Application of Sequence-based Methods in Human Microbial Ecology

Li Weng, Edward M. Rubin, and James Bristow*

Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598

Address correspondence to:
James Bristow
Joint Genome Institute
2800 Mitchell Dr.
Walnut Creek, CA 94598
(925) 296-5609 office
(415) 378-8643 cell
jbristow@LBL.gov

Running title: "Human microbial ecology"

Abstract

Ecologists studying microbial life in the environment have recognized the enormous complexity of microbial diversity for many years (Woese 1987), 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 (Handelsman 2004; Harris et al. 2004; Hugenholtz et al. 1998; Moreira and Lopez-Garcia 2002; Rappe and Giovannoni 2003). 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 behind. 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" (Relman and Falkow 2001).

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 to 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. 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 (Gupta et al. 1983; Woese and Fox 1977; Woese and Olsen 1986), but this method was subsequently applied to libraries of rRNA genes PCR-amplified from environmental DNA samples without cultivation (Giovannoni et al. 1990; Schmidt et al. 1991; Stahl et al. 1984; Stahl et al. 1985; Ward et al. 1990). 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 (Rappe and Giovannoni 2003). To maximize the utility of 16S rRNA gene analysis for species determination, it is now routine to amplify the entire 16S rRNA gene, which spans ~1500 bp, and is thus readily sequenced in its entirety through bi-directional 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).

While 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 not amplify at all or do so less efficiently, so that the representation in cloned libraries of amplified 16S rRNA genes may be skewed, especially if the number of 16S rRNA

sequences sampled is small (Kroes et al. 1999). Such 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 (Amann et al. 1990; Amann et al. 1995;

Bosshard et al. 2000; DeLong et al. 1989; Giovannoni et al. 1988).

A second 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 to

hybridization-based methods. As an alternative, several strategies employing 16S rRNA gene

microarrays have been presented and offer some advantage in speed compared to sequencing when

analysis of many samples is required (Guschin et al. 1997; Loy et al. 2002; Loy et al. 2005; Rudi et al.

2000; Small et al. 2001). 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 differentiating related species, but it seems reasonable to expect

further improvement in this technology and eventual application to clinical materials from humans.

Identification of Novel Human Pathogens

An exciting application of culture-independent methods is the identification of uncultivated

organisms causing human disease. Because DNA can be extracted from any potentially infected

material and used as a substrate for 16S rRNA gene amplification, Relman predicted in 1996 that a rash

of claims for disease causation by new pathogens would follow application of this method to human

tissues, and laid out a clear strategy for proving disease causation for organisms that might be difficult or

impossible to grow (Fredericks and Relman 1996). 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 due to their fastidious growth requirements (Perkocha et al. 1990). In 1990, Relman et al. amplified a partial 16S r RNA 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, *Rochalimea*-like bacillius 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 potential infectious diseases and led to the identification of *Ehrlichia chaffeensis*, a new species associated with tick bites and causing a febrile illness. Ehrlichiosis is clinically similar to Rocky Mountain spotted fever, another tick-borne disease caused by the intercellular parasite, *Rickettsia rickettsii*. While 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, George Hoyt noted "rodlike bacilli in a small node" (Whipple 1907). Eighty-four years passed before the identification of the etiologic agent despite its consistent observation in affected tissues (Chears 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. whippleii* predicted deficiencies in amino acid synthesis (Bentley et al. 2003; Raoult et al. 2003), and with this information, Renesto et al. successfully designed a complete medium that allowed cell-free cultivation of *T. whippleii* (Renesto et al. 2003). This was the first demonstration that genomic information could guide rational design of media for axenic cultivation of fastidious bacteria.

While most human tissues are normally sterile, 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 the 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 has been studied intensively through both culture-dependent and culture-independent techniques. About 500 bacterial species have been found in the human oral cavity (Meyer and Fives-Taylor 1998; Paster et al. 2001; Thoden van Velzen et al. 1984) and 40-60% of these species are uncultivated "phylotypes" (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 andTreponema 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(Brinig et al. 2003; Choi et al. 1994; Dewhirst et al. 2001; Hutter et al. 2003; Ouverney et al. 2003; Paster et al. 2001).

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 Lepp's 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 in question the nature of their role in this polymicrobial disease. Interestingly, the abundance of archaea and *Treponema 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

While 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 hours for growth and identification of most bacterial species. Clinical practice has long favored the use of antibiotics to cover the most likely organisms to be present when infection is suspected, but this practice has contributed greatly to the spread of antibiotic resistance. 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 body compartments that are normally sterile 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%. (Domann et al. 2003; Moumile et al. 2004; Nikkari et al. 2002; Qin and Urdahl 2001; 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 and yet some 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 (Suau et al. 1999; Wilson and Blitchington 1996). 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 less diversity in 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 to wildtype littermates, 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 (Backhed et al. 2004; Backhed et al. 2005; Falk et al. 1998; Guarner and Malagelada 2003; Rawls et al. 2004). For example, in the recent study by Rawls et al., gene expression patterns were compared between germ-free and normal gut in both mice and zebrafish, identifying 59 genes with concordant changes in the two species. One gene that was potently suppressed by the presence of gut microbiota in both mice and fish is *Fiaf,* an inhibitor of lipoprotin-lipase, a gene important for fat accumulation. Reconstitution of gut microbiata 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, while 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 the functions is complicated for several

reasons. First, the 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, tools

for analyzing the 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. for analysis of several environmental

microbial communities (Tringe et al. 2005) (and Fig. 2). This analysis represents the community not

from the viewpoint of the individual organisms present, but rather from the collective gene content of

the constituent organisms. For their 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 assigned to each read based on its best BLAST hit. EGTs were then binned by

function and the prevalence of specific functions 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). Phylogenetic trees constructed from these EGT fingerprints showed that samples obtained from similar environments clustered together. Furthermore, analysis of EGTs over-represented in specific environments 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, one might image the value of studying microbial communities in genetically tractable model organisms.   For example, examination of the EGT profiles of gut microbes in mice carrying genetic modifications that alter the immune system or gut development, might provide novel insights into the host-microbe interaction. While the study of Tringe et al 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.

Conclusion 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 snap-

shots 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.

Reference:

Amann, R.I., L. Krumholz, and D.A. Stahl. 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* **172:** 762-770.

Amann, R.I., W. Ludwig, and K.H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143-169.

Anderson, B.E., J.E. Dawson, D.C. Jones, and K.H. Wilson. 1991. Ehrlichia chaffeensis, a new species associated with human ehrlichiosis. *J Clin Microbiol* **29:** 2838-2842.

Andries, K., P. Verhasselt, J. Guillemont, H.W. Gohlmann, J.M. Neefs, H. Winkler, J. Van Gestel, P. Timmerman, M. Zhu, E. Lee, P. Williams, D. de Chaffoy, E. Huitric, S. Hoffner, E. Cambau, C. Truffot-Pernot, N. Lounis, and V. Jarlier. 2005. A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science* **307:** 223-227.

Backhed, F., H. Ding, T. Wang, L.V. Hooper, G.Y. Koh, A. Nagy, C.F. Semenkovich, and J.I. Gordon. 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* **101:** 15718-15723.

Backhed, F., R.E. Ley, J.L. Sonnenburg, D.A. Peterson, and J.I. Gordon. 2005. Host-bacterial mutualism in the human intestine. *Science* **307:** 1915-1920.

Bentley, S.D., M. Maiwald, L.D. Murphy, M.J. Pallen, C.A. Yeats, L.G. Dover, H.T. Norbertczak, G.S. Besra, M.A. Quail, D.E. Harris, A. von Herbay, A. Goble, S. Rutter, R. Squares, S. Squares, B.G. Barrell, J. Parkhill, and D.A. Relman. 2003. Sequencing and analysis of the genome of the Whipple's disease bacterium Tropheryma whipplei. *Lancet* **361:** 637-644.

Bosshard, P.P., Y. Santini, D. Gruter, R. Stettler, and R. Bachofen. 2000. Bacterial diversity and community composition in the chemocline of the meromictic alpine Lake Cadagno as revealed by 16S rDNA analysis. *FEMS Microbiol Ecol* **31:** 173-182.

Brinig, M.M., P.W. Lepp, C.C. Ouverney, G.C. Armitage, and D.A. Relman. 2003. Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. *Appl Environ Microbiol* **69:** 1687-1694.

Chang, Y., E. Cesarman, M.S. Pessin, F. Lee, J. Culpepper, D.M. Knowles, and P.S. Moore. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266:** 1865-1869.

Chears, W.C., Jr. and C.T. Ashworth. 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., B.J. Paster, F.E. Dewhirst, and U.B. Gobel. 1994. Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. *Infect Immun* **62:** 1889-1895.

Choo, Q.L., G. Kuo, A.J. Weiner, L.R. Overby, D.W. Bradley, and M. Houghton. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244:** 359-362.

Dawson, J.E., B.E. Anderson, D.B. Fishbein, J.L. Sanchez, C.S. Goldsmith, K.H. Wilson, and C.W. Duntley. 1991. Isolation and characterization of an Ehrlichia sp. from a patient diagnosed with human ehrlichiosis. *J Clin Microbiol* **29:** 2741-2745.

DeLong, E.F., G.S. Wickham, and N.R. Pace. 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243:** 1360-1363.

Dewhirst, F.E., B.J. Paster, N. Tzellas, B. Coleman, J. Downes, D.A. Spratt, and W.G. Wade. 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 profusa sp. nov. *Int J Syst Evol Microbiol* **51:** 1797-1804.

Domann, E., G. Hong, C. Imirzalioglu, S. Turschner, J. Kuhle, C. Watzel, T. Hain, H. Hossain, and T. Chakraborty. 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., E.M. Bik, C.N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S.R. Gill, K.E. Nelson, and D.A. Relman. 2005. Diversity of the Human Intestinal Microbial Flora. *Science*.

Falk, P.G., L.V. Hooper, T. Midtvedt, and J.I. Gordon. 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., H.J. Harmsen, G.C. Raangs, G.J. Jansen, F. Schut, and G.W. Welling. 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 D.A. Relman. 1996. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* **9:** 18-33.

Giovannoni, S.J., T.B. Britschgi, C.L. Moyer, and K.G. Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345:** 60-63.

Giovannoni, S.J., E.F. DeLong, G.J. Olsen, and N.R. Pace. 1988. Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J Bacteriol* **170:** 720-726.

Guarner, F. and J.R. Malagelada. 2003. Gut flora in health and disease. *Lancet* **361:** 512-519.

Gupta, R., J.M. Lanter, and C.R. Woese. 1983. Sequence of the 16S Ribosomal RNA from Halobacterium volcanii, an Archaebacterium. *Science* **221:** 656-659.

Guschin, D.Y., B.K. Mobarry, D. Proudnikov, D.A. Stahl, B.E. Rittmann, and A.D. Mirzabekov. 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.

Harris, J.K., S.T. Kelley, and N.R. Pace. 2004. New perspective on uncultured bacterial phylogenetic division OP11. *Appl Environ Microbiol* **70:** 845-849.

Hugenholtz, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3:** REVIEWS0003.

Hugenholtz, P., B.M. Goebel, and N.R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180:** 4765-4774.

Hutter, G., U. Schlagenhauf, G. Valenza, M. Horn, S. Burgemeister, H. Claus, and U. Vogel. 2003. Molecular analysis of bacteria in periodontitis: evaluation of clone libraries, novel phylotypes and putative pathogens. *Microbiology* **149:** 67-75.

Kolenbrander, P.E., R.N. Andersen, and L.V. Moore. 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., P.W. Lepp, and D.A. Relman. 1999. Bacterial diversity within the human subgingival crevice. *Proc Natl Acad Sci U S A* **96:** 14547-14552.

Ksiazek, T.G., D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, P.E. Rollin, S.F. Dowell, A.E. Ling, C.D. Humphrey, W.J. Shieh, J. Guarner, C.D. Paddock, P. Rota, B. Fields, J. DeRisi, J.Y. Yang, N. Cox, J.M. Hughes, J.W. LeDuc, W.J. Bellini, and L.J. Anderson. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348:** 1953-1966.

Kulik, E.M., H. Sandmeier, K. Hinni, and J. Meyer. 2001. Identification of archaeal rDNA from subgingival dental plaque by PCR amplification and sequence analysis. *FEMS Microbiol Lett* **196:** 129-133.

Langendijk, P.S., F. Schut, G.J. Jansen, G.C. Raangs, G.R. Kamphuis, M.H. Wilkinson, and G.W. Welling. 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., M.M. Brinig, C.C. Ouverney, K. Palm, G.C. Armitage, and D.A. Relman. 2004. Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A* **101:** 6176-6181.

Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K.H. Schleifer, and M. Wagner. 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., C. Schulz, S. Lucker, A. Schopfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner, and M. Wagner. 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". *Appl Environ Microbiol* **71:** 1373-1386.

Maeda, K., N. Markowitz, R.C. Hawley, M. Ristic, D. Cox, and J.E. McDade. 1987. Human infection with Ehrlichia canis, a leukocytic rickettsia. *N Engl J Med* **316:** 853-856.

Meyer, D.H. and P.M. Fives-Taylor. 1998. Oral pathogens: from dental plaque to cardiac disease. *Curr Opin Microbiol* **1:** 88-95.

Moreira, D. and P. Lopez-Garcia. 2002. The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol* **10:** 31-38.

Moumile, K., E. Carbonnelle, P. Dessemme, D. Tamisier, F. Iserin, V. Houdouin, X. Nassif, and P. Berche. 2004. Culture-negative pericarditis caused by Neisseria meningitidis serogroup C. *J Clin Microbiol* **42:** 923-924.

Nikkari, S., F.A. Lopez, P.W. Lepp, P.R. Cieslak, S. Ladd-Wilson, D. Passaro, R. Danila, and D.A. Relman. 2002. Broad-range bacterial detection and the analysis of unexplained death and critical illness. *Emerg Infect Dis* **8:** 188-194.

Ouverney, C.C., G.C. Armitage, and D.A. Relman. 2003. Single-cell enumeration of an uncultivated TM7 subgroup in the human subgingival crevice. *Appl Environ Microbiol* **69:** 6294-6298.

Paster, B.J., S.K. Boches, J.L. Galvin, R.E. Ericson, C.N. Lau, V.A. Levanos, A. Sahasrabudhe, and F.E. Dewhirst. 2001. Bacterial diversity in human subgingival plaque. *J Bacteriol* **183:** 3770-3783.

Perkocha, L.A., S.M. Geaghan, T.S. Yen, S.L. Nishimura, S.P. Chan, R. Garcia-Kennedy, G. Honda, A.C. Stoloff, H.Z. Klein, R.L. Goldman, and 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 K.B. Urdahl. 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., M.L. Birg, B. La Scola, P.E. Fournier, M. Enea, H. Lepidi, V. Roux, J.C. Piette, F. Vandenesch, D. Vital-Durand, and T.J. Marrie. 2000. Cultivation of the bacillus of Whipple's disease. *N Engl J Med* **342:** 620-625.

Raoult, D., H. Ogata, S. Audic, C. Robert, K. Suhre, M. Drancourt, and J.M. Claverie. 2003. Tropheryma whipplei Twist: a human pathogenic Actinobacteria with a reduced genome. *Genome Res* **13:** 1800-1809.

Rappe, M.S. and S.J. Giovannoni. 2003. The uncultured microbial majority. *Annu Rev Microbiol* **57:** 369-394.

Rawls, J.F., B.S. Samuel, and J.I. Gordon. 2004. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proc Natl Acad Sci U S A* **101:** 4596-4601.

Regnery, R.L., B.E. Anderson, J.E. Clarridge, 3rd, M.C. Rodriguez-Barradas, D.C. Jones, and J.H. Carr. 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:** S254-258.

Relman, D.A. and S. Falkow. 2001. The meaning and impact of the human genome sequence for microbiology. *Trends Microbiol* **9:** 206-208.

Relman, D.A., J.S. Loutit, T.M. Schmidt, S. Falkow, and L.S. Tompkins. 1990. The agent of bacillary angiomatosis. An approach to the identification of uncultured pathogens. *N Engl J Med* **323:** 1573-1580.

Relman, D.A., T.M. Schmidt, R.P. MacDermott, and S. Falkow. 1992. Identification of the uncultured bacillus of Whipple's disease. *N Engl J Med* **327:** 293-301.

Renesto, P., N. Crapoulet, H. Ogata, B. La Scola, G. Vestris, J.M. Claverie, and D. Raoult. 2003. Genome-based design of a cell-free culture medium for Tropheryma whipplei. *Lancet* **362:** 447-449.

Rudi, K., O.M. Skulberg, R. Skulberg, and K.S. Jakobsen. 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., E.F. DeLong, and N.R. Pace. 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* **173:** 4371-4378.

Schuurman, T., R.F. de Boer, A.M. Kooistra-Smid, and A.A. van Zwet. 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., D.F. Welch, D. Hensel, and D.W. Coody. 1990. A newly recognized fastidious gram-negative pathogen as a cause of fever and bacteremia. *N Engl J Med* **323:** 1587-1593.

Small, J., D.R. Call, F.J. Brockman, T.M. Straub, and D.P. Chandler. 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol* **67:** 4708-4716.

Stahl, D.A., D.J. Lane, G.J. Olsen, and N.R. Pace. 1984. Analysis of Hydrothermal Vent-Associated Symbionts by Ribosomal RNA Sequences. *Science* **224:** 409-411.

Stahl, D.A., D.J. Lane, G.J. Olsen, and N.R. Pace. 1985. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl Environ Microbiol* **49:** 1379-1384.

Suau, A., R. Bonnet, M. Sutren, J.J. Godon, G.R. Gibson, M.D. Collins, and J. Dore. 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., B. Meek, Y. Doi, M. Muramatsu, T. Chiba, T. Honjo, and S. Fagarasan. 2004. Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc Natl Acad Sci U S A* **101:** 1981-1986.

Thoden van Velzen, S.K., L. Abraham-Inpijn, and W.R. Moorer. 1984. Plaque and systemic disease: a reappraisal of the focal infection concept. *J Clin Periodontol* **11:** 209-220.

Trier, J.S., P.C. Phelps, S. Eidelman, and C.E. Rubin. 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., C. von Mering, A. Kobayashi, A.A. Salamov, K. Chen, H.W. Chang, M. Podar, J.M. Short, E.J. Mathur, J.C. Detter, P. Bork, P. Hugenholtz, and E.M. Rubin. 2005. Comparative metagenomics of microbial communities. *Science* **308:** 554-557.

Tyson, G.W., J. Chapman, P. Hugenholtz, E.E. Allen, R.J. Ram, P.M. Richardson, V.V. Solovyev, E.M. Rubin, D.S. Rokhsar, and J.F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37-43.

Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D. Wu, I. Paulsen, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.H. Rogers, and H.O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74.

Ward, D.M., R. Weller, and M.M. Bateson. 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., C.M. Klier, and P.E. Kolenbrander. 1996. Mechanisms of adhesion by oral bacteria. *Annu Rev Microbiol* **50:** 513-552.

Wilson, K.H., R. Blitchington, R. Frothingham, and J.A. Wilson. 1991. Phylogeny of the Whipple's-disease-associated bacterium. *Lancet* **338:** 474-475.

Wilson, K.H. and R.B. Blitchington. 1996. Human colonic biota studied by ribosomal DNA sequence analysis. *Appl Environ Microbiol* **62:** 2273-2278.

Woese, C.R. 1987. Bacterial evolution. *Microbiol Rev* **51:** 221-271.

Woese, C.R. and G.E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74:** 5088-5090.

Woese, C.R. and G.J. Olsen. 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. *Syst Appl Microbiol* **7:** 161-177.

Yardley, J.H. and T.R. Hendrix. 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., A.D. Akkermans, W.M. Akkermans-van Vliet, J.A.G.M. De Visser, and W.M. De Vos. 2001. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microb. Ecol. Health Dis.* **13:** 129-134.

Zoetendal, E.G., A.D. Akkermans, and W.M. De Vos. 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.

Figure Legends

Fig 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 1.5kb in length, are cloned into a standard vector and both ends sequenced. The aligned sequences are first clustered into groups, and the representatives from each cluster are compared to 16S rRNA gene databases for phylogenetic classification.

Fig 2. Environmental genomic tags for functional analysis of complex microbial communities. Genomomic 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.