



Application of sequence-based methods in human microbial ecology

Li Weng, Edward M. Rubin and James Bristow

Genome Res. 2006 16: 316-322

Access the most recent version at doi:[10.1101/gr.3676406](https://doi.org/10.1101/gr.3676406)

References This article cites 82 articles, 48 of which can be accessed free at:
<http://genome.cshlp.org/content/16/3/316.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text "CRISPR and RNAi Genetic Screening. Your new superpower." is written in white. In the center, there is a white-bordered box containing the words "LEARN MORE" in blue. On the right, there is a photograph of a woman wearing a red superhero mask and a red cape over a white shirt. To the right of the photo is the Cellecta logo, which consists of a cluster of green dots of varying sizes, with the word "CELLECTA" in white capital letters below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Application of sequence-based methods in human microbial ecology

Li Weng, Edward M. Rubin, and James Bristow¹

Joint Genome Institute, Walnut Creek, California 94598, USA

Ecologists studying microbial life in the environment have recognized the enormous complexity of microbial diversity for many years, and the development of a variety of culture-independent methods, many of them coupled with high-throughput DNA sequencing, has allowed this diversity to be explored in ever-greater detail. Despite the widespread application of these new techniques to the characterization of uncultivated microbes and microbial communities in the environment, their application to human health and disease has lagged. Because DNA-based techniques for defining uncultured microbes allow not only cataloging of microbial diversity but also insight into microbial functions, investigators are beginning to apply these tools to the microbial communities that abound on and within us, in what has aptly been called “the second Human Genome Project.” In this review we discuss the sequence-based methods for microbial analysis that are currently available and their application to identify novel human pathogens, improve diagnosis of known infectious diseases, and advance understanding of our relationship with microbial communities that normally reside in and on the human body.

Sequence-based methods

It has long been recognized that standard culture methods fail to adequately represent the enormous microbial diversity that exists in nature because of the fastidious growth requirements of many microorganisms. Even when growth conditions are altered to mimic environmental nutrient composition, up to 80% of organisms identified by culture-independent methods fail to grow in culture (Connon and Giovannoni 2002). To avoid reliance on cultivation, many culture-independent methods have been developed to search for novel bacterial species, including pathogens. These methods include screening of expression libraries with immune serum, nucleic acid subtractive methods, and small molecule detection with mass spectroscopy, among others, and these methods have been reviewed elsewhere (Relman 2002). This review will focus primarily on sequence-based methods because of their general applicability and the continued expansion of high-throughput, low-cost sequencing capacity.

The cornerstone of culture-independent identification of bacterial and archaeal species is sequence analysis of ribosomal RNA genes that are sufficiently well conserved across species that they can be amplified using PCR primers based on highly conserved sequences, yet are sufficiently diverse to differentiate bacterial or archaeal species. Carl Woese, in a series of seminal studies, initially used small subunit (16S) rRNA gene sequences for construction of phylogenetic trees of cultivated organisms (Woese and Fox 1977; Woese 1982; Gupta et al. 1983; Woese and Olsen 1986), but this method was subsequently applied to libraries of rRNA genes PCR-amplified from environmental DNA samples without cultivation (Stahl et al. 1984, 1985; Giovannoni et al. 1990; Ward et al. 1990; Schmidt et al. 1991). A striking collective result from the application of this technique to numerous environmental samples was the realization that cultivated organisms represent a tiny fraction of species present in most

environmental samples. Indeed only half of 52 currently recognized bacterial phyla contain cultivated members (Hugenholtz et al. 1998; Rappe and Giovannoni 2003). To maximize the utility of 16S rRNA gene analysis for species determination, it is now routine to amplify almost the entire 16S rRNA gene, which spans ~1500 bp and is thus readily sequenced in its entirety through bidirectional sequencing of cloned 16S amplicons (Hugenholtz 2002). After sequencing, 16S sequences are clustered into groups and a threshold of sequence similarity is established (usually 98 or 99%) to distinguish species (Fig. 1).

Although PCR amplification of 16S sequences has been of enormous value, there are caveats to this approach. One is that organisms that carry sequence differences within the highly conserved regions used for primer design may amplify less efficiently or not amplify at all. For example, the 16S rRNA gene sequence of the Nanoarchaeota is so divergent that PCR with the “universal” primers failed to detect this species even from cultured organisms (Huber et al. 2002). Second, PCR conditions, such as annealing temperature or extension time, may allow formation of chimeras or produce amplification bias that skews the representation of each species in cloned libraries (Wang and Wang 1997; Kroes et al. 1999; Ishii and Fukui 2001). Some of these errors may be recognized and corrected by hybridization-based methods such as *in situ* hybridization with species- or strain-specific 16S oligonucleotides applied to the original (or similar) sample.

A relative drawback of 16S rRNA gene sequencing is the need for significant sequencing capacity that, except in high-throughput sequencing centers, may be relatively slow compared with hybridization-based methods. As an alternative, several strategies employing 16S rRNA gene microarrays have been presented and offer some advantage in speed compared with sequencing when analysis of many samples is required (Guschin et al. 1997; Rudi et al. 2000; Small et al. 2001; Loy et al. 2002, 2005). For the most part these studies employed oligonucleotide probes that were designed for detection of specific organisms, such as sulfate-reducing bacteria or beta-proteobacteria, and have offered acceptable sensitivity. Application to highly complex environmental samples has been limited by sensitivity and difficulties in

¹Corresponding author.

E-mail jbristow@LBL.gov; fax (925) 296-5752.

Article published online ahead of print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.3676406>.

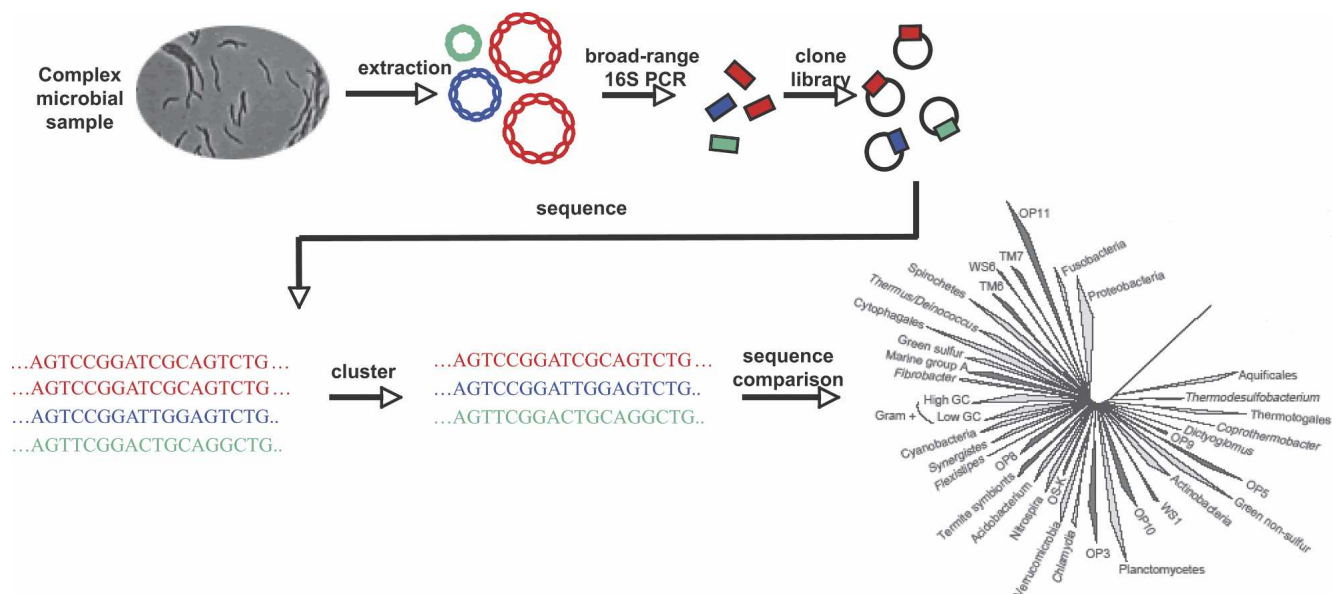


Figure 1. Broad-range PCR amplification and sequencing of microbial 16S rRNA genes. Genomic DNA extracted from a microbial community is used as a template for 16S rRNA PCR with “universal” primers specific for Archaea and Bacteria. The PCR products, which are about 1.5 kb in length, are cloned into a standard vector and both ends are sequenced. The aligned sequences are first clustered into groups, and the representatives from each cluster are compared with 16S rRNA gene databases for phylogenetic classification.

differentiating related species, but it seems reasonable to expect further improvement in this technology and eventual application to clinical materials from humans.

A final drawback to 16S rRNA gene sequencing is the absence of functional genomic information obtained. Recently genomic libraries have been created directly from DNA extracted from environmental samples and subjected to functional screens or to shotgun sequencing with the goal of assembly for the most abundant genome(s) present (Handelsman 2004). These “metagenomic” methods circumvent the needs for cultivation or PCR amplification. Recent metagenomic studies of an acid mine biofilm and the Sargasso Sea yielded significant insight into species diversity and ecology of uncultivated microbial communities (Tyson et al. 2004; Venter et al. 2004; Tringe et al. 2005). While not yet applied to clinical environments, metagenomic methods have the potential to provide functional characterization of complex, human-associated microbial communities.

Identification of novel human pathogens

An exciting application of culture-independent methods is the identification of uncultivated organisms that cause human disease. Because DNA can be extracted from any potentially infected material and used as a substrate for 16S rRNA gene amplification, Fredericks and Relman (1996) predicted that a rash of claims for disease causation by new pathogens would follow application of this method to human tissues and laid out a strategy for proving disease causation for organisms that might be difficult or impossible to grow. Reiman and Falkow (2001) later amplified these criteria. Remarkably, this has not come to pass and only a few new pathogens have been identified through 16S rRNA gene amplification. This may be because present methods for cultivation are sufficient for the vast majority of pathogens capable of growth on human tissues, or it may be that we have yet to apply culture-independent methods to many clinical conditions that

have an infectious component. Whatever the reason, these new infectious agents are notable and worthy of review here.

The first novel pathogen to be identified by sequence-based methods was *Rochalimaea henselae*, the organism responsible for bacillary angiomatosis (BA). The hallmark of BA is abnormal proliferation of small blood vessels in the skin and visceral organs of immunocompromised patients. Although bacteria had been found in tissue sections by Warthin-Starry staining, they could not be cultured because of their fastidious growth requirements (Perkocha et al. 1990). In 1990, Relman et al. amplified a partial 16S rRNA gene sequence from tissue samples obtained from bacillary angiomatosis patients, but not from normal tissues, using broad-range PCR (Relman et al. 1990). Analysis of this 16S sequence suggested a novel species most closely related to *Rochalimaea* spp. Further evidence for causation was provided by the isolation of a slow-growing, *Rochalimaea*-like bacillus from a single BA patient in an independent study (Slater et al. 1990). Two years later, genotypic analysis of the complete 16S rRNA gene and other genomic loci further confirmed the novelty of the isolated BA agent, *Rochalimaea henselae* (later moved into the genus *Bartonella*) (Regnery et al. 1992).

The same strategy was soon applied to other potentially infectious diseases and led to the identification of *Ehrlichia chaffeensis*, a new species associated with tick bites that causes a febrile illness. Ehrlichiosis is clinically similar to Rocky Mountain spotted fever, another tick-borne disease caused by the intracellular parasite, *Rickettsia rickettsii*. Although testing of the index patient’s serum for antibodies against *R. rickettsii* was negative, patient serum contained antibodies reactive to *E. canis*, a well-described canine pathogen (Maeda et al. 1987). This suggested that *E. canis* or a related species was responsible for disease, and 16S rRNA gene amplification from infected macrophages led to identification of *E. chaffeensis* (Anderson et al. 1991; Dawson et al. 1991). Causation is supported by concordance of *E. chaffeensis* 16S rRNA and serologic findings in patients with fever, leukocyte

inclusions, and history of a tick bite, as well as a salutary clinical response to appropriate antibiotics accompanied by loss of *E. chaffeensis* 16S rRNA gene from leukocytes.

A third example of success with 16S rRNA gene amplification is Whipple's disease. Whipple's disease is a rare disease first described in 1907 in a missionary who died of an illness marked by chronic joint pain, weight loss, and severe abdominal pain. In the report of this patient, "rod-like bacilli in a small node" were noted (Whipple 1907). Eighty-four years passed before the identification of the etiologic agent, despite its consistent observation in affected tissues (Chears Jr. and Ashworth 1961; Yardley and Hendrix 1961) and patients' improvement with antibiotic treatment (Trier et al. 1965). In 1991, a partial 16S rRNA gene sequence was amplified from a small-bowel biopsy specimen taken from a patient with Whipple's disease (Wilson et al. 1991), and the complete 16S rRNA gene sequence was determined a year later, revealing it to be an actinomycete not closely related to any known genus. It was therefore given a new genus and species name, *Tropheryma whipplei*, based on the unusual features of the disease and the distinct morphological characteristics of the bacillus (Relman et al. 1992). It is worth noting that *T. whipplei* was particularly recalcitrant to cultivation (Raoult et al. 2000). The complete genome sequence of *T. whipplei* predicted deficiencies in amino acid synthesis (Bentley et al. 2003; Raoult et al. 2003), and with this information, Renesto et al. (2003) successfully designed a complete medium that allowed cell-free cultivation of *T. whipplei*. This was the first demonstration that genomic information could guide rational design of media for axenic cultivation of fastidious bacteria.

Whereas most human tissues are normally devoid of cultivable microorganisms, many epithelial-lined cavities of the human body in contiguity with the environment harbor microbial communities, the complexities of which are just beginning to be understood. These include the skin, mouth, ear, gastrointestinal tract, and vagina. Identifying pathogens within this complex bacterial background is more difficult than identifying them in normally sterile compartments. One of the most successful examples of this involves the study of dental plaque. Because of their known role in dental caries and periodontal disease, human oral flora have been studied intensively through both culture-dependent and culture-independent techniques. About 500 bacterial species have been found in the human oral cavity (Thoden van Velzen et al. 1984; Meyer and Fives-Taylor 1998; Paster et al. 2001) and 40%–60% of these species are uncultivated "phylo-types" (Kroes et al. 1999; Paster et al. 2001). Studies using conventional culture methods have established that early colonizing streptococci play a key role in initiating the formation of dental caries (Kolenbrander et al. 1990; Whittaker et al. 1996), while *Actinobacillus actinomycetemcomitans*, *Porphyromonas gingivalis* and *Treponema denticola* contribute to the development of periodontal disease (Meyer and Fives-Taylor 1998). Recent 16S rDNA sequence analysis uncovered that in addition to these known pathogens, several additional organisms, including organisms assigned to uncultivated phyla *OP11* and *TM7*, are strongly associated with periodontitis or acute necrotizing ulcerative gingivitis in humans (Choi et al. 1994; Dewhirst et al. 2001; Paster et al. 2001; Brinig et al. 2003; Hutter et al. 2003; Ouverney et al. 2003).

Recently, methanogenic Archaea have also been linked to periodontal disease based on 16S rRNA sequencing and FISH analysis (Kulik et al. 2001; Lepp et al. 2004), providing the first example of archaeal disease association in humans. In the latter

study, the relative abundance of methanogenic Archaea was associated with disease severity and decreased with effective treatment (Lepp et al. 2004). However, Archaea were not uniformly found in the subgingival space of severely affected individuals, leaving open the question of whether the Archaea are playing a causative role in this polymicrobial disease. Interestingly, the abundance of Archaea and *T. denticola* were inversely correlated, suggesting that they might compete for the same niche in the community. It was hypothesized that both organisms serve as "hydrogen sinks" in the highly reduced environment of the subgingival space, allowing acid-producing members of the community to grow to higher density than they might in the absence of Archaea or Treponemes. While additional genomic and metabolic studies will be required to fully understand the role of methanogenic Archaea in periodontal disease, this example clearly illustrates how culture-independent methods may do more than increase our appreciation of the number of known pathogens. These methods also stand to teach us a great deal about the mechanisms of disease, especially in circumstances where the paradigm of a single causative agent seems not to hold. Other clinical circumstances that may benefit from similar analyses include sinusitis, ventilator-associated pneumonia, small-bowel overgrowth syndromes, inflammatory bowel disease, and bacterial vaginosis.

Novel viral pathogens

Although it is unclear how many new bacterial pathogens remain to be identified, it seems likely that a larger number of viral pathogens have thus far escaped detection. This is because growth conditions are far harder to determine and nucleic acid techniques based on sequence conservation are not available. Nonetheless there are several examples of successful application of culture-independent methods to the identification of novel viral pathogens. These include the use of DNA subtraction techniques to identify the Kaposi's sarcoma virus (Chang et al. 1994), expression screening of a cDNA library with patient serum to identify the hepatitis-C agent (Choo et al. 1989), and more recently hybridization-based screening to identify the SARS virus as a coronavirus (Ksiazek et al. 2003). The last example was particularly notable because novel viral sequences were recovered from individual elements of the array to which they hybridized and because the analysis was accomplished with remarkable speed.

The diversity of viral genomes clearly complicates the search for novel pathogens and development of new strategies may be required. One such strategy might employ large-scale sequencing of appropriately extracted clinical materials on a massively parallel, single-molecule sequencing apparatus. Sequencing machines capable of producing >200,000 short sequencing reads in a single run, from small amounts of DNA without prior cloning, are now commercially available (Andries et al. 2005). Such machines produce sequence at ~20% the cost of Sanger sequencing on capillary sequencers and the short reads produced by single-molecule techniques provide an efficient way of producing molecular tags for microbial species. Infected tissues or body fluids could be extracted, reverse transcribed if needed (for RNA viruses), and sequenced along with contaminating host or bacterial DNA. The resulting sequence reads would then be searched for homology to known viral (and bacterial) sequences. Viral sequences found could then be linked by subsequent PCR or used in hybridization-based strategies to recover complete viral genomes.

Sequence-based methods for rapid diagnosis of bacterial infections

Sequence-based methods have also found application to the rapid identification of human pathogens that can be cultured. For fastidious or slow-growing organisms the advantage is obvious, but there may be significant value in their application to more common infectious agents because standard culture techniques require 24–48 h for growth and identification of most bacterial species. Clinical practice has long favored the use of antibiotics to cover the organisms most likely to be present when infection is suspected, but this practice has contributed greatly to the spread of antibiotic resistance (Neu 1992; Cohen 2000). Hence, the use of culture-independent methods for rapid diagnosis has the potential to provide more specific antibiotic therapy from the outset or to withhold it altogether if there is no infection present. Because of the continuing need for analysis of antibiotic sensitivity, culture is still required for every serious infection, but one can envision a day when resistance is sufficiently well understood that this might also be rapidly predicted from nucleic acid-based assays.

Both hybridization and PCR-based strategies have been used for rapid diagnosis. 16S rRNA gene amplification can be carried out using universal primers followed by detection with group-specific fluorescent probes or a second group-specific PCR. When applied to tissues or fluids that are normally devoid of cultivable organisms, including blood, urine, cerebrospinal fluid, wounds, and indwelling intravascular catheters, these assays have generally been able to identify more than 95% of infected samples with false-positive rates of ~10%. (Qin and Urdahl 2001; Nikkari et al. 2002; Domann et al. 2003; Moumille et al. 2004; Schuurman et al. 2004). While these studies demonstrate acceptable performance, to our knowledge, no one has yet fulfilled the promise of these methods to guide therapy prospectively or to reduce the administration of broad-spectrum antibiotics to potentially infected patients while awaiting culture results.

Understanding complex microbial communities in human ecology

It is increasingly clear that humans have a symbiotic relationship with several microbial consortia living on and within us, but understanding the details of these relationships is a major challenge. One important application of culture-independent methods has been preliminary evaluation of these communities. In the discussion below we will focus on the gut microbial consortia because it clearly illustrates the complexity of the problem in a context where recent progress has been made.

It has long been appreciated that the human gastrointestinal tract is colonized by a complex microbial community whose numbers are believed to far exceed the total number of cells in the human body. Although this community has been studied in detail by relatively efficient culture techniques, culture-independent studies suggest that 40%–80% of the total microscopic counts are uncultivated species (Langendijk et al. 1995; Suau et al. 1999). Several studies have used 16S rRNA gene amplification to uncover numerous novel species (Wilson and Blitchington 1996; Suau et al. 1999). Quantitative assays of human fecal samples collected from several individuals have shown that each person has a unique collection of bacteria, and an individual's dominant flora is relatively stable over time (Franks et al. 1998; Zoetendal et al. 1998). These findings were recently confirmed with a large-scale 16S survey, which also found 62% of

bacterial phylotypes in the gut were novel and 80% represented uncultivated species, mostly from the *Firmicutes* and *Bacteroidetes* phyla (Eckburg et al. 2005).

The finding of greater similarity between gut microbes of monozygotic twins living apart than genetically unrelated individuals living together suggests that host genetics may significantly affect the composition of gut microbial flora (Zoetendal et al. 2001). A recent study demonstrated expansion of segmented filamentous bacteria in the gut of IgA-deficient mice compared with wild-type litter mates, confirming that host genotype molds the gut microbial consortium and suggesting that the host immune system is responsible for regulating normal flora in addition to protecting against pathogens (Suzuki et al. 2004).

Studies in mice, rats, and fish raised in sterile environments, and therefore lacking this community in the gut, have demonstrated the importance of this community for normal gut structure and function, nutrient absorption, fat deposition, and development of normal immunity (Falk et al. 1998; Guarner and Malagelada 2003; Backhed et al. 2004, 2005; Rawls et al. 2004). For example, in the recent study by Rawls et al. (2004), gene expression patterns were compared between germ-free and normal gut in both mice and zebrafish, and 59 genes were identified with concordant changes in the two species. One gene that was potentially suppressed by the presence of gut microbiota in both mice and fish is *Fiaf*, an inhibitor of lipoprotein-lipase, a gene important for fat accumulation. Reconstitution of gut microbiota increased body fat by more than 50% in germ-free mice. However, inactivation of *Fiaf* by gene targeting eliminated the normal microbiota-induced deposition of fat, demonstrating an important role for this protein in microbe-induced fat deposition. The mechanism by which gut microbial flora suppress *Fiaf* expression is unknown (Backhed et al. 2004).

Thus, it appears that the human gut exhibits a true symbiosis in which the microbial community enjoys a stable, nutrient-rich environment with a limited host immune response, and in turn, the microbes selected to exist in this environment facilitate normal gut function. Unfortunately, although 16S rRNA gene surveys have successfully defined microbial diversity in the gut, these studies provide little functional information to help us understand specific microbial functions relevant to human health. Dissecting these functions is complicated for several reasons. First, microbial diversity within and between individuals makes complete description of the community difficult. A complete description is important because minor species may provide important functions. Second, the number of sequenced genomes from the microbial community relative to the total number of species is relatively limited, and third, genomic tools for predicting novel functions of complex microbial communities have not existed.

An alternative strategy for understanding potential functions of microbial communities that circumvents these problems was recently reported by Tringe et al. (2005) for analysis of several environmental microbial communities (and Fig. 2). This analysis represents the microbial community not from the viewpoint of the individual organisms present but rather from the collective gene content of the constituent organisms. For this analysis, small-insert libraries were created for soil and several deep-ocean whale carcasses and approximately 25–100 Mb of sequence was generated from each library. No attempt was made to assemble the sequencing reads (each ~700 bp in length). Instead, the reads (referred to as “environmental gene tags” or EGTs) were individually annotated and when possible a putative gene function was assigned to each read based on its best BLAST

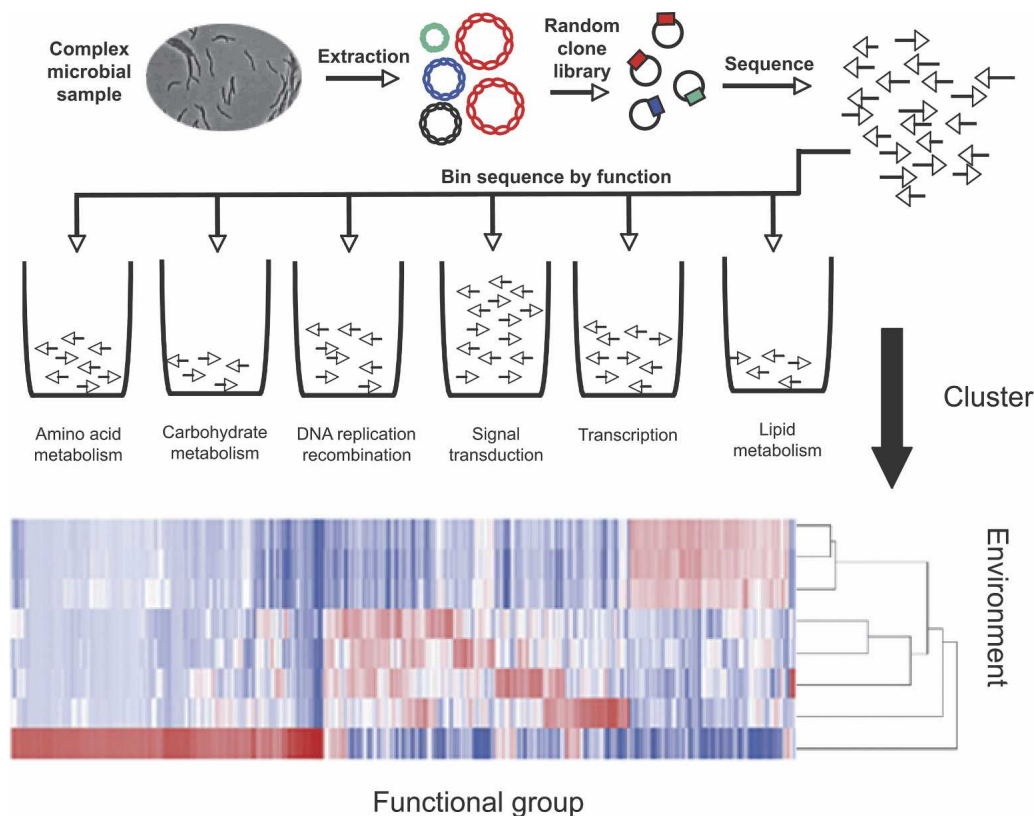


Figure 2. Environmental genomic tags for functional analysis of complex microbial communities. Genomic DNA is extracted from the community, sheared, and cloned into a standard sequencing vector. Clones are sequenced and function assigned to individual reads based on highly significant BLAST hits. These reads are then binned by function, and the relative abundance of reads in each functional category can be clustered and displayed graphically so that multiple environments can be compared. The clustering algorithm also computes and displays branch lengths that represent relatedness of individual samples.

hit. EGTs were then binned by function and the prevalence of specific functions was compared between four environments: soil, whale falls, and two previously reported environmental samples, a biofilm obtained from acid mine runoff (Tyson et al. 2004) and seawater collected from the Sargasso Sea (Venter et al. 2004). Trees constructed from these EGTs showed that samples obtained from similar environments clustered together. Furthermore, analysis of EGTs overrepresented in specific environments indicates they perform functions important for survival in that environment (e.g., sodium transporters in seawater). It is easy to imagine how such an analysis could be used to understand the function of microbial community in the gut and other human environments. In a fashion analogous to expression profiling of mammalian tissues, EGT fingerprints of microbial communities from diseased and normal individuals might be compared, or those from single individuals subjected to different diets.

With expanding sequence capacity and new technologies like EGT fingerprints, the value of studying microbial communities in genetically tractable model organisms seems clear. Very recently, Ley et al. (2005) reported the finding of the cecal microbiota composition change in genetically obese (*ob/ob*) mice using 16S rRNA sequence analysis. The obese mice, which carry a homozygous mutation in the leptin gene (*ob/ob*) locus, displayed a statistically significant reduction in Bacteroidetes when compared with their lean (*ob/+* or *+/+*) siblings. This result suggested that obesity may affect the diversity of the gut microbiota at

division level in nearly isogenic mice. However, the mechanism for such changes remains unknown. Comparing gene content of gut microbes in *ob/ob* mice and lean siblings through a metagenomic approach might provide novel insights in understanding how changes of community structure at division level may affect energy homeostasis. While the study of Tringe et al. (2005) compared quite different microbial communities, we expect that the number of EGTs altered in congenic mice, raised together and eating the same diet but differing in the presence or absence of a single gene, should be much more limited and could be readily related to phenotypic changes in the host.

Conclusions and future directions

Culture-independent methods, while largely developed for analysis of environmental microbes, have found broad utility in human ecology and have led to an appreciation for the diversity of microbial communities inhabiting normal humans. Quantitative PCR methods show particular promise for providing rapid identification of human pathogens that may allow clinicians to narrow and limit antibiotic use, which could in turn limit the spread of antibiotic resistance. As DNA-sequencing capacity grows and costs fall, sequence-based methods of analysis will be expanded to provide snapshots of the microbes present in many body fluids and tissues and the functions encoded in their genes. Specifically, we predict that expansion of EGT-based methods to complex microbial communities in the mouth, gut, skin, and

vagina may lead to an understanding of their role in human health and disease.

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological, and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098, and Los Alamos National Laboratory under contract No. W-7405-ENG-36 and was supported by NIH-NHLBI THL007279F. We thank Phil Hugenholz, Susannah Tringe, Tanja Woyke, and members of the Rubin laboratory for their critical reading of the manuscript.

References

- Anderson, B.E., Dawson, J.E., Jones, D.C., and Wilson, K.H. 1991. *Ehrlichia chaffeensis*, a new species associated with human ehrlichiosis. *J. Clin. Microbiol.* **29**: 2838–2842.
- Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H.W., Neefs, J.M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., et al. 2005. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**: 223–227.
- Backhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A., Semenkovich, C.F., and Gordon, J.I. 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci.* **101**: 15718–15723.
- Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. 2005. Host-bacterial mutualism in the human intestine. *Science* **307**: 1915–1920.
- Bentley, S.D., Maiwald, M., Murphy, L.D., Pallen, M.J., Yeats, C.A., Dover, L.G., Norbertczak, H.T., Besra, G.S., Quail, M.A., Harris, D.E., et al. 2003. Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whippelii*. *Lancet* **361**: 637–644.
- Brinig, M.M., Lepp, P.W., Ouverney, C.C., Armitage, G.C., and Relman, D.A. 2003. Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. *Appl. Environ. Microbiol.* **69**: 1687–1694.
- Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J., Knowles, D.M., and Moore, P.S. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**: 1865–1869.
- Cheers Jr., W.C. and Ashworth, C.T. 1961. Electron microscopic study of the intestinal mucosa in Whipple's disease. Demonstration of encapsulated bacilliform bodies in the lesion. *Gastroenterology* **41**: 129–138.
- Choi, B.K., Paster, B.J., Dewhirst, F.E., and Gobel, U.B. 1994. Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. *Infect. Immun.* **62**: 1889–1895.
- Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**: 359–362.
- Cohen, M.L. 2000. Changing patterns of infectious disease. *Nature* **406**: 762–767.
- Connon, S.A. and Giovannoni, S.J. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* **68**: 3878–3885.
- Dawson, J.E., Anderson, B.E., Fishbein, D.B., Sanchez, J.L., Goldsmith, C.S., Wilson, K.H., and Duntley, C.W. 1991. Isolation and characterization of an *Ehrlichia* sp. from a patient diagnosed with human ehrlichiosis. *J. Clin. Microbiol.* **29**: 2741–2745.
- Dewhirst, F.E., Paster, B.J., Tzellas, N., Coleman, B., Downes, J., Spratt, D.A., and Wade, W.G. 2001. Characterization of novel human oral isolates and cloned 16S rDNA sequences that fall in the family *Coriobacteriaceae*: Description of *olsenella* gen. nov., reclassification of *Lactobacillus uli* as *Olsenella uli* comb. nov. and description of *Olsenella profuse* sp. nov. *Int. J. Syst. Evol. Microbiol.* **51**: 1797–1804.
- Domann, E., Hong, G., Imirzalioglu, C., Turschner, S., Kuhle, J., Watzel, C., Hain, T., Hossain, H., and Chakraborty, T. 2003. Culture-independent identification of pathogenic bacteria and polymicrobial infections in the genitourinary tract of renal transplant recipients. *J. Clin. Microbiol.* **41**: 5500–5510.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. 2005. Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.
- Falk, P.G., Hooper, L.V., Midtvedt, T., and Gordon, J.I. 1998. Creating and maintaining the gastrointestinal ecosystem: What we know and need to know from gnotobiology. *Microbiol. Mol. Biol. Rev.* **62**: 1157–1170.
- Franks, A.H., Harmsen, H.J., Raangs, G.C., Jansen, G.J., Schut, F., and Welling, G.W. 1998. Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.* **64**: 3336–3345.
- Fredericks, D.N. and Relman, D.A. 1996. Sequence-based identification of microbial pathogens: A reconsideration of Koch's postulates. *Clin. Microbiol. Rev.* **9**: 18–33.
- Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60–63.
- Guarner, F. and Malagelada, J.R. 2003. Gut flora in health and disease. *Lancet* **361**: 512–519.
- Gupta, R., Lanter, J.M., and Woese, C.R. 1983. Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an Archaeobacterium. *Science* **221**: 656–659.
- Guschin, D.Y., Mobarry, B.K., Proudnikov, D., Stahl, D.A., Rittmann, B.E., and Mirzabekov, A.D. 1997. Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl. Environ. Microbiol.* **63**: 2397–2402.
- Handelsman, J. 2004. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**: 669–685.
- Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63–67.
- Hugenholz, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3**: reviews0003.
- Hugenholz, P., Goebel, B.M., and Pace, N.R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**: 4765–4774.
- Hutter, G., Schlagenhauf, U., Valenza, G., Horn, M., Burgemeister, S., Claus, H., and Vogel, U. 2003. Molecular analysis of bacteria in periodontitis: Evaluation of clone libraries, novel phylotypes and putative pathogens. *Microbiology* **149**: 67–75.
- Ishii, K. and Fukui, M. 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl. Environ. Microbiol.* **67**: 3753–3755.
- Kolenbrander, P.E., Andersen, R.N., and Moore, L.V. 1990. Intrageneric coaggregation among strains of human oral bacteria: Potential role in primary colonization of the tooth surface. *Appl. Environ. Microbiol.* **56**: 3890–3894.
- Kroes, I., Lepp, P.W., and Relman, D.A. 1999. Bacterial diversity within the human subgingival crevice. *Proc. Natl. Acad. Sci.* **96**: 14547–14552.
- Ksiazek, T.G., Erdman, D., Goldsmith, D.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**: 1953–1966.
- Kulik, E.M., Sandmeier, H., Hinni, K., and Meyer, J. 2001. Identification of archaeal rDNA from subgingival dental plaque by PCR amplification and sequence analysis. *FEMS Microbiol. Lett.* **196**: 129–133.
- Langendijk, P.S., Schut, F., Jansen, G.J., Raangs, G.C., Kamphuis, G.R., Wilkinson, M.H., and Welling, G.W. 1995. Quantitative fluorescence in situ hybridization of *Bifidobacterium* spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples. *Appl. Environ. Microbiol.* **61**: 3069–3075.
- Lepp, P.W., Brinig, M.M., Ouverney, C.C., Palm, K., Armitage, G.C., and Relman, D.A. 2004. Methanogenic Archaea and human periodontal disease. *Proc. Natl. Acad. Sci.* **101**: 6176–6181.
- Ley, R.E., Backhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. 2005. Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci.* **102**: 11070–11075.
- Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., Schleifer, K.H., and Wagner, M. 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol.* **68**: 5064–5081.
- Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoecker, K., Baranyi, C., Lehner, A., and Wagner, M. 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the β -proteobacterial order "Rhodocyclales". *Appl. Environ. Microbiol.* **71**: 1373–1386.

- Maeda, K., Markowitz, N., Hawley, R.C., Ristic, M., Cox, D., and McDade, J.E. 1987. Human infection with *Ehrlichia canis*, a leukocytic rickettsia. *N. Engl. J. Med.* **316**: 853–856.
- Meyer, D.H. and Fives-Taylor, P.M. 1998. Oral pathogens: From dental plaque to cardiac disease. *Curr. Opin. Microbiol.* **1**: 88–95.
- Moumille, K., Carbonnelle, E., Dessemme, P., Tamisier, D., Iserin, F., Houdouin, V., Nassif, X., and Berche, P. 2004. Culture-negative pericarditis caused by *Neisseria meningitidis* serogroup C. *J. Clin. Microbiol.* **42**: 923–924.
- Neu, H.C. 1992. The crisis in antibiotic resistance. *Science* **257**: 1064–1073.
- Nikkari, S., Lopez, F.A., Lepp, P.W., Cieslak, P.R., Ladd-Wilson, S., Passaro, D., Danila, R., and Relman, D.A. 2002. Broad-range bacterial detection and the analysis of unexplained death and critical illness. *Emerg. Infect. Dis.* **8**: 188–194.
- Ouverney, C.C., Armitage, G.C., and Relman, D.A. 2003. Single-cell enumeration of an uncultivated TM7 subgroup in the human subgingival crevice. *Appl. Environ. Microbiol.* **69**: 6294–6298.
- Paster, B.J., Boches, S.K., Galvin, J.L., Ericson, R.E., Lau, C.N., Levanos, V.A., Sahasrabudhe, A., and Dewhirst, F.E. 2001. Bacterial diversity in human subgingival plaque. *J. Bacteriol.* **183**: 3770–3783.
- Perkocha, L.A., Geaghan, S.M., Yen, T.S., Nishimura, S.L., Chan, S.P., Garcia-Kennedy, R., Honda, G., Stoloff, A.C., Klein, H.Z., Goldman, R.L., et al. 1990. Clinical and pathological features of bacillary peliosis hepatis in association with human immunodeficiency virus infection. *N. Engl. J. Med.* **323**: 1581–1586.
- Qin, X. and Urdahl, K.B. 2001. PCR and sequencing of independent genetic targets for the diagnosis of culture negative bacterial endocarditis. *Diagn. Microbiol. Infect. Dis.* **40**: 145–149.
- Raoult, D., Birg, M.L., La Scola, B., Fournier, P.E., Enea, M., Lepidi, H., Roux, V., Piette, J.C., Vandenesch, F., Vital-Durand, D., et al. 2000. Cultivation of the bacillus of Whipple's disease. *N. Engl. J. Med.* **342**: 620–625.
- Raoult, D., Ogata, H., Audic, S., Robert, C., Suhre, K., Drancourt, M., and Claverie, J.M. 2003. *Tropheryma whippelii* Twist: A human pathogenic Actinobacteria with a reduced genome. *Genome Res.* **13**: 1800–1809.
- Rappe, M.S. and Giovannoni, S.J. 2003. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**: 369–394.
- Rawls, J.F., Samuel, B.S., and Gordon, J.I. 2004. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proc. Natl. Acad. Sci.* **101**: 4596–4601.
- Regnery, R.L., Anderson, B.A., Clarridge III, J.E., Rodriguez-Barradas, M.C., Jones, D.C., and Carr, J.H. 1992. Characterization of a novel *Rochalimaea* species, *R. henselae* sp. nov., isolated from blood of a febrile, human immunodeficiency virus-positive patient. *J. Clin. Microbiol.* **30**: 265–274.
- Relman, D.A. 2002. New technologies, human-microbe interactions, and the search for previously unrecognized pathogens. *J. Infect. Dis.* **186** Suppl 2: S254–S258.
- Relman, D.A. and Falkow, S. 2001. The meaning and impact of the human genome sequence for microbiology. *Trends Microbiol.* **9**: 206–208.
- Relman, D.A., Loutit, J.S., Schmidt, T.M., Falkow, S., and Tompkins, L.S. 1990. The agent of bacillary angiomatosis. An approach to the identification of uncultured pathogens. *N. Engl. J. Med.* **323**: 1573–1580.
- Relman, D.A., Schmidt, T.M., MacDermott, R.P., and Falkow, S. 1992. Identification of the uncultured bacillus of Whipple's disease. *N. Engl. J. Med.* **327**: 293–301.
- Renesto, P., Crapoulet, N., Ogata, H., La Scola, B., Vestris, G., Claverie, J.M., and Raoult, D. 2003. Genome-based design of a cell-free culture medium for *Tropheryma whippelii*. *Lancet* **362**: 447–449.
- Rudi, K., Skulberg, O.M., Skulberg, R., and Jakobsen, K.S. 2000. Application of sequence-specific labeled 16S rRNA gene oligonucleotide probes for genetic profiling of cyanobacterial abundance and diversity by array hybridization. *Appl. Environ. Microbiol.* **66**: 4004–4011.
- Schmidt, T.M., DeLong, E.F., and Pace, N.R. 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**: 4371–4378.
- Schuurman, T., de Boer, R.F., Kooistra-Smid, A.M., and van Zwet, A.A. 2004. Prospective study of use of PCR amplification and sequencing of 16S ribosomal DNA from cerebrospinal fluid for diagnosis of bacterial meningitis in a clinical setting. *J. Clin. Microbiol.* **42**: 734–740.
- Slater, L.N., Welch, D.F., Hensel, D., and Coody, D.W. 1990. A newly recognized fastidious gram-negative pathogen as a cause of fever and bacteremia. *N. Engl. J. Med.* **323**: 1587–1593.
- Small, J., Call, D.R., Brockman, F.J., Straub, T.M., and Chandler, D.P. 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl. Environ. Microbiol.* **67**: 4708–4716.
- Stahl, D.A., Lane, D.J., Olsen, G.J., and Pace, N.R. 1984. Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science* **224**: 409–411.
- . 1985. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl. Environ. Microbiol.* **49**: 1379–1384.
- Suau, A., Bonnet, R., Sutren, M., Godon, J.J., Gibson, G.R., Collins, M.D., and Dore, J. 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.
- Suzuki, K., Meek, B., Doi, Y., Muramatsu, M., Chiba, T., Honjo, T., and Fagarasan, S. 2004. Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc. Natl. Acad. Sci.* **101**: 1981–1986.
- Thoden van Velzen, S.K., Abraham-Inpijn, L., and Moorer, W.R. 1984. Plaque and systemic disease: A reappraisal of the focal infection concept. *J. Clin. Periodontol.* **11**: 209–220.
- Trier, J.S., Phelps, P.C., Eidelman, S., and Rubin, C.E. 1965. Whipple's Disease: Light and electron microscope correlation of jejunal mucosal histology with antibiotic treatment and clinical status. *Gastroenterology* **48**: 684–707.
- Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., et al. 2005. Comparative metagenomics of microbial communities. *Science* **308**: 554–557.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Wang, G.C. and Wang, Y. 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* **63**: 4645–4650.
- Ward, D.M., Weller, R., and Bateson, M.M. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**: 63–65.
- Whipple, G.H. 1907. A hitherto undescribed disease characterized anatomically by deposits of fat and fatty acids in the intestinal and mesenteric lymphatic tissues. *Johns Hopkins Hosp. Bull.* **18**: 382–391.
- Whittaker, C.J., Klier, C.M., and Kolenbrander, P.E. 1996. Mechanisms of adhesion by oral bacteria. *Annu. Rev. Microbiol.* **50**: 513–552.
- Wilson, K.H. and Blichington, R.B. 1996. Human colonic biota studied by ribosomal DNA sequence analysis. *Appl. Environ. Microbiol.* **62**: 2273–2278.
- Wilson, K.H., Blichington, R., Frothingham, R., and Wilson, J.A. 1991. Phylogeny of the Whipple's-disease-associated bacterium. *Lancet* **338**: 474–475.
- Woese, C.R. 1982. Archaeobacteria and cellular origins: An overview. *Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. I. Orig. Reihe C* **3**: 1–17.
- Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* **74**: 5088–5090.
- Woese, C.R. and Olsen, G.J. 1986. Archaeobacterial phylogeny: Perspectives on the urkingdoms. *Syst. Appl. Microbiol.* **7**: 161–177.
- Yardley, J.H. and Hendrix, T.R. 1961. Combined electron and light microscopy in Whipple's disease. Demonstration of "bacillary bodies" in the intestine. *Bull. Johns Hopkins Hosp.* **109**: 80–98.
- Zoetendal, E.G., Akkermans, A.D., and De Vos, W.M. 1998. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl. Environ. Microbiol.* **64**: 3854–3859.
- Zoetendal, E.G., Akkermans, A.D., Akkermans-van Vliet, W.M., De Visser, J.A.G.M., and De Vos, W.M. 2001. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microb. Ecol. Health Dis.* **13**: 129–134.

Received January 10, 2005; accepted in revised form October 23, 2005.