515
- Sabree, Z. L., Hansen, A. K., and Moran, N. A. (2012). Independent studies using deep sequencing resolve the same set of core bacterial species dominating gut communities of honey bees. *PLoS ONE* 7:e41250. doi: 10.1371/journal.pone.0041250
- Sanders, J. G., Beichman, A. C., Roman, J., Scott, J. J., Emerson, D., McCarthy, J. J., et al. (2015). Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat. Commun.* 6:8285. doi: 10.1038/ncomms9285
- Sapp, J., and Fox, G. E. (2013). The singular quest for a universal tree of life. *Microbiol. Mol. Biol. Rev.* 77, 541–550. doi: 10.1128/MMBR.00038-13
- Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., et al. (2013). Genomic variation landscape of the human gut microbiome. *Nature* 493, 45–50. doi: 10.1038/nature11711
- Schluter, J., and Foster, K. R. (2012). The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biol.* 10:e1001424. doi: 10.1371/journal.pbio.1001424
- Schmidt, T. S., Matias Rodrigues, J. F., and von Mering, C. (2014). Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. *PLoS Comput. Biol.* 10:e1003594. doi: 10.1371/journal.pcbi.1003594
- Scholz, M., Ward, D. V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., et al. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat. Methods* 13, 435–438. doi: 10.1038/nmeth.3802
- Seedorf, H., Griffin, N. W., Ridaura, V. K., Reyes, A., Cheng, J. Y., Rey, F. E., et al. (2014). Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* 159, 253–266. doi: 10.1016/j.cell.2014.09.008
- Seeley, T. D. (1985). *Honeybee Ecology*. Princeton, NJ: Princeton University Press.
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066
- Sekirov, I., Russell, S. L., Antunes, L. C., and Finlay, B. B. (2010). Gut microbiota in health and disease. *Physiol. Rev.* 90, 859–904. doi: 10.1152/physrev.00045.2009
- Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabo, G., et al. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51. doi: 10.1126/science.1218198
- Sharon, I., Kertesz, M., Hug, L. A., Pushkarev, D., Blauwkamp, T. A., Castelle, C. J., et al. (2015). Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res.* 25, 534–543. doi: 10.1101/gr.183012.114
- Sharon, I., Morowitz, M. J., Thomas, B. C., Costello, E. K., Relman, D. A., and Banfield, J. F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res.* 23, 111–120. doi: 10.1101/gr.142315.112
- Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244. doi: 10.1038/nature10571
- Stackebrandt, E., and Goebel, B. M. (1994). A place for DNA-DNA reassociation and 16S ribosomal-RNA sequence-analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44, 846–849. doi: 10.1099/00207713-44-4-846

- Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D., et al. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl. Environ. Microbiol.* 65, 4799–4807.
- Sullam, K. E., Essinger, S. D., Lozupone, C. A., O'Connor, M. P., Rosen, G. L., Knight, R., et al. (2012). Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol. Ecol.* 21, 3363–3378. doi: 10.1111/j.1365-294X.2012.05552.x
- Takeuchi, N., Cordero, O. X., Koonin, E. V., and Kaneko, K. (2015). Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection. *BMC Biol.* 13:20. doi: 10.1186/s12915-015-0131-7
- Tannock, G. W., Wilson, C. M., Loach, D., Cook, G. M., Eason, J., O'Toole, P. W., et al. (2012). Resource partitioning in relation to cohabitation of *Lactobacillus* species in the mouse forestomach. *ISME J.* 6, 927–938. doi: 10.1038/ismej.2011.161
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102
- Thingstad, T. F., Vage, S., Storesund, J. E., Sandaa, R. A., and Giske, J. (2014). A theoretical analysis of how strain-specific viruses can control microbial species diversity. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7813–7818. doi: 10.1073/pnas.1400909111
- Treangen, T. J., and Rocha, E. P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7:e1001284. doi: 10.1371/journal.pgen.1001284
- Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903. doi: 10.1038/nmeth.3589
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi: 10.1126/science.1093857
- Waite, D. W., and Taylor, M. W. (2014). Characterizing the avian gut microbiota: membership, driving influences, and potential function. *Front. Microbiol.* 5:223. doi: 10.3389/fmicb.2014.00223
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565. doi: 10.1038/nature06269
- Wayne, L. G. (1988). International committee on systematic bacteriology: announcement of the report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Zentralbl. Bakteriol. Mikrobiol. Hyg. A* 268, 433–434.
- Wilson, K. H., and Blitchington, R. B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Appl. Environ. Microbiol.* 62, 2273–2278.
- Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., et al. (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.* 5:1574–1586. doi: 10.1371/journal.pbio.0050156
- Yachi, S., and Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1463–1468. doi: 10.1073/pnas.96.4.1463
- Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6449–6454. doi: 10.1073/pnas.1421834112
- Zhu, A., Sunagawa, S., Mende, D. R., and Bork, P. (2015). Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol.* 16:82. doi: 10.1186/s13059-015-0646-9

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ellegaard and Engel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.