Published in final edited form as:
Curr Opin Chem Biol. 2008 February ; 12(1): 109–114.

Host-bacterial coevolution and the search for new drug targets

Jesse Zaneveld[1], **Peter J. Turnbaugh**[2], **Catherine Lozupone**[2,3], **Ruth E. Ley**[2], **Micah Hamady**[4], **Jeffrey I. Gordon**[2], and **Rob Knight**[3,*]

1 *Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309*

2 *Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO 63108*

3 *Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309*

4 *Department of Computer Science, University of Colorado, Boulder, CO 80309*

Abstract

Understanding coevolution between humans and our microbial symbionts and pathogens requires complementary approaches, ranging from community analysis to in-depth analysis of individual genomes. Here we review the evidence for coevolution between symbionts and their hosts, the role of horizontal gene transfer in coevolution, and genomic and metagenomic approaches to identifying drug targets. Recent studies have shown that our symbiotic microbes confer many metabolic capabilities that our mammalian genomes lack, and that targeting mechanisms of horizontal gene transfer is a promising new direction for drug discovery. Gnotobiotic ("germ-free") mice are an especially exciting new tool for unraveling the function of microbes, whether individually or in the context of complex communities.

Keywords

coevolution; genomics; metagenomics; symbiosis

Introduction

The ~10^{14} microbes that live in and on each of our bodies belong to all three domains of life on earth — bacteria, archaea and eukarya. They outnumber our own cells by a factor of 10, and contribute many physiological capabilities, including metabolism of glycans and amino acids, synthesis of vitamins and isoprenoids, and biotransformation of xenobiotics[1••]. A deeper understanding of our human biology thus requires understanding of our microbial communities and the genes that they harbor ('our' microbiome) (Fig. 1). The notion that we have a 'meta-genome' composed of microbial and human components, and a 'meta-metabolome' that reflects metabolic activities carried out by both our microbial and our *H. sapiens* cells, has implications for the definition of health, discernment of disease susceptibilities, and diagnosis of human pathologies. This view of ourselves also opens up another dimension to therapeutics, including treatment strategies that accommodate microbial metabolism of drugs that target our human cells, and a new generation of therapeutics that affect the structure and function of our indigenous microbial communities. The vast majority

* To whom correspondence should be addressed: Rob Knight: Rob.Knight@colorado.edu, 303-492-1984 (phone), 303-492-7744 (fax).

of our microbes live in the gut. Thus, the current challenge is to understand the extent to which each individual's gut microbiota affects the bioavailability and host/microbial responses to orally or parenterally administered drugs, and the impact of interventions that alter our microbial ecology.

The concepts of coevolution and co-differentiation

One key step in understanding our microbiota is to identify lineages that have coevolved with humans (or with mammals in general), and to identify the genomic consequences of this coevolution. Coevolution between a host and a beneficial symbiont, or a pathogen, is defined as reciprocal adaptation of each lineage in response to the other[2••]. For example, genetic changes that increase production of a metabolite by an intestinal bacterium may trigger selection of changes in the host genome that promote uptake or prevent synthesis of that metabolite. Coevolution can also result in co-differentiation. Co-differentiation is defined as the diversification of host and symbiont lineages in parallel through a history of constant association; however, coevolution and co-differentiation can occur independently of one another [2••]. Co-differentiation can be detected by showing that host and symbiont phylogenies match (see [3•,4•] for detailed reviews of methods). Some methods for detecting co-differentiation can also generate hypotheses about processes causing specific differences between host and symbiont phylogenies. These differences (Fig. 2) include (i) the absence of a symbiont in a host lineage, due to extinction of the microbial species in the microbiota that occupies a given body habitat in the host, or due to under-sampling; (ii) speciation of a symbiont within the same host, so that the host contains two closely related species of the symbiont; (iii) failure of the symbiont to speciate when the host speciates, so that two closely related species of the host contain the same symbiont; and (iv) host-switching, i.e. transfer of a symbiont to a different host[3]. For example, in the case of (iv), unrelated mammal species that independently acquired similar diets may support the same symbiont species, which evolved in one and then found conditions in the other to be suitable.

An extension of the concept of co-differentiation between host and symbiont is co-differentiation of an entire microbial community with a host animal lineage. Here, an entire microbial community would be passed vertically from host to offspring. Over the course of speciation events in the host lineage, the microbial communities would differentiate in a way that would mirror host phylogeny. Such a scenario would be expected in host lineages where parents inoculate their offspring with a microbial consortium that is highly adapted to a specialized diet. An example of such hosts is the Koala bear: mothers inoculate their young with "pap", a specialized dropping that allows the young to make the transition from milk to a folivorous diet of *Eucalyptus* leaves and branches[5].

Coevolution has been invoked to describe the relationship between mammals and their gut microbial communities, because communities differ between species (mouse, cow, pig and human[6]). However, these differences could instead stem from selection of microbial lineages by a host's diet or immune system. Little is known about variation within each species, so differences between samples could also primarily reflect differences between individuals rather than between species. Similarly, gut microbial communities may be composed of environmental microbes pre-adapted to the chemical milieu of the gut, rather than microbes that have co-evolved with their hosts. Unambiguous demonstration of co-diversification of mammals and their gut microbes has not yet been achieved: for example, patterns of community similarity obtained by comparing the gut microbiotas of a range of mammals (e.g. using distance measurement algorithms such as UniFrac[7]), might mirror the phylogeny of the mammals. Such a test requires a systematic survey of the microbial communities associated with animal hosts representing a range of taxonomic orders and diets.

Identifying genes critical for symbiosis

Functional and comparative genomic analyses of human gut symbionts are revealing genes critical for adaptation to the gut environment, and mechanisms for horizontal transfer of these genes. These studies, along with in-depth analyses of symbionts of invertebrate hosts, provide a necessary framework for designing and interpreting metagenomic studies of the human microbiome.

In contrast to the human gut, which houses a diverse microbial community, many invertebrates (e.g. aphids, sharpshooters, and stinkbugs) have simple communities that are either maternally transmitted directly to the offspring[8,9••] or are eaten as maternally-deposited capsules shortly after hatching[10]. These symbiont genomes have dramatically reduced gene content, but retain genes for key metabolic capabilities that complement host physiology, including vitamin and amino biosynthesis[8,9••]. However, some human gut symbionts, including members of a large division of bacteria known as the Bacteroidetes, have maintained a larger genome size[11•], perhaps because they must survive outside the host to be transmitted.

Studies of mice raised to adulthood in sterile isolators without any exposure to microbes ('germ-free' animals) are especially useful complements to genomic approaches for understanding the function of gut microbes. For example, germ-free mice colonized with a prominent human gut symbiont, *Bacteroides thetaiotaomicron*, demonstrate that this bacterium can selectively induce a set of its genes that degrade otherwise indigestible dietary polysaccharides[12••]. Genomic analysis of *Methanobrevibacter smithii* showed that this species likely promotes energy harvest in hosts by consuming a range of fermentation products of other gut bacteria, and may be a good target for anti-obesity drugs[13]. Comparative genomic analyses of gut and non-gut Bacteroidetes[11•] revealed that gut Bacteroidetes possess large arsenals of genes that sense the nutrient environment. These nutrient sensors are linked to gene clusters encoding proteins involved in acquiring specific classes of glycans, and degrading these glycans by glycoside hydrolases and polysaccharide lyases. The products of these polysaccharide utilization gene clusters are used by other members of the microbiota that are ill-equipped to degrade complex glycans, but are well-endowed with genes involved in importing monosaccharides and converting them to fermentation products that can be utilized by other components of the microbiota, and the host[14•]. These types of studies can be expanded to model communities of sequenced gut symbionts that are introduced into normal or genetically engineered germ-free mice: the effects of diet and or drugs can be carefully monitored in these 'gnotobiotic' mouse models under conditions where potentially confounding variables, such as host genotype and diet, can be constrained. Gnotobiotic mouse studies will provide better understanding of the rules and forces that govern the assembly and operations of microbial communities, proof-of-principle experiments that ascertain the contributions of specified groups of microbes to community and host operations, and proof-of-concept tests of the efficacy of new types of anti-microbial drugs that target horizontal gene transfer between members of a microbiota or the activities of virulence factors embedded in a microbiome (see below).

A preliminary study of microbial gene content in the fecal microbiota, which mirrors the microbiota of the distal gut, of several healthy humans showed that compared to our *H. sapiens* genome and the genomes of all sequenced microbes, there is an enrichment of the representation of genes involved in vitamin biosynthesis, degradation of diet- and host-derived polysaccharides as well as xenobiotic metabolism[1••]. Further analysis revealed that the gut microbiome is also enriched in a family of conjugative transposons, consistent with a pronounced role of horizontal gene transfer (HGT) in shaping gut microbial genomes[15].

Horizontal gene transfer and coevolution

HGT is an important factor in the evolution of microbial communities that promotes adaptation to novel or changing environments, including mammalian host environments. HGT is of intense medical interest, not only because of its contribution to the spread of antibiotic resistance genes, but also because it can cause closely related strains to differ drastically in clinical parameters. For example, type III secreted effectors may cause differences in host specificity between strains of *Salmonella enterica*[16]. On a longer time scale, the acquisition of the type III secretory systems encoded by the SPI-1 and SPI-2 pathogenicity islands is a defining feature of host adaptation for *S. enterica* as a whole[17].

Several novel strategies for drug development are being pursued in response to the challenge posed by the horizontal transfer of genes involved in both antibiotic resistance and virulence. These strategies include the development of compounds that directly inhibit gene transfer [18••] or virulence[19].

Targeting virulence factors with small-molecule inhibitors directly presents several potential advantages. Such targeting may cause less collateral damage to the indigenous microbiome than traditional antibiotics, may exert less selective pressure for the evolution and transfer of resistance, and may be effective against divergent organisms that have acquired a particular virulence factor by HGT.

Genomic islands contain a rich source of genes of unknown function[20•] that may yield novel virulence factors appropriate for small-molecule inhibitors. For example, a recent screen for *Salmonella enterica* serovar Typhimurium genes involved in survival and replication within macrophages (a key feature of persistent infection by *Salmonella*) found that such genes were dramatically overrepresented within putatively transferred regions, such as prophages and pathogenicity islands[21].

Gene transfer systems themselves are also being targeted with small-molecule inhibitors. Such inhibitors could be co-administered with antibiotics to prevent the *in-vivo* acquisition of resistance factors by susceptible pathogens during the course of antibiotic therapy. For example, the bisphosphate-containing compounds clodronate and etidronate inhibit the F plasmid TraI relaxase *in vitro* and conjugative transfer of F plasmid *in vivo*. These findings are particularly significant because relaxases are essential components of conjugative transfer systems, and the F plasmid TraI relaxase is closely related (~99% sequence identity) to the relaxases of many plasmids known to transfer antibiotic resistance genes[18••].

The promise of metagenomic approaches

The vast majority of phylogenetic diversity in microbial communities associated with the human body (and other ecosystems) is represented by organisms that are difficult or impossible to culture in the laboratory using currently available methods. Standard culture methods are especially problematic for understanding symbiosis, because microbes may express completely different sets of genes or may not grow at all outside the host. Metagenomics allows us to observe the genes contained in this vast uncultured majority through the isolation and sequencing of DNA directly from the community. Typically, 16S rRNA gene sequences are used as a phylogenetic marker to probe community structure and diversity (‘who’s there, and in what abundance?’). The rest of the genes in the microbiome are characterized through shotgun sequencing of whole microbial community DNA. Although the short 200–250 nucleotide reads currently obtained by the latest generation of massively parallel DNA sequencers (i.e. pyrosequencers) are sufficient for characterizing communities based on their 16S rRNA gene content[22], characterizing other genes typically requires a combination of short pyrosequencing reads, longer Sanger-sequencing reads, and, ideally, complete reference

genomes (i.e. genomes of cultured representatives of major phylogenetic lineages present in the community) (Fig. 3). In simple communities, such as those present in certain invertebrates or in the environment, metagenomic data can be sufficient for assembling the genomes of their constituent microbes. Nonetheless, assembly of sequenced genomic DNA fragments from a microbiome into relatively large contiguous sequences of physically linked genes derived from a given organismal genome remains challenging, especially in complex communities such as the human gut microbiota [1••,15,23•]. Despite these challenges, metagenomic studies have revealed specific genes that are enriched in the gut microbiome of humans (see above), as well as microbial genes whose representation is enriched in mouse models of human diseases, including obesity[23•].

Because assembly is difficult, metagenomic analysis of complex communities is currently 'gene-centric': DNA sequences are mapped to known genes and genomes to infer the relative abundance of different genes and metabolic pathways[23•,24••]. One major challenge in metagenomics is to link these 'gene centric' functional predictions to the organism that contained each gene. This goal is complicated by a lack of complete reference genomes and complete/consensus reference taxonomies, and by the short DNA fragment generated by pyrosequencing. Functional assignments typically rely on homology searches. Taxonomic assignment is more complex, and usually involves aligning homologs, and building phylogenetic trees. Annotations can then be assigned based on the best matches or closest homologs in a reference set of genes. Thus, taxonomic assignment is challenging and computationally expensive even with full-length sequences, and results are strongly affected by the reference set.

Another major challenge is that individual labs usually do not have the capability to completely characterize a complex community through metagenomics: many key analysis tools do not scale to large datasets, and the cost of sequencing and the computational resources can be formidable. The most successful studies have come from collaborations spanning a range of disciplinary expertise, including physiology, microbiology, molecular biology, statistics, molecular evolution, ecology and high-performance computing. Collaboration and coordination at different project scales are clearly needed[14•].

Conclusions

Coevolution between mammals and their microbial symbionts is probably more complex than co-divergence, and HGT appears to play a key role. 16S rRNA and metagenomic profiling of mammal-associated microbial communities associated with health and disease, in combination with carefully chosen complete genomes, will likely reveal new drug targets for the manipulation of microbial community structure and function. This is one of the hoped-for outcomes of the international human microbiome project, which is now being launched in several countries [14•].

Acknowledgements

This work was supported by NIH grant P01DK078669, the NIH/CU Molecular Biophysics Training Program T32GM065103, and the NIH/CU Signaling and Cell Regulation Training Program T32 GM08759.

References

1••. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. Science 2006;312:1355–1359. [PubMed: 16741115]The first large-scale analysis of a human-associated microbiome.

2••. Moran NA. Symbiosis. Curr Biol 2006;16:R866–871. [PubMed: 17055966]This 'primer' provides definitions and examples for different forms of symbiosis, and as such lays a framework for

comparative studies of different biological systems where coevolution or co-diversification are thought to occur.

3•. Charleston MA, Perkins SL. Traversing the tangle: algorithms and applications for cophylogenetic studies. J Biomed Inform 2006;39:62–71. [PubMed: 16226921]Useful recent review of methods for testing for co-divergence, including discussion of the strengths and weaknesses of the conceptual underpinnings of several of the methods.

4•. Stevens J. Computational aspects of host-parasite phylogenies. Brief Bioinform 2004;5:339–349. [PubMed: 15606970]Exceptionally clear review of methods for testing for co-divergence between hosts and parasites (or symbionts). Includes notes on the availability of the different methods.

5. Osawa R, Blanshard WH, Ocallaghan PG. Microbiologial Studies of the Intestinal Microflora of the Koala, *Phascolarctos cinereus*. 2. Pap, a Special Maternal Feces Consumed by Juvenile Koalas. Australian Journal of Zoology 1993;41:611–620.

6. Dethlefsen L, McFall-Ngai M, Relman DA. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature 2007;449:811–818. [PubMed: 17943117]

7. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71:8228–8235. [PubMed: 16332807]

8. Ochman H, Moran NA. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 2001;292:1096–1099. [PubMed: 11352062]

9••. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, et al. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol 2006;4:e188. [PubMed: 16729848]A fascinating study of the genomic consequences of host-symbiont evolution in insects, showing how the symbiont can allow the host to become a crop pest.

10. Hosokawa T, Kikuchi Y, Shimada M, Fukatsu T. Obligate symbiont involved in pest status of host insect. Proc Biol Sci 2007;274:1979–1984. [PubMed: 17567556]

11•. Xu J, Mahowald MA, Ley RE, Lozupone CA, Hamady M, Martens EC, Henrissat B, Coutinho PM, Minx P, Latreille P, et al. Evolution of Symbiotic Bacteria in the Distal Human Intestine. PLoS Biol 2007;5:e156. [PubMed: 17579514]A comparative genomic analysis of multiple human gut symbionts that demonstrates the importance of horizontal gene transfer in shaping gut genomes.

12••. Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 2005;307:1955–1959. [PubMed: 15790854]Functional genomic analyses of a prominent human gut symbiont in gnotobiotic mice.

13. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille P, Kim K, Wilson RK, Gordon JI. Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. Proc Natl Acad Sci U S A 2007;104:10643–10648. [PubMed: 17563350]

14•. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature 2007;449:804–810. [PubMed: 17943116]A overview and prospectus of the recently initiated international effort to understand the microbiome of multiple human-associated habitats.

15. Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res 2007;14:169–181. [PubMed: 17916580]

16. Hansen-Wester I, Stecher B, Hensel M. Analyses of the evolutionary distribution of Salmonella translocated effectors. Infect Immun 2002;70:1619–1622. [PubMed: 11854253]

17. Hensel M. Evolution of pathogenicity islands of Salmonella enterica. Int J Med Microbiol 2004;294:95–102. [PubMed: 15493819]

18••. Lujan SA, Guogas LM, Ragonese H, Matson SW, Redinbo MR. Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proc Natl Acad Sci U S A 2007;104:12282–12287. [PubMed: 17630285]This study demonstrates that several bis-phosphate containing compounds, including two already in clinical use in other applications, serve as inhibitors of the TraI relaxase *in vitro* and conjugative F plasmid transfer *in vivo*.

19. Dahlgren MK, Kauppi AM, Olsson IM, Linusson A, Elofsson M. Design, Synthesis, and Multivariate Quantitative Structure-Activity Relationship of Salicylanilides-Potent Inhibitors of Type III Secretion in Yersinia. J Med Chem 2007;50:6177–6188. [PubMed: 17975903]

20•. Hsiao WW, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FS. Evidence of a large novel gene pool associated with prokaryotic genomic islands. PLoS Genet 2005;1:e62. [PubMed: 16299586]This bioinformatics study demonstrates that genomic islands are enriched in genes of unknown function.

21. Klumpp J, Fuchs TM. Identification of novel genes in genomic islands that contribute to Salmonella typhimurium replication in macrophages. Microbiology 2007;153:1207–1220. [PubMed: 17379730]

22. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res 2007;35:e120. [PubMed: 17881377]

23•. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 2006;444:1027–1031. [PubMed: 17183312]A demonstration that comparative metagenomics of complex microbial communities using short pyrosequencing reads, coupled to gnotobiotics, can be used to identify and test novel hypotheses.

24••. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, et al. Comparative metagenomics of microbial communities. Science 2005;308:554–557. [PubMed: 15845853]An early demonstration that comparative metagenomics can reveal specific genes that are important for the microorganisms in a specific environment.

Fig. 1. Understanding the microbial part of ourselves

A key part of understanding human metabolic capabilities is to understand our microbial symbionts, and the genomes of those symbionts.

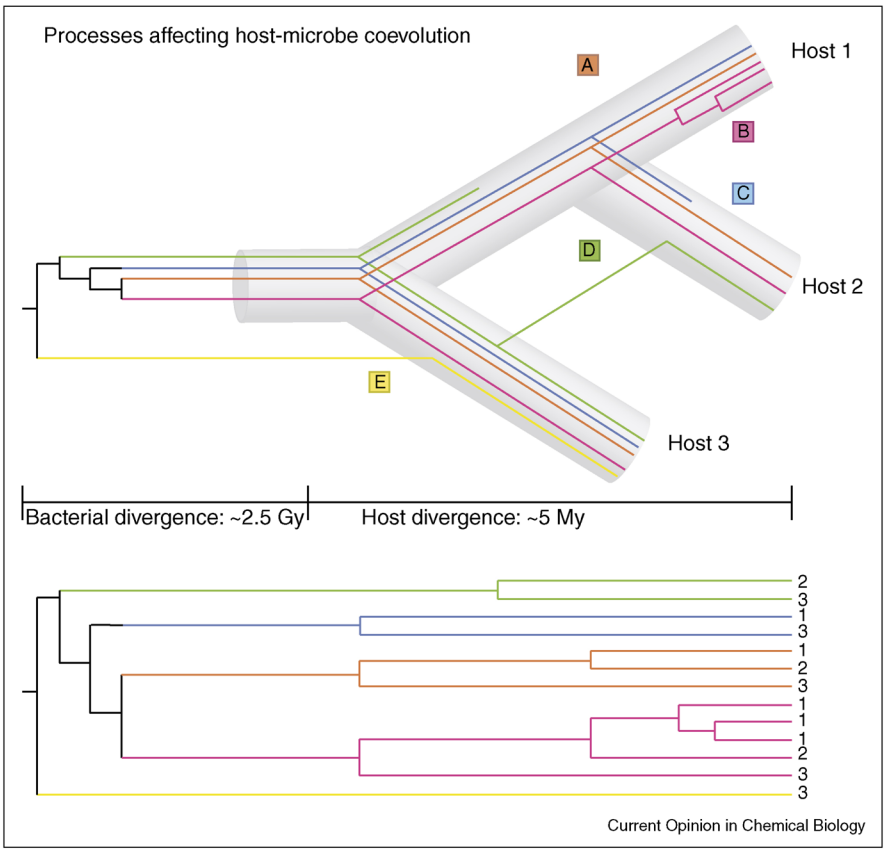

Fig 2. Processes affecting host-microbe coevolution

Large gray tubes indicate relationships between host lineages, while thin colored lines indicate relationships between microbial lineages (**top tree**). Importantly, divergence between bacterial divisions (black lines) occurred over a much longer time-scale than divergence between metazoan host lineages such as, for example, humans, chimpanzees and bonobos (see scale). Processes depicted include co-divergence of host and microbial lineages **(A)**, diversification of a microbial lineage while associated with a host **(B)**, extinction **(C)**, host-swapping **(D)** and association of a free-living microbial lineage with a host **(E)**. Co-divergence, adaptation to a novel host, host-swapping or diversification within a lineage all produce splits on the microbial phylogenetic tree (**bottom tree**). Although great caution and deep sampling are required, studies of the distribution of microbial lineages among hosts (numbers at tips of bottom tree), and comparison of the microbial phylogeny to the host phylogeny, may help to resolve which evolutionary processes are responsible for observed divergence between host-associated microbial lineages.

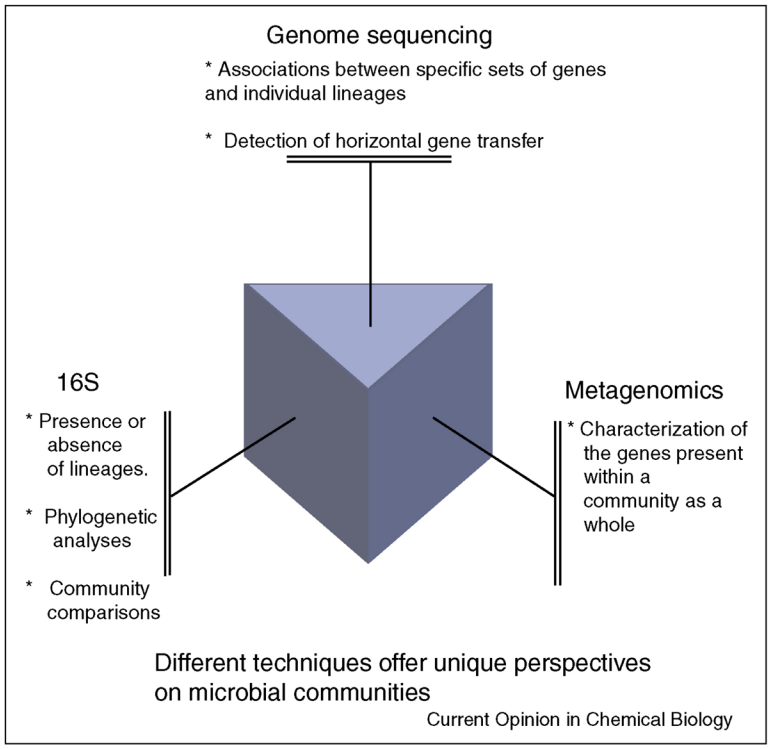

Fig. 3. A comprehensive understanding of our microbial ecology requires integration of many data sources

Community profiles with 16S rRNA indicate which types of organisms are present, metagenomic profiling allows us to identify specific functional categories of genes that are critical for differences in function, and complete genomes act as scaffolds for understanding changes in gene content through loss, amplification, and HGT that allow microbes to adapt to functional roles in different environments.