Population Genetics of Host-Associated Microbiomes

Article *in* Current Molecular Biology Reports · September 2019

DOI: 10.1007/s40610-019-00122-y

2 authors:

Louis-Marie Bobay
University of North Carolina at Greensboro
53 PUBLICATIONS **1,052** CITATIONS

SEE PROFILE

Kasie Raymann
University of North Carolina at Greensboro
63 PUBLICATIONS **1,353** CITATIONS

SEE PROFILE

Current Molecular Biology Reports
https://doi.org/10.1007/s40610-019-00122-y

POPULATION GENETICS (E LEWALLEN AND C BONIN, SECTION EDITORS)

Population Genetics of Host-Associated Microbiomes

Louis-Marie Bobay[1] · Kasie Raymann[1]

Abstract

Purpose of Review Host-associated microbiomes can play key roles in the health of animals and plants, but fundamental aspects of the dynamics and evolution of microbial communities are not fully understood.

Recent Findings Several recent studies have sequenced and analyzed the entire diversity of microbial species and strains in different host-associated microbiomes. These studies analyze the population genetics of host-associated microbes, yet many questions remain unanswered.

Summary In this review, we describe the key insights that have been gained by recent microbiome population genetics studies and how they have contributed to our understanding of the fundamental mechanisms that alter the population dynamics of entire microbial communities. We further discuss the technical limitations of current approaches and how new methods and model systems would allow for better genetic characterization of host-associated microbiome populations.

Keywords Microbiome · Population genetics · Metagenomics · Prokaryotes · Evolution · Host-associated microbes

Introduction

The study of microbial evolution and ecology has been rapidly changing in recent years. The continuous improvement of molecular and sequencing techniques is providing new tools to address fundamental questions regarding the evolution and population biology of prokaryotes. In less than two decades, techniques evolved from gel electrophoresis of several proteins of single microbial isolates to metagenomic sequencing of natural communities. These technological advancements have greatly enhanced our understanding of microbial populations. In particular, it is now possible to analyze the population dynamics of whole communities allowing even uncultivable species to be analyzed in their natural environment. Despite these improvements, many fundamental questions regarding the dynamics of microbial populations remain to be answered. Thus far, the application of metagenomic approaches has been heavily focused on host-associated communities, because these microorganisms can impact the health and biology of their hosts, including humans. Many studies have focused on understanding the potential influence of these microbes on their hosts with only a limited number of studies concentrating on the dynamics of the microbes themselves. However, understanding host-associated microbiomes will reveal fundamental aspects of the population dynamics of microbial species. In this review, we summarize and discuss recent findings on the population genetics of host-associated microbiomes.

Population Genetics and Prokaryotes

Population genetics has a long history and represents the theoretical framework upon which most evolutionary models are based. However, population genetics paradigms are rooted in theories that have been developed for sexual organisms, particularly animals and plants. Therefore, the direct application of population genetics models to prokaryotes is not always straightforward.

One common assumption of population genetics models is that individuals undergo sexual reproduction through meiosis at each generation. Therefore, modeling the rates of homologous recombination is comparatively easy for sexual

This article is part of the Topical Collection on *Population Genetics*

✉ Kasie Raymann
ktrayman@uncg.edu

Louis-Marie Bobay
ljbobay@uncg.edu

[1] Biology Department, University of North Carolina at Greensboro, 321 McIver St. 321 Eberhart Building, Greensboro, NC 27412, USA

Published online: 17 July 2019

organisms. In contrast, understanding the impact of occasional recombination on prokaryote evolution has proven difficult [1, 2]. Prokaryotes reproduce by binary fission, a process that does not involve the exchange of genetic material, and were originally thought to evolve clonally with the occasional introduction of mutations [1]. In recent years, there has been growing evidence that many prokaryote species frequently engage in homologous recombination [3, 4]. However, the rates at which different prokaryotes engage in recombination vary and most species appear neither strictly clonal nor strictly sexual [3]. The lack of a generation-dependent process, like meiosis, makes recombination more difficult to calibrate in prokaryotes. Indeed, recombination rates vary over time due to population structure, selection, and contingency. Therefore, the overall impact of recombination on prokaryote evolution is unclear.

Perhaps the most important limitation to the application of population genetics theory to prokaryotes is the assumption that neutral evolution represents a null hypothesis of evolution. Microbial organisms can reach gigantic populations and large effective population sizes (*Ne*) enhance the effectiveness of selection [5, 6]. For these reasons, it has been questioned whether neutral or nearly neutral models of evolution are pertinent to describe microbial evolution [7, 8]. Multiple lines of evidence suggest that synonymous codon positions are substantially affected by selection due to strong codon usage preference and nucleotide composition (i.e., GC content) [9]. One study has shown that adaptive evolution—rather than drift—is responsible for the fixation of at least 50% of amino acid substitutions since the divergence of two closely related species: *Escherichia coli* and *Salmonella enterica* [10]. These results and other observations have motivated researchers to question whether neutrality constitutes a reasonable assumption for microbes, even at seemingly non-functional positions [8]. For instance, demography inference, which is based on neutral models of evolution, has been shown to be systematically biased in bacteria [8]. In summary, the development of more solid population genetics foundations is needed for the appropriate investigation of microbial organisms and populations.

The Prokaryotic Species Problem

One obvious problem when analyzing populations is the definition of what constitutes the population under study. Species definitions aim to delimit cohesive populations that share a common history and—to some extent—common traits, but defining species is a challenging and highly debated task in prokaryotes [4]. Again, this is a lesser concern in sexual organisms where species are usually defined as groups of randomly mating individuals. Approaches used to define species in prokaryotes rely on implicit assumptions of their population dynamics. For instance, the stable ecotype model (SEM) assumes that each prokaryotic species represents an ecological species adapted to a specific niche [11]. Under this model, the genetic cohesion of the species is maintained by periodic selection: a new variant acquires a beneficial mutation and sweeps through the population (through "genome sweep"). This process would result in strong recurrent bottlenecks, where the loss of most variants would maintain the genetic cohesion of the species. The SEM suffers several shortcomings: (i) it does not easily account for the high variability in gene content observed in most prokaryotic species (see "Assessing strain diversity" below) and (ii) the theoretical framework of the SEM is best suited for clonal populations, but recent works have pointed out the central role of gene flow for an increasing number of prokaryotic species [2, 3]. The SEM constitutes an ecology-centered view of species, which is extremely difficult—if even possible—to characterize. In fact, each species might occupy a specific niche, where strains and other subpopulations could occupy sub-niches [12]. Currently, our lack of knowledge regarding the ecology of prokaryotes precludes the development of a solid ecology-centered species definition for prokaryotes.

Several studies suggest that gene flow could constitute a force that maintains genetic cohesion in prokaryotic species, creating the possibility to define microbial species in a similar way as used for sexual organisms, i.e., the biological species concept (BSC) [2, 13–16]. The BSC not only offers a theoretical framework to classify microbial species but also provides concrete metrics to assess species borders from genomic data. Although the BSC cannot be applied to strictly clonal species, recent evidence suggests that only a minority of prokaryotes are truly clonal [2]. Despite these advantages, a BSC-anchored definition of prokaryotic species remains challenging, since prokaryotes exhibit various degrees of gene flow [3]. Also, the extent to which microbial species remain "sexually" isolated from one another remains to be precisely identified. Nevertheless, the accumulation of recent results indicates growing support for a key role of gene flow in maintaining cohesive populations in prokaryotes.

In order to circumvent the species definition issue, it has become standard practice to delimit species borders based on sequence thresholds. Two strains are usually considered members of the same species if their genomes differ by less than 5% sequence identity [17]. This convenient threshold is based on empirical observations rather than theory and offers a way to define prokaryotic species under a unique and simple framework. However, it should be stressed that this practice potentially introduces important biases when studying population dynamics and evolution. Some have argued that uneven evolutionary rates across prokaryote lineages do not seem compatible with the application of a single sequence identity threshold [18]. Defining species borders without a foundation of population genetics is problematic when studying prokaryotic organisms.

Prokaryotic Population Structure

Analyzing the spatial and temporal dynamics of prokaryotes is notoriously difficult. One common limitation is the potential presence of strains that remain hidden at low—or very low frequency—in a given area, thereby remaining undetected by most methods. It is nearly impossible to prove that a given strain is absent from a geographic location due to sampling and sequencing limitations. As a result, it is challenging to determine whether the strains of a population have been acquired by migration or whether they result from the increase in frequency of a cryptic strain. Additionally, it is difficult to delimit strict geographical areas and niches for microbes and the homogeneity of defined areas can always be questioned [19]. Microbes are often associated with particles; therefore, even a simple puddle of water could represent a complex environment structured into thousands of niches. It is possible to sample a habitat at different depths and at different locations, but from a microbial perspective, sampling a bacterial strain in a puddle of water could be equivalent to sampling a single fish from the entire ocean. Despite the vastness of geographical areas on Earth, several studies reported that microbes can be highly cohesive over global ranges [20–22], suggesting that geography might have limited impact on population structure. Nevertheless, modeling studies have suggested that geographic distribution can be an important force driving genetic structure in bacteria [23], especially for species engaging in recombination. Conversely, populations living in the same location might remain physically isolated due to the presence of sub-habitats, a process named *mosaic sympatry* [4, 19]. Overall, these results give a paradoxical picture of population structure in microbes, where populations separated by thousands of kilometers can appear surprisingly similar, but other population members found in the exact same sample can be very diverse. Understanding the evolutionary and ecological forces responsible for these patterns remains challenging, but new methodologies and new study systems can shed new lights into these questions.

Methods for Evaluating Prokaryotic Strain Diversity in Microbiomes

Although the definition of prokaryotic species is questionable due to the abovementioned reasons, they appear much more diverse than eukaryotes. Sexual organisms typically differ by <2% nucleotide divergence [24], whereas prokaryotes often display higher levels of divergence [25]. However, the most dramatic variations in prokaryote strain diversity do not come from sequence divergence but from variations in gene content. Indeed, different strains comprising the same species typically only share 50–60% of their genes with other members of the species (i.e., the core genome) [25]. This implies that different strains contain diverse sets of genes, and while many of these genes might actually correspond to mobile elements, a substantial number of accessory genes appears involved in various functions that range from metabolic enzymes to warfare systems mediating strain competition [26–28]. These observations highlight that different strains belonging to the same species can display diverse metabolisms; therefore, it is important to characterize strain diversity in order to understand prokaryotic population dynamics and their impact on their host or their environment.

The methods and tools used to characterize prokaryotic populations have been rapidly evolving. Originally, strains were characterized with molecular markers (multilocus enzyme electrophoresis (MLEE) and multilocus strain typing (MLST)), which consists of characterizing the sequence of a handful of protein coding in individual isolates [29, 30]. Profiles or sequences are then used to classify strains into sequence types (STs). These approaches allow for the characterization of strains based on several housekeeping genes that are part of the core genome but do not provide information regarding the diversity of accessory genes that can vary drastically across strains. Since the cost of sequencing has decreased, these techniques are less frequently used and whole genome sequencing (WGS) has started to replace these practices. WGS provides a more accurate description of strain diversity and this approach is now used routinely to characterize entire genomes. The only limitation of WGS is that it requires isolation and cultivation of each strain. Cultivation can be circumvented by single-cell sequencing, but this method also has limitations: single-cell sequencing often results in contaminated and incomplete genome assemblies [31].

Metagenomic approaches allow for the sequencing of entire communities in their natural environment. This opens the possibility to assess populations of prokaryotic species that cannot be cultivated in the laboratory. The most popular application of metagenomics consists of 16S rRNA amplicon sequencing. Because this gene is universal across prokaryotes and has a slow evolution rate, a single pair of primers can be used to assess virtually all prokaryotic diversity in a sample. Unfortunately, the slow rate of evolution of 16S rRNA prevents analysis of strain-level diversity within natural populations. As a result, this method has provided little insight on the population genetics of prokaryotes. Shotgun metagenomics, on the other hand, consists of sequencing all the genomes in a given sample. This approach is more expensive because it requires deep coverage to capture all genomes but has the potential to capture strain diversity in a community. Recently, much effort has been dedicated to the development of algorithms that can reconstruct strain genotypes from the large bulk of reads produced by metagenomic sequencing [32•, 33, 34, 35••, 36, 37, 33–]. Specificities of these tools are detailed in Table 1. Most methods require a database of reference genomes to identify variants, which are typically

assigned to the same genotype when found at the same frequency. As a result, most of current methods perform better when analyzing a large number of samples with high genome coverage. The recent development of algorithms capable of reconstructing entire strain genotypes from metagenomic data has led to key insights regarding the population dynamic of host-associated microbiomes.

Population Genetics of the Human Gut Microbiome

Many population genetics studies have concentrated on gastric bacterial pathogens of humans, such as *Clostridium difficile* [38–41], *Vibrio cholerae* [42], *Helicobacter pylori* [43–46], and *E. coli* [47–53]. In terms of population genetics of commensal organisms, in addition to the development of the methods cited above (Table 1), other recent large-scale studies have been conducted with the specific goal of studying the population genetics of the human gut microbiome [54••, 55••, 56••, 57••, 58•, 59, 60•, 61–64]. Consistently, human gut population genetics studies have shown that individual hosts are more similar in microbial species and strain compositions over time than host-to-host comparisons [35, 54••, 55••, 58•, 62–67], suggesting that microbial populations persist within hosts for at least one year (most longitudinal studies are limited to one year). Despite the recent increase in strain-level analyses of human gut microbes, studies examining population genetics of microbiomes are limited.

One of the most interesting recent findings is the observation of within-host adaptation and diversification (i.e., microevolution) in some gut microbiome species [54••, 55••, 56••]. Garud et al. [55••] followed the evolution of 40 prevalent species of the human gut microbiome using metagenomics and found evidence that gut bacteria can evolve over short timescales and as well as evidence for microbiome-wide signatures of adaptive evolution. Unique to this study, the authors investigated the role of recombination in the evolution of gut bacterial strains and demonstrated that recombination and mutations are common drivers of population dynamics over short timescales, but replacement events dominate on multidecade time-scales [55••]. One major limitation of this study is that a large amount of data was discarded in an attempt to reduce false positives: a common practice to ensure that the inferred variants do not result from sequencing errors. Discarding bulk data prevents the analysis of low-frequency variants and hinders the ability to understand evolutionary events occurring within the population.

Consistent with Garud et al. [55••], Zhao et al. [54••] found evidence for within-host adaptation of *Bacteroides fragilis* in the human gut microbiome using isolate sequencing and metagenomics. This large-scale and in-depth study consisted of culturing and sequencing 602 *B. fragilis* isolates from 30 fecal samples and metagenomic sequencing of 319 fecal samples (longitudinal samples from seven individuals spanning two years and single samples for five subjects). In addition, the authors complemented their analysis by searching for adaptive mutations in *B. fragilis* in over 1000 publicly available metagenomes from multiple geographic locations. Aside from the finding of rapid evolution of *B. fragilis* within hosts, the authors found evidence for parallel evolution in sixteen genes, several of which are involved in polysaccharide utilization and cell-envelope biosynthesis. Interestingly, one adaptive mutation common in their 12 US-derived samples was found to be frequent only in Western microbiomes [54••]. Taken together, these results suggest that *B. fragilis* populations diversify within an individual host to form co-existing sublineages that acquire beneficial de novo mutations in the absence of obvious selective pressure (e.g., antibiotics) [54••]. This groundbreaking study was limited to a single species, so it is unclear if these results can be generalized to all gut commensals. However, the metagenomic study of Garud et al. [55••] suggests that this phenomenon of rapid adaptive evolution might occur in many other gut species.

Another study found evidence for within-host microevolution of commensal *E. coli* [68••]. Ghalayini et al. [68••] investigated the evolution of *E. coli* in the human gut by sequencing 24 isolates sampled at three different time points within a single individual over a year. It was predicted that the annual mutation rate of *E. coli* strain ED1a is 6.90×10^{-7} [68••], which is three times higher than what has been found in vitro during long-term evolution experiments (LTEEs) [69]. Despite rapid evolution of *E. coli* within a single host, the authors did not find evidence for selection, suggesting that *E. coli* ED1a evolution is neutral in the gut [68••]. Because *E. coli* is predicted to have a small effective population within human gut microbiomes, genetic drift could have a stronger influence on the evolution of this species [68••]. This finding indicates that the population dynamics of different members of the human gut microbiome can vary greatly from one species to another.

Very recently, Poyet et al. isolated and sequenced a very large collection of >3500 strains from the human gut microbiome [56]. Their study demonstrated that many more gut bacteria can be cultivated in lab conditions than previously thought. Using longitudinal sampling of the same individuals, they were able to retrieve more accurate information concerning the composition of gut microbial communities and how they evolve within individuals [56]. Some species appeared to be same stable over time and composed of the same strain(s), while others experienced strain turnover. Interestingly, they found multiple drivers of microbiome composition: amino acids levels seemed to the be the main factor driving within host variation, whereas differences among individuals appeared primarily driven by bile acid levels [56].

Table 1 Overview of recently developed methods for assessing strain diversity in metagenomic datasets

Program	Method overview	Requirements	Datasets analyzed	Major results	Reference
ConStrains	SNP calling in a set of universal genes and strain phylogenetic reconstruction using reference guided method	One sequenced genome (complete or draft), minimum of tenfold coverage for SNP calling	In silico–simulated datasets, *Burkholderia dolosa* in cystic fibrosis patients, and infant gut dataset containing 54 samples from nine subjects collected over the first three years of life	Recovered strain-level diversity from 75 species from the human gut microbiome. Identified strains of *Bifidobacterium longum* containing the human milk oligosaccharide utilization cluster in higher relative abundance in the infant gut microbiome	Luo et al. 2015 Nat Biotechnol.
WG-FAST	SNP calling by alignment to reference genome and placement of samples in a phylogenetic context	Reference species with high quality genomic data (e.g. 50× reference genome coverage)	53 datasets from a metagenomic survey of stool samples from the 2011 *E. coli* O104:H4 outbreak	Confirmed samples as positive for O104:H4 as well as identifying additional samples as positive for O104:H4 that were previously reported as negative	Sahl et al. 2015 Genome Med.
MetaPhlAn2	SNP calling in species-specific markers and strain clustering based on presence absence of markers	Strain-level markers created from reference genomes. Strain identification requires at least 70% of the markers for a strain are present in the sample	24 synthetic metagenomes comprising 656 million reads and 1,295 species, four elbow-skin samples from three subjects, extended HMP samples from human gut	Identified *Bacteroides* strains in the human gut (HMP samples) and found strong patterns of strain retention of *Prevotella copri*, three non-*Bacteroides* species, and the three *Bacteroides* species most abundant in the human gut	Truong et al. 2015 Nat Methods
MIDAS	Quantification of gene content and SNPs using a database of >30,000 bacterial reference genomes clustered into species groups	Strain analysis requires high coverage (> 10×) and the availability of a reference genome	98 stool metagenomes from Swedish mothers and their infants over one year, 98 globally distributed marine metagenomes	Followed mother-to-infant strain transmission and identified strains that are vertically transmitted from mother to infant. Captured genome content variability across strains through analysis of metagenomic samples from the Tara Oceans Project	Nayfach et al. 2016 Genome Res.
PanPhlAn	Presence or absence of genes based on the species' pangenome, pangenome-based phylogenomic analysis, and transcriptional activity of strains	Reference genomes required to build pangenomes	Synthetic data sets, unknown bacterial strains in metagenomic data from the 2011 *E. coli* outbreak, 1,830 gut metagenomes from eight large-scale studies	Identified outbreak-positive samples (when pathogenic *E. coli* genomes were excluded from the reference genome database) and analyzed the population structure of two prevalent gut species *Eubacterium rectale* (only one reference genome) and *Akkermansia muciniphila* (only two reference genome)	Scholz et al. 2016 Nat Methods
metaMLST	Reconstructs MLST loci of microbes present in metagenomic data to determine sequence type (ST) profiles	Database of MLST loci for species of interest. Limited to tracking the most abundant strain	Synthetic and semi-synthetic metagenomes, 531 gut microbiome samples, 473 skin metagenomic samples, metagenomic samples extracted from a 5300 year-old mummy	Tracked *E. coli* strain ST-678 (2011 outbreak in Germany), identified 79 ST of *S. epidermidis* from 100 skin metagenomes and confirmed body type specificity of this species, and identified the presence of a new ST of *H. pylori* in mummy stomach metagenomes	Zolfo et al. 2017 Nucleic Acids Res.
StrainPhlAn	Reconstruction of consensus sequence variants within species-specific marker genes and estimation of strain-level phylogenies	At least one reference genome	1500 gut metagenomes drawn from populations spanning North and South American, European, Asian, and African countries	Phylogenetically profiled thousands of strains from 125 undercharacterized intestinal species, identified that one strain generally dominates each species in the human gut and is retained over time, and found strains from	Truong et al. 2017 Genome Res.

Table 1 (continued)

Program	Method overview	Requirements	Datasets analyzed	Major results	Reference
				the same species in different individuals are genetically distinct	
Strain Finder	Strain genotyping, SNP frequency, and strain phylogenetic reconstruction	At least one reference genome and high coverage	In silico datasets, 79 stool samples from donors and recipients following fecal microbiota transplantation (FMT)	Tracked FMT donors and recipients and identified that bacterial abundance and phylogeny dictate engraftment and either no strains or complete sets of strains colonize the patient	Smillie et al. 2018. Cell Host Microbe
mOTUs2	Nucleotide variation profiles (centroid clustering) based on single copy phylogenetic marker genes from >3100 metagenomic samples	Universal phylogenetic marker gene-based operational taxonomic units (mOTUs) generated from metagenomic datasets	2807 human microbiome samples and 139 prokaryote-enriched metagenomes from the Tara Oceans Project	Compared strain population similarities across body site samples from HMP project and found that stool and vaginal samples are the locations that display the most across individual variation	Milanese et al. 2019 Nat Commun

Limitations of Studying Population Genetics in Complex Host-Associated Microbial Communities

Most insights on the population genetics of host-associated microbiomes have been obtained by shotgun metagenomic sequencing followed by strain genotype reconstruction. However, there are several limitations to these approaches. One major issue is the ability to resolve (or "phase") the different reads in order to ascribe alleles and accessory genes to each strain genotype of the community. Moreover, these analyses are computationally expensive and require high coverage to robustly infer genotypes. As a result, rare variants of the community are overlooked, and, unless the coverage is extremely deep, only a picture of the dominant members of the is obtained. Let us assume a simple hypothetical community composed of two species A and B found at a 1:100 ratio, and within each of these hypothetical species, multiple strains co-exist, with some present at low frequencies. The presence of a rare variant at 1% frequency in the minor species would require a sequencing depth of 10,000× of this simple community to obtain a coverage of 1× of the rare variant. Characterizing the complete genome of all the strains of each species would therefore require unrealistically deep sequencing to reconstruct the complete genomes of low frequency strains and species, especially for complex communities, like the human and mouse gut microbiomes, that contain hundreds of species. For these reasons, shotgun metagenomic sequencing only provides a picture of the most common strains of the most abundant species in the population and the associated costs usually prevent the analysis of large numbers of samples. These limitations raise major concerns about the results of human, or other complex microbiome, studies. Many cryptic strains are potentially overlooked, and it is possible that the observed gain or transfer of a strain might in fact be the result of the rise in frequency of a cryptic strain. These limitations also apply to the detection of de novo mutations in metagenomic samples. Determining how much cryptic strain diversity remains hidden in host-associated microbiomes and how they contribute to microbiome evolution is an important issue that cannot be easily addressed with shotgun metagenomic approaches.

In order to characterize low-frequency strains in microbial communities, we [70•] and others [71–73] have recently used a more targeted approach, which we will hereafter refer to as metagenomic amplicon strain typing (MAST). This approach combines MLST and metagenomic approaches but differs from other metagenomic MLST-like approaches, such as MetaPhlAn2 [74], metaMLST [34••], StrainPhlAn [35], mOTUs2 [37], in that it does not use shotgun metagenomic data but instead consists of amplicon

sequencing of variable regions in protein-coding genes specific to each species. Protein-coding genes present more variations than the 16S rRNA gene, and this offers a higher resolution to characterize strain diversity. By deep sequencing only a few protein coding gene markers, it is possible to assess strain diversity and capture rare variants (e.g., 50,000× for each species) [70•]. MAST has two obvious limitations: (i) it does not provide any information regarding the overall gene content of each strain; (ii) it requires designing a set of primers for each species, precluding the analysis of complex communities and species without reference genomes. For these reasons, MAST is best suited for following individual species or simple communities. Despite these limitations, MAST provides a cost-effective method for uncovering a more accurate picture of strain dynamics in host-associated microbiomes across many samples and experimental conditions.

Aside from technical issues (methods and cost), there are other problems with using complex communities to study the population genetics of prokaryotes. One key problem is a lack of reference genomes for all members of the community. Despite the recent generation of thousands of reference genomes from the human gut microbiome, approximately 20% of the microbiome can still not be mapped to reference genomes [75]. In some studies, de novo assembly of entire communities has been conducted [76•], but the computational cost of these approaches is rather dissuasive and usually provides incomplete assemblies. Although genomic sequencing can provide a lot of information, there is still a need to cultivate organisms to perform controlled experiments. Over 1000 uncultured organisms were recently identified in metagenomic datasets of the human gut microbiome [77]. Lack of cultivated community members is not only a problem for the human microbiome, but for all complex microbiome systems and models, like the mouse. Mammalian models are also costly to maintain and have long generation times, resulting in a low experimental tractability.

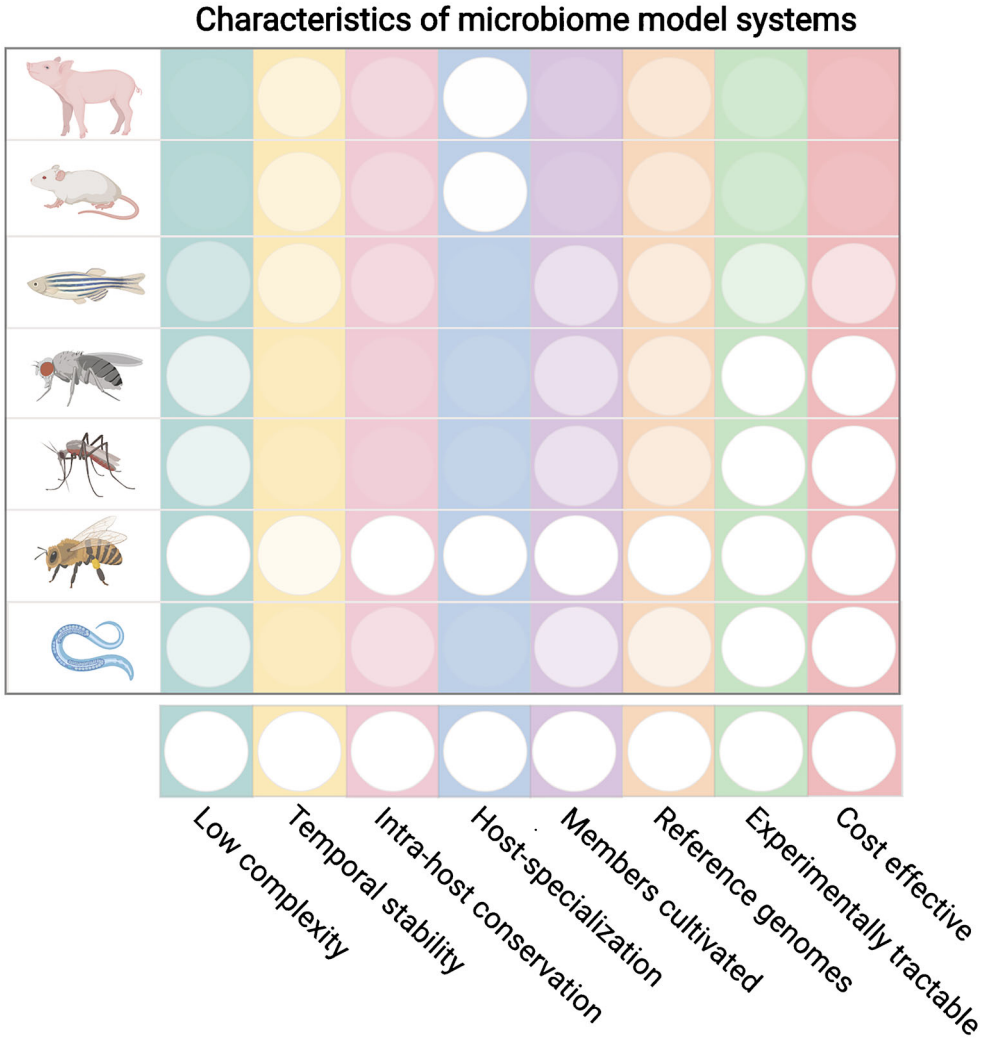

Fig. 1 Comparison of the most commonly used model organisms for microbiome research based on the eight characteristics desirable for population genetics studies. Color gradients represent the degree to which each model system fulfills the listed characteristic, with white representing the characteristic is fully present. Created with BioRender (https://biorender.com/)

Model Systems for Microbial Population Genetics

Here, we argue that there are eight desirable characteristics that contribute to the usefulness of a model system for studying microbiome population genetics: (1) low complexity; (2) temporal stability; (3) inter-host conservation (i.e., a set of species shared across all individuals); (4) host-specialization; (5) the ability to cultivate all members; (6) the existence of reference genomes for all members; (7) the experimental tractability of the host organism (ability to manipulate the host organism, use large sample sizes, and replicate experiments over short time-scales); and (8) low cost associated with maintaining and working with the host organism. Two common mammalian model systems that are used to study the microbiomes are mice and piglets (Fig. 1 [78–81]). It is argued that mammalian models are more accurate for studying the human gut microbiome because of their close relationship and similarities. For example, many of the same prokaryotic taxa are somewhat conserved across mammals [81, 82] and human gut microbes are capable of colonizing mice and piglets [78]. In fact, many human microbiome studies are conducted using axenic mice colonized with human microbes [78, 80, 81]. Although colonizing mice or other mammals with human microbes might be helpful for understanding how the microbiome contributes to human health, the population genetics of prokaryotes cannot accurately be studied when these microbes are not in their natural host environment. In addition, these models also have very complex natural microbiomes that have not been well characterized (reference genomes and cultivated microbes are limited) and, like humans, their microbiome composition can vary over time and across individuals [81]. Mice and piglets are comparatively costly to maintain and manipulate, particularly over the course of longer experiments. To gain a more detailed understanding of the population dynamics of host-associated microbiomes, simpler model systems are needed. Indeed, many fundamental questions regarding the population genetics of host-associated microbiomes could be answered by using model organisms with less complex and more stable communities.

Non-mammalian model systems frequently used to study host-associated microbes include: zebrafish [83], fruit flies [84], mosquitoes [85], honey bees [86•], and nematodes [87] (Fig. 1). All these systems have been utilized due to their tractability and simplicity. However, the natural microbiome of zebrafish [88], mosquitoes [89], and nematodes [90] are complex, with many different species and low conservation across individuals. On the other hand, honey bees and fruit flies have been reported to possess less than 10 species of bacteria in their gut, and their microbiomes have been very well characterized [84, 86•]. Both fruit flies and honey bees are experimentally tractable and cost effective [84, 86•], but flies lack host-specialization, temporal stability, and conservation across individuals [91].

In contrast to fruit files, the honey bee gut microbiome consists of five core species (found in all bees worldwide) that make up approximately 90% of the total population, their microbiome is stable over time and across individuals, the community members are host-specialized (not found in the environment or in other hosts), all members (even the few non-core species) can be cultivated and multiple strains of each species have been sequenced [86•, 92–94]. For these reasons, the honey bee is a particularly well-suited model system for microbiome research [86], especially to study microbiome population genetics [70•]. In fact, we previously showed that using an MAST approach, we could track strain dynamics of two of the most dominant core species following antibiotic exposure [70]. Using this approach, we were able to provide an accurate picture of the fine-scale dynamics of microbial communities over time. In addition, several other studies have looked at strain-level diversity of the bee gut microbiome using shotgun metagenomics [95, 96••, 97••], metabolomics [98•], marker gene sequencing [71], and isolate-based approaches [99–106]. From these studies, an extensive amount of strain-level diversity has been observed within each of the five core bee gut species, which present different functional capabilities [107]. Due to its overall simplicity and stability, the honey bee gut microbiome—as a whole—is perhaps one of the most accurately understood host-associated microbiomes. Thus, future studies on population genetics of microbiomes could greatly benefit from this model system.

Conclusions

Recent advances, mostly coming from shotgun metagenomic sequencing, have contributed to a better understanding of the population dynamics of host-associated microbiomes. These works have been mainly directed toward the human and mouse gut microbiomes. However, using shotgun metagenomic sequencing in complex systems is far from ideal when studying fundamental population processes inherent to microbial communities. We argue that more targeted approaches, like MAST, and simpler model systems, like the honey bee, potentially offer best-suited tools to study host-associated microbiomes in experimental settings.

Acknowledgments We would like to thank Mathieu Groussin for reading and providing feedback on the manuscript.

Funding information This work was supported by the National Science Foundation (DEB-1831730) to L-M.B. and USDA NIFA (2018-02589) to K.R.

Compliance with Ethical Standards

Conflict of Interest Kasie Raymann and Louis-Marie Bobay declare no conflicts of interest.

Human and Animal Rights and Informed Consent This article contains no studies with human and animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? Proc Natl Acad Sci U S A. 1993;90:4384–8. https://doi.org/10.1073/pnas.90.10.4384.
2. Bobay L-M, Ochman H. Biological species are universal across life's domains. Genome Biol Evol. 2017;9:491–501. https://doi.org/10.1093/gbe/evx026.
3. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 2009;3:199–208. https://doi.org/10.1038/ismej.2008.93.
4. Shapiro BJ, Polz MF. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol. 2014;22:235–47. https://doi.org/10.1016/j.tim.2014.02.006.
5. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science. 2014;344:416–20. https://doi.org/10.1126/science.
6. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 2008;89:311–36. https://doi.org/10.1017/S001667230800949X.
7. Rocha EPC, Feil EJ. Mutational patterns cannot explain genome composition: are there any neutral sites in the genomes of bacteria? PLoS Genet. 2010;6:e1001104. https://doi.org/10.1371/journal.pgen.1001104.
8. Lapierre M, Blin C, Lambert A, Achaz G, Rocha EPC. The impact of selection, gene conversion, and biased sampling on the assessment of microbial demography. Mol Biol Evol. 2016;33:1711–25. https://doi.org/10.1093/molbev/msw048.
9. Hershberg R, Petrov DA. General rules for optimal codon choice. PLoS Genet. 2009;5:e1000556. https://doi.org/10.1371/journal.pgen.1000556.
10. Charlesworth J, Eyre-Walker A. The rate of adaptive evolution in enteric bacteria. Mol Biol Evol. 2006;23:1348–56. https://doi.org/10.1093/molbev/msk025.
11. Cohan FM. Bacterial species and speciation. Syst Biol. 2001;50(4):513–24. https://doi.org/10.1080/10635150118398.
12. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev. 2011;35:957–76. https://doi.org/10.1111/j.1574-6976.2011.00292.x.
13. Dykhuizen DE, Green L. Recombination in Escherichia coli and the definition of biological species. J Bacteriol. 1991;173(22):7257–68. https://doi.org/10.1128/jb.173.22.7257-7268.1991.
14. Vulic M, Dionisio F, Taddei F, Radman M. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc Natl Acad Sci U S A. 2002;94(18):9763–7. https://doi.org/10.1073/pnas.94.18.9763.
15. Eppley JM, Tyson GW, Getz WM, Banfield JF. Genetic exchange across a species boundary in the archaeal genus ferroplasma. Genetics. 2007;177:407–16. https://doi.org/10.1534/genetics.107.072892.
16. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabo G, et al. Population genomics of early events in the ecological differentiation of bacteria. Science. 2012;336:48–51. https://doi.org/10.1126/science.1218198.
17. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102:2567–72. https://doi.org/10.1073/pnas.0409727102.
18. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004. https://doi.org/10.1038/nbt.4229.
19. Mallet J. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos Trans R Soc B biol Sci. 2008;363:2971–86. https://doi.org/10.1098/rstb.2008.0081.
20. Papke RT, Zhaxybayeva O, Feil EJ, Sommerfeld K, Muise D, Doolittle WF. Searching for species in haloarchaea. Proc Natl Acad Sci U S A. 2007;104:14092–7. https://doi.org/10.1073/pnas.0706358104.
21. Coleman ML, Chisholm SW. Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci U S A. 2010;107:18634–9. https://doi.org/10.1073/pnas.1009480107.
22. Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Bapteste E, et al. Local mobile gene pools rapidly cross species boundaries to create endemicity within global Vibrio cholerae populations. MBio. 2011;2:e00335–10. https://doi.org/10.1128/mBio.00335-10.
23. Marttinen P, Hanage WP. Speciation trajectories in recombining bacterial species. PLoS Comput Biol. 2017;13:e1005640. https://doi.org/10.1371/journal.pcbi.1005640.
24. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515:261–3. https://doi.org/10.1038/nature13685.
25. Bobay L-M, Ochman H. Factors driving effective population size and pan-genome evolution in bacteria. BMC Evol Biol. 2018;18:153. https://doi.org/10.1186/s12862-018-1272-4.
26. Bobay LM, Rocha EPC, Touchon M. The adaptation of temperate bacteriophages to their host genomes. Mol Biol Evol. 2013;30:737–51. https://doi.org/10.1093/molbev/mss279.
27. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet. 2009;5:e1000344. https://doi.org/10.1371/journal.pgen.1000344.
28. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein secretion systems in bacterial genomes. Sci Rep. 1615;2016:1–21. https://doi.org/10.1007/978-1-4939-7033-9_1.
29. Baptist JN, Shaw CR, Mandel M. Comparative zone electrophoresis of enzymes of Pseudomonas solanacearum and Pseudomonas cepacia. J Bacteriol. 1971;108(2):799–803.
30. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998;95:3140–5. https://doi.org/10.1073/pnas.95.6.3140.
31. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 2016;17:175–88. https://doi.org/10.1038/nrg.2015.16.

32.• Luo C, Knight R, Siljander H, Knip M, Xavier RJ, Gevers D. ConStrains identifies microbial strains in metagenomic datasets. Nat Biotechnol. 2015;33:1045–52. https://doi.org/10.1038/nbt.3319 **This study recovered strain diversity from 75 species of the human gut microbiome and showed that strains of Bifidobacterium longum contain the human milk oligosaccharide utilization cluster and are found in higher relative abundance in the infant gut microbiome**.

33. Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, et al. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat Methods. 2016;13:435–8. https://doi.org/10.1038/nmeth.3802.

34. Zolfo M, Tett A, Jousson O, Donati C, Segata N. MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples. Nucleic Acids Res. 2017;45:e7. https://doi.org/10.1093/nar/gkw837.

35.•• Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure & genetic diversity from metagenomes. Genome Res. 2017;27:626–38. https://doi.org/10.1101/gr.216242.116 **This study characterized thousands of strains for 125 species of the human gut microbiome and showed that a single strain usually dominates each species.** Different individuals rarely present the same strains.

36.•• Smillie CS, Sauk J, Gevers D, Friedman J, Sung J, Youngster I, et al. Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. Cell Host Microbe. 2018;23:229–240.e5. https://doi.org/10.1016/j.chom.2018.01.003 **Analysis of fecal transplants from donors to patients.** Results showed that either no strain or the complete set of strains colonizes patients.

37. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10:1014. https://doi.org/10.1038/s41467-019-08844-4.

38. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and evolution in the genome of Clostridium difficile. Clin Microbiol Rev. 2015;28:721–41. https://doi.org/10.1128/CMR.00127-14.

39. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, et al. Evolutionary history of the Clostridium difficile pathogenicity locus. Genome Biol Evol. 2014;6:36–52. https://doi.org/10.1093/gbe/evt204.

40. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. Nat Genet. 2013;45:109–13. https://doi.org/10.1038/ng.2478.

41. Yan Q, Zhang J, Chen C, Zhou H, Du P, Cui Z, et al. Multilocus sequence typing (MLST) analysis of 104 Clostridium difficile strains isolated from China. Epidemiol Infect. 2013;141:195–9. https://doi.org/10.1017/S0950268812000453.

42. Levade I, Terrat Y, Leducq J-B, Weil AA, Mayo-Smith LM, Chowdhury F, et al. Vibrio cholerae genomic diversity within and between patients. Microb Genomics. 2017;3. https://doi.org/10.1099/mgen.0.000142.

43. Suzuki R, Shiota S, Yamaoka Y. Molecular epidemiology, population genetics, and pathogenic role of Helicobacter pylori. Infect Genet Evol. 2012;12:203–13. https://doi.org/10.1016/j.meegid.2011.12.002.

44. Secka O, Moodley Y, Antonio M, Berg DE, Tapgun M, Walton R, et al. Population genetic analyses of Helicobacter pylori isolates from Gambian adults and children. PLoS One. 2014;9:e109466. https://doi.org/10.1371/journal.pone.0109466.

45. Linz B, Vololonantenainab CRR, Seck A, Carod JF, Dia D, Garin B, et al. Population genetic structure and isolation by distance of Helicobacter pylori in Senegal and Madagascar. PLoS One. 2014;9:e87355. https://doi.org/10.1371/journal.pone.0087355.

46. Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, et al. Worldwide population structure, long-term demography, and local adaptation of Helicobacter pylori. Genetics. 2015;200:947–63. https://doi.org/10.1534/genetics.115.176404.

47. Sahl JW, Schupp JM, Rasko DA, Colman RE, Foster JT, Keim P. Phylogenetically typing bacterial strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data. Genome Med. 2015;7:52. https://doi.org/10.1186/s13073-015-0176-9.

48. Ferdous M, Zhou K, De Boer RF, Friedrich AW, Kooistra-Smid AMD, Rossen JWA. Comprehensive characterization of Escherichia coli O104: H4 isolated from patients in the Netherlands. Front Microbiol. 2015;6:1348. https://doi.org/10.3389/fmicb.2015.01348.

49. Guy L, Jernberg C, Arvén Norling J, Ivarsson S, Hedenström I, Melefors Ö, et al. Adaptive mutations and replacements of virulence traits in the Escherichia coli O104:H4 outbreak population. PLoS One. 2013;8:e63027. https://doi.org/10.1371/journal.pone.0063027.

50. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, FitzGerald M, et al. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A. 2012;109:3065–70. https://doi.org/10.1073/pnas.1121491109.

51. Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, et al. Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. Genome Res. 2015;25:119–28. https://doi.org/10.1101/gr.180190.114.

52. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol. 2015;53:2410–26. https://doi.org/10.1128/JCM.00008-15.

53. Loman NJ, Constantinidou C, Christner M, Chan JZM, Quick J, Weir JC, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4. JAMA - J Am Med Assoc. 2013;309:1502–10. https://doi.org/10.1001/jama.2013.3231.

54.•• Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, et al. Adaptive evolution within gut microbiomes of healthy people. Cell Host Microbe. 2019;25:656–667.e8. https://doi.org/10.1016/j.chom.2019.03.007 **The authors characterized within-host evolution of a member of the human gut microbiome: *B. fragilis*. Adaptive mutations in *B. fragilis* was shown to lead to the establishment of multiple co-existing lineages within the same host**.

55.•• Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol. 2019;17:e3000102. https://doi.org/10.1371/journal.pbio.3000102 **This study showed evidence of within-host evolution through mutations, recombination and adaptation across 40 species of the human gut microbiome**.

56. Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. Nat Medicine. 2019 *in-press.*

57. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9:811–4. https://doi.org/10.1038/nmeth.2066.

58.• Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. Nature. 2012;493:45–50. https://doi.org/10.1038/nature11711 One of the first studies to analyze the patterns of selection of the human gut microbiome using shotgun metagenomics.

59. Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. mSystems. 2017;2: e00164–16. https://doi.org/10.1128/mSystems.00164-16.
60. Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. Cell. 2015;160:583–94. https://doi.org/10.1016/j.cell.2014.12.038.
61. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJM, et al. Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci U S A. 2015;12: E2930–8. https://doi.org/10.1073/pnas.1423854112.
62. Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, Sunagawa S, et al. Temporal and technical variability of human gut metagenomes. Genome Biol. 2015;16:73. https://doi.org/10.1186/s13059-015-0639-8.
63. Zhu A, Sunagawa S, Mende DR, Bork P. Inter-individual differences in the gene content of human gut bacterial species. Genome Biol. 2015;16:82. https://doi.org/10.1186/s13059-015-0646-9.
64. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. Nature. 2017;550:61–6. https://doi.org/10.1038/nature23889.
65. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 2016;26:1612–25. https://doi.org/10.1101/gr.201863.115.
66. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, et al. Structural variation in the gut microbiome associates with host health. Nature. 2019;568:43–8. https://doi.org/10.1038/s41586-019-1065-y.
67. Costea PI, Coelho LP, Sunagawa S, Munch R, Huerta-Cepas J, Forslund K, et al. Subspecies in the global human gut microbiome. Mol Syst Biol. 2017;13:960. https://doi.org/10.15252/msb.20177589.
68.•• Ghalayini M, Launay A, Bridier-Nahmias A, Clermont O, Denamur E, Lescat M, et al. Evolution of a dominant natural isolate of *Escherichia coli* in the human gut over the course of a year suggests a neutral evolution with reduced effective population size. Appl Environ Microbiol. 2018;84:e02377–17. https://doi.org/10.1128/AEM.02377-17 **The authors analyzed the evolution of *E. coli* within human hosts and estimated a significantly higher mutation rate than reported by *in vitro* analyses. In contrast to other studies, no signs of selection were detected**.
69. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, et al. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. Proc Natl Acad Sci U S A. 2013;110:222–7. https://doi.org/10.1073/pnas.1219574110.
70.• Raymann K, Bobay L-M, Moran NA. Antibiotics reduce genetic diversity of core species in the honeybee gut microbiome. Mol Ecol. 2017;27:2057–66. https://doi.org/10.1111/mec.14434 **Population genetics analysis of two species of the honey bee gut microbiome following antibiotic treatments using the MAST approach**.
71. Powell JE, Ratnayeke N, Moran NA. Strain diversity and host specificity in a specialized gut symbiont of honey bees and bumble bees. Mol Ecol. 2016;25(18):4461–71. https://doi.org/10.1111/mec.13787.
72. Caro-Quintero A, Ochman H. Assessing the unseen bacterial diversity in microbial communities. Genome Biol Evol. 2015;7: 3416–25. https://doi.org/10.1093/gbe/evv234.
73. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, et al. Cospeciation of gut microbiota with hominids. Science (80-). 2016;353:380–2. https://doi.org/10.1126/science.aaf3951.
74. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2016;12:902–3. https://doi.org/10.1038/nmeth.3589.
75. Zou Y, Xue W, Luo G, Deng Z, Qin P, Guo R, et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. Nat Biotechnol. 2019;37:179–85. https://doi.org/10.1038/s41587-018-0008-8.
76.• Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 2016;10:1589–601. https://doi.org/10.1038/ismej.2015.241 **This study followed the evolution of the microbial community of the same lake sampled over seven years.** Results revealed that most species evolve by gene sweeps rather than genome sweeps.
77. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. Nature. 2019;568:499–504. https://doi.org/10.1038/s41586-019-0965-1.
78. Alper J, Anestidou L, Ogilvie J. Animal models for microbiome research: advancing basic translational science. Proceedings of a Workshop. Natl. Acad. Press. 2018.
79. Charbonneau MR, O'Donnell D, Blanton LV, Totten SM, Davis JCC, Barratt MJ, et al. Sialylated milk oligosaccharides promote microbiota-dependent growth in models of infant undernutrition. Cell. 2016;164:859–71. https://doi.org/10.1016/j.cell.2016.01.024.
80. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? Dis Model Mech. 2015;8:1–16. https://doi.org/10.1242/dmm.017400.
81. Hugenholtz F, de Vos WM. Mouse models for human intestinal microbiota research: a critical evaluation. Cell Mol Life Sci. 2018;75:149–60. https://doi.org/10.1007/s00018-017-2693-8.
82. Gaulke CA, Arnold HK, Humphreys IR, Kembel SW, O'Dwyer JP, Sharpton TJ. Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota. MBio. 2018;9:e01348–18. https://doi.org/10.1128/mBio.01348-18.
83. Burns AR, Guillemin K. The scales of the zebrafish: host–microbiota interactions from proteins to populations. Curr Opin Microbiol. 2017;38:137–41. https://doi.org/10.1016/j.mib.2017.05.011.
84. Douglas AE. The Drosophila model for microbiome research. Lab Anim (NY). 2018;47:157–64. https://doi.org/10.1038/s41684-018-0065-0.
85. Correa MA, Matusovsky B, Brackney DE, Steven B. Generation of axenic Aedes aegypti demonstrate live bacteria are not required for mosquito development. Nat Commun. 2018;9:4464. https://doi.org/10.1038/s41467-018-07014-2.
86.• Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA. Honey bees as models for gut microbiota research. Lab Anim (NY). 2018;47:317–25. https://doi.org/10.1038/s41684-018-0173-x **This review highlights the tractability and usefulness of the honey bee as a model system for gut microbiome research**.
87. Zhang F, Berg M, Dierking K, Félix MA, Shapira M, Samuel BS, et al. Caenorhabditis elegans as a model for microbiome research. Front Microbiol. 2017;8:485. https://doi.org/10.3389/fmicb.2017.00485.
88. Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, et al. Evidence for a core gut microbiota in the zebrafish. ISME J. 2011;5:1595–608. https://doi.org/10.1038/ismej.2011.38.
89. Guégan M, Zouache K, Démichel C, Minard G, Tran Van V, Potier P, et al. The mosquito holobiont: fresh insight into mosquito-microbiota interactions. Microbiome. 2018; 6:49. doi: https://doi.org/10.1186/s40168-018-0435-2.

90. Dirksen P, Marsh SA, Braker I, Heitland N, Wagner S, Nakad R, et al. The native microbiome of the nematode Caenorhabditis elegans: gateway to a new host-microbiome model. BMC Biol. 2016;4:38. https://doi.org/10.1186/s12915-016-0258-1.
91. Broderick NA, Lemaitre B. Gut-associated microbes of Drosophila melanogaster. Gut Microbes. 2012;3:307–21. https://doi.org/10.4161/gmic.19896.
92. Raymann K, Moran NA. The role of the gut microbiome in health and disease of adult honey bee workers. Curr Opin Insect Sci. 2018;26:97–104. https://doi.org/10.1016/j.cois.2018.02.012.
93. Kwong WK, Moran NA. Gut microbial communities of social bees. Nat Rev Microbiol. 2016;14:374–84. https://doi.org/10.1038/nrmicro.2016.43.
94. Moran NA. Genomics of the honey bee microbiome. Curr Opin insect Sci. 2015;10:22–8. https://doi.org/10.1016/j.cois.2015.04.003.
95. Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A. 2012;109:11002–7. https://doi.org/10.1073/pnas.1202970109.
96.•• Ellegaard KM, Engel P. Genomic diversity landscape of the honey bee gut microbiota. Nat Commun. 2019;10:446. https://doi.org/10.1038/s41467-019-08303-0 **Authors use shotgun metagenomics to access strain diversity within the bee gut and find that each honey bee harbors a distinct community at the functional level. Findings highlight the need to move beyond phylotype-level characterizations to understand the function, and illustrates the potential for using the honey bee as a model system for strain-level analysis**.
97. Ellegaard KM, Engel P. Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. Front Microbiol. 2016;7:1475. https://doi.org/10.3389/fmicb.2016.01475.
98.• Kešnerová L, RAT M, Ellegaard KM, Troilo M, Sauer U, Engel P. Disentangling metabolic functions of bacteria in the honey bee gut. PLoS Biol. 2017;15:e2003467. https://doi.org/10.1371/journal.pbio.2003467 **This study used untargeted metabolomics to profiled metabolic changes in gnotobiotic bees that were colonized with a complete microbiota. The authors found that honey bee gut bacteria can metabolize a wide range of compounds in the gut and provide insights into how metabolic activities are divided within gut communities**.
99. Kwong WK, Moran NA. 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 memb. Int J Syst Evol Microbiol. 2013;63:2008–18. https://doi.org/10.1099/ijs.0.044875-0.
100. Engel P, Kwong WK, Moran NA. Frischella perrara gen. nov., sp. nov., a gammaproteobacterium isolated from the gut of the honeybee, Apis mellifera. Int J Syst Evol Microbiol. 2013;63:3646–51. https://doi.org/10.1099/ijs.0.049569-0.
101. Kwong WK, Engel P, Koch H, Moran NA. Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc Natl Acad Sci U S A. 2014;111:11509–14. https://doi.org/10.1073/pnas.1405838111.
102. Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, et al. Metabolism of toxic sugars by strains of the bee gut symbiont Gilliamella apicola. MBio. 2016;7:e01326–16. https://doi.org/10.1128/mBio.01326-16.
103. Powell JE, Leonard SP, Kwong WK, Engel P, Moran NA. Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. Proc Natl Acad Sci U S A. 2016;113:13887–92. https://doi.org/10.1073/pnas.1610856113.
104. Steele MI, Kwong WK, Whiteley M, Moran NA. Diversification of type VI secretion system toxins reveals ancient antagonism among bee gut microbes. MBio. 2017;8:e01630–17. https://doi.org/10.1128/mBio.01630-17.
105. Engel P, Stepanauskas R, Moran NA. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. PLoS Genet. 2014;10:e1004596. https://doi.org/10.1371/journal.pgen.1004596.
106.• Ellegaard KM, Brochet S, Bonilla-Rosso G, Emery O, Glover N, Hadadi N, et al. Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. Mol Ecol. 2019;28:2224–37. https://doi.org/10.1111/mec.15075 **The authors study host specialization of gut symbionts by investigating Lactobacillus Firm5, a bacteria found in the gut of both honey bees and bumble bees that separates into deep-branching host-specific phylogenetic lineages. Results revealed genomic changes underlying host specialization**.
107. Bonilla-Rosso G, Engel P. Functional roles and metabolic niches in the honey bee gut microbiota. Curr Opin Microbiol. 2018;43:69–76. https://doi.org/10.1016/j.mib.2017.12.009.