Chapter 16
Implications of Human Microbiome Research for the Developing World

Appolinaire Djikeng, Barbara Jones Nelson, and Karen E. Nelson

Introduction

New high-throughput sequencing and data analysis approaches (Costello et al., 2009; Turnbaugh et al., 2009), along with novel diversity screens and even more intrinsic single cell approaches to isolating new species (Lasken, 2009), have presented the sciences with a unique opportunity to investigate and interrogate the microorganisms that are associated with the human body, all at a greater depth than previously appreciated.

From the earliest studies, the greater scientific community has recognized a high level of microbial diversity in nature beyond imaginations. This includes observations on the oceans, soils, and on animals. With respect to humans, it became increasingly apparent that the species on and in our bodies make significant contributions to our health and development. These species maintain normal cell function in the gastrointestinal tract (for example, see Eckburg et al., 2005; Bik et al., 2006; Gill et al., 2006). We are also dependent on these species for the efficient digestion of food components, such as plant material and xenobiotics (Gill et al., 2006), and to ward off certain diseases. In parallel, these microbes have been associated with, and can result in, many common diseases such as cavities, stomach ulcers, bacterial vaginosis (BV), and irritable bowel syndrome (Foster et al., 2008; Dorer et al., 2009).

Most of the initial studies on the microbial diversity associated with the human body focused either on traditional culturing approaches or on sequencing and phylogenetic analysis of the 16S ribosomal (r)RNA genes derived from microbial samples taken from the human body (Eckburg et al., 2005; Bik et al., 2006). The limit to culturing or 16S rRNA sequencing was primarily a reflection of the availability of molecular tools and approaches, and the cost associated with earlier versions of available sequencing technologies. The 16S rRNA sequencing and analysis invariably revealed a higher level of microbial diversity than that seen with conventional

K.E. Nelson (✉)
Department of Human Genomic Medicine, The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
e-mail: kenelson@jcvi.org

K.E. Nelson (ed.), *Metagenomics of the Human Body*,
DOI 10.1007/978-1-4419-7089-3_16, © Springer Science+Business Media, LLC 2011

culture techniques (Gao et al., 2007; Gao et al., 2008). From the human stomach, for example, although the highly acidic environment was thought to only contain *Helicobacter* types, Bik et al. (2006) used 1,833 16S rRNA sequences obtained from 23 gastric endoscopic biopsy samples to identify 128 phylotypes of bacteria that potentially reside in the human stomach. The majority of these phylotypes was shown to belong to the Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Fusobacteria. This work also described that 10% of the clones represented organisms that were previously uncharacterized.

Subsequent ongoing studies continue to reveal high levels of diversity from the microbial species that inhabit the human body, with high levels of both intra- and interspecies diversity (Costello et al., 2009; Turnbaugh et al., 2009). Most of these studies that have investigated the diversity of the microbial species associated with the human body have however left important questions unanswered such as the identity of the nondominant community members and their biological roles, and which metabolic processes the populations that are present encode and carry out. In addition, the past 15 years of genomic research have made it clear that closely related species, and even species that are identical at the 16S rRNA level, can have wide variation in gene content (Perna et al., 2001; Kudva et al., 2002). Terms such as pan-genome and core-genome have been coined over the years to address the variations that are apparent in closely related species and have now been applied in a similar fashion to metagenomic populations (Callister et al., 2008; Bentley, 2009).

The initial publication of a shotgun sequencing of the human microbiome focused on the analysis of fecal samples from the human gastrointestinal tract (Gill et al., 2006). This study along with subsequent applications of shotgun techniques to the study of the human microbiome again highlighted the extent of microbial diversity associated with the human body (Grice et al., 2009). Gordon and co-workers, for example, investigated the interplay between the gastrointestinal ecology and energy balance of animals on a Western diet. Here they found that obesity that was induced by the diet resulted in an increased proportion of one class of the Firmicutes and that this same population could be reduced by manipulation of the diet. Transplantation of the microbial populations from the obese mice to lean germ-free mice resulted in a higher level of the deposition of fat than when these microbial populations were taken from lean donors (Turnbaugh et al., 2008). More recently, Gordon's group presented the results of a monozygotic and dizygotic twin pair study, where twins were concordant for leanness or obesity, and their mothers (Turnbaugh et al., 2009). The aim of this study was to address how host genotype, environmental exposure, and host adiposity influence the gut microbiome (Turnbaugh et al., 2009). The comparative analysis of fecal samples that were derived from 154 individuals yielded 9,920 near full-length and 1,937,461 partial bacterial 16S rRNA sequences. In addition, 2.14 Gb of metagenomic data was obtained from their microbiomes. The results from this analysis suggest that the gastrointestinal microbiome is shared among family members, but variations are present within each individual with respect to the lineages that can be observed. This variation was evident in both the monozygotic and dizygotic twin pairs. The results however suggest that there is a core functional microbiome that can vary depending on physiological states.

The Promise of the Human Microbiome

The genomic era created the possibility of studying the detailed genetics of many microbial species. These include pathogens and nonpathogens and species that have both positive and negative impacts on the environment. The developments in the genomics arena have taken advantage of emerging and improving sequencing technologies, as well as improved assembly algorithms and approaches coupled with reduced costs for generating genomic data. The developments also placed the greater scientific community in positions to ask in-depth gene-based questions and get real answers.

The advent of metagenomics was a natural progression of the genomics field, and in particular took advantage of the ability to sequence DNA that was derived from a mixed community and assemble this genetic information to reconstitute metabolic and physiological information of any community of choice. Metagenomic approaches have by now been successfully applied to environments as diverse as soils, the oceans (Nealson and Venter, 2007; Rusch et al., 2007; Yooseph et al., 2007; Yutin et al., 2007), and the human body (Gill et al., 2006; Costello et al., 2009; Turnbaugh et al., 2009) in an attempt to describe and decipher the microbial species that are present in these niches. Entire systems can now be studied with respect to viral, microbial, and fungal diversity, over varying time courses, and before and after perturbation (Costello et al., 2009). On the human side, when coupled with 16S rRNA analyses for detailed estimates of microbial diversity, metagenomics approaches present the perfect opportunity to address questions related to human health and associated problems. This is particularly relevant in the developing world, which presents its own series of challenges, some of which are presented below.

One of the most valuable examples to date of the potential benefits from knowledge gained with human microbiome studies comes from a series of studies performed by Dore and colleagues at INRA. The significance of the studies conducted by this group relates to how the evolution of initial studies focusing on the microbiome can result in recommendations to improve human health conditions. Their results evolved from initial metagenomic studies on human gastric samples (Manichanh et al., 2006) using fosmid libraries from six healthy patients, and six patients with Crohn's disease (CD). The group was able to identify 125 nonredundant ribotypes mainly represented by the phyla Bacteroidetes and Firmicutes, of which 43 distinct ribotypes were identified in the healthy microbiota, and only 13 in CD. This metagenomic approach that was initially published gave the first insight into the reduced microbial diversity in patients with CD.

Their work continued to focus on microbiology of CD (Seksik et al., 2006; Sokol et al., 2006; Sokol et al., 2007; Seksik et al., 2009). They most recently compared fecal samples of 22 patients active for CD (A-CD) patients, 10 CD patients in remission (R-CD), 13 active ulcerative colitis (A-UC) patients, four UC patients in remission (R-UC), eight infectious colitis (IC) patients, and 27 by 16S PCR and found that members of Firmicutes (*Clostridium leptum* and *C. coccoides* groups in particular) were less represented in A-IBD patients compared to healthy subjects (HS) with *Faecalibacterium prausnitzii* species (a major representative of the

C. leptum group) in lower abundance in A-IBD and IC patients compared with HS. As a result of the initial work, the group proposed that *F. prausnitzii* was important for gut homeostasis.

In 2008 members of the same group had published on the composition of the mucosa-associated microbiota of CD patients at the time of surgical resection and 6 months following using FISH analysis (Sokol et al., 2008), and again found reduced abundance of *F. prausnitzii* being correlated with an increased risk of postoperative recurrence of ileal CD. They further studied the anti-inflammatory effects of *F. prausnitzii* and showed that the bacterium exhibited anti-inflammatory effects on cellular and TNBS colitis models.

NIH-Funded HMP and International Components of the HMP

The US National Institutes of Health (NIH) initiated a roadmap program focused on the human microbiome (Peterson et al., 2009). The project has been described as a community resource, with overarching aims to help determine the core human microbiome, to understand the changes in the human microbiome that can be correlated with changes in human health, to develop new technological and bioinformatics tools to support these goals, and to address the ethical, legal, and social issues raised by human microbiome research (http://nihroadmap.nih.gov/hmp/). The project has a heavy sequencing and data analysis component that currently is underway at the four large NHGRI/NIAID-funded sequencing centers (J. Craig Venter Institute (JCVI), Baylor College of Medicine, The Broad Institute and Washington University). The current sequencing focus includes generating at least 3000 reference genomes at various levels of finishing (Chain et al., 2009), coupled with significant 16S rRNA sequencing and metagenomic sequencing from 15 to 18 body sites on 300 individuals some of which would be repeat sample donors (http://nihroadmap.nih.gov/hmp/). A number of "Investigator"-driven demonstration projects have also been awarded. These demonstration projects aim to understand the implications of a number of diseases including CD, BV, psoriasis, and esophageal cancer to name a few (Peterson et al., 2009). It is anticipated that the results from these demonstration projects will give additional insights into the relationship between human health and disease and changes in the human microbiome. Finally, the Human Microbiome Project (HMP) roadmap initiative has awarded funds to investigate new technologies for improving knowledge of the human microbiome, as well as for the development of computational tools that will increase the value of metagenomic data (http://nihroadmap.nih.gov/hmp/fundedresearch.asp), and the ethical, legal, and social implications of this work.

In summary, and as captured in the recent publication by Peterson et al. (2009), the goals of the HMP are to demonstrate the characterization of the human microbiome with population, genotype, disease, age, etc., and also catalog the influence of disease. The aim also is to present a standardized data resource and technological benefits. The project will go over 5 years at a funding level of close to 157 million US dollars.

Since the launch of this roadmap initiative in 2007, a number of other HMPs have been described. An overview of available projects as of mid-2008 was presented in an editorial (Mullard, 2008). Projects beyond the large NIH USA based human microbiome efforts include work in Europe, China, Australia, and Canada. In 2007, the European Commission committed close to 31 million US dollars to a 4-year initiative called the Metagenomics of the Human Intestinal Tract (METAHIT) where the primary focus is the microorganisms that inhabit the gut, and how they contribute to obesity and inflammatory bowel disease (Mullard, 2008). A review of this effort is presented in another chapter written by Ehrlich and colleagues.

We are all cognizant of the fact that age, diet, and geographical location contribute to variations in the human microbiome. Consequently, the more initiatives that we have in diverse parts of the world, the better positioned we will be to fully understand the key components of the microbiome and their interactions with the host under various environmental and physiological cues.

Implications of the HMP for the Developing World

Because of a slow rate of progress in the areas of scientific research, along with low levels of available funding and investment in sciences in most developing countries, there has been very little scientific contribution toward solving major problems that hinder their global development. As Coloma and Harris (2009) nicely put it, "researchers in many developing countries will not be participating in genomics research, mainly because of their technological isolation and their limited resources and capacity for genomics research combined with the urgency of many other health priorities." Areas such as public health, emerging infectious diseases, and agricultural development, which are key to long-lasting and sustainable national development, still lack the funding and innovation required to mitigate their inability to contribute to global development. The global health sector is of particular importance given the increasing number of diseases that plague the developing world (some of which are making a comeback after several years under effective control). Examples of some of these are detailed below. Consequently, in most developing countries throughout the world, and specifically in sub-Saharan Africa, South America, and Asia, there is a serious need to improve public health. In these countries, communicable diseases caused by known and even unknown pathogens (see below) remain a leading cause of mortality. Emerging infectious diseases are a major cause for alarm, and malnutrition and associated effects are also major issues that need to be effectively addressed.

If one takes emerging infectious disease as an example, this captures many viral and bacterial agents. Severe acute respiratory syndrome (SARS) was the first infectious disease to emerge in the 21st century, and other emerging viral infectious diseases according to The World Health Organisation (WHO) include but are not limited to Ebola and Marburg hemorrhagic fevers. In addition, in an earlier publication by WHO ("New WHO office to help developing countries control emerging

diseases"; J Environ Health 63, 2001) it was stated that in 1998 alone, communicable diseases caused the death of over 13 million people worldwide, mainly among the poorest populations of developing countries. Since then, more than 30 new communicable diseases have been identified, and this list includes several diseases that were thought to be almost extinct that apparently have come back into the human population. Certain food-borne diseases are also considered to be emerging as they now occur more often, and that list includes outbreaks of salmonellosis, which have increased significantly in the past 25 years. *Listeria monocytogenes* also falls into this category as its major role in food-borne diseases has become more recently appreciated, and some food-borne trematodes are also emerging as a serious public health concern. Although food-borne infections with *Escherichia coli* serotype O157:H7 were first described in 1982, it has rapidly emerged to be a leading cause of infections, which in turn result as a major cause of bloody diarrhea and acute renal failure, with an infection that is sometimes fatal. Finally, while cholera devastated much of Asia and Africa for years, its introduction for the first time in almost a century on the South American continent in 1991 makes it another example of an emerging infectious disease.

In addition, very little to none of the successful metagenomics stories in understanding the human microbiome and its role in aspects of human health have been conducted or duplicated in developing countries. Notwithstanding ongoing efforts focusing on vaccines, better diagnostics, and improved treatment of many of these diseases, it is becoming increasingly essential to complement such approaches with an investigation of the role of the human microbiome on human health. Several areas of anticipated important contribution of the human microbiome include zoonotic diseases and other emerging and re-emerging infectious diseases, sexually transmitted diseases, diarrheal diseases, respiratory diseases, eukaryotic diseases, malnutrition, and the integration of probiotics for improving human health. In addition to the translation of findings into practical approaches for improving human health, other opportunities offered by the human microbiome initiative are related to the transfer of technology to developing countries and the associated long-term benefits to training local populations in these developing sciences so that nations can retain the benefits. This will further strengthen capacity in genomics and bioinformatics and all the associated downstream applications that come with these areas of research.

It is now appreciated that there is a resident (which constitutes the core microbiome) and a nonresident microbiota. The nonresident microbiota contains known and unknown microbes that cause a wide range of human diseases, most of which remain to be effectively controlled in both the developed and the developing world. Human losses in the developing world in terms of mortality and contributions to economic development appear to be greater, however. Currently, for example, communicable diseases caused by eukaryotic parasites such as *Plasmodium* spp., *Leishmania* spp., *Trypanosoma* spp., and various viruses, among others, remain serious public health concerns in the developing world and affect more than 1.2 billion people (Mahmoud and Zerhouni, 2009). In this context, scientific challenges that include genomics, metagenomics, proteomics, and metabolomics-related activities

need to be expanded to ultimately encompass system and ecological understanding of communities of microbes and their interactions with humans. It has in fact been anticipated that the control of these diseases may be accelerated by the complete understanding of the genomes of these species, coupled with an understanding of the changes of the human microbiome that favor or reduce/eliminate their virulence and/or transmissibility. The adaptation processes, for example, by which zoonotic microorganisms that enter the human population adapt to become pathogens overtime can also be accelerated by longitudinal studies that focus on the populations on the bodies of both healthy and diseased individuals.

The Promise for Technology Development in the Developing World

As with most advanced technological and scientific approaches, and as is evident from the developing countries reports presented above, the development and applications of technological advances probably will take a significant amount of time to trickle to the developing world. In the realm of genomics and metagenomics as applied to human health, there is limited evidence that this will happen soon enough to allow developing countries to actively participate and shape research in these new fields. However, a recent award from the Bill & Melinda Gates Foundation (BMGF) to Dr. Jeffrey Gordon at The Washington State University in St. Louis to study childhood malnutrition in developing countries (http://www.gatesfoundation.org/press-releases/Pages/child-malnutrition-microbial-cells-study-090331.aspx) suggests that there will be more movement in the direction of applying these technologies to problems in the developing world. For the above-mentioned study, Gordon's group will investigate the microbes in severe malnutrition with a major focus on severely malnourished infants living in Malawi and Bangladesh, and whether their microbial flora varies from that found in healthy infants who live in the same environment. It is anticipated that as a result of these studies, we will be better positioned to develop effective treatment regimes for these disease conditions. This award is part of an initiative by the BMGF to fund research on malnutrition (http://mednews.wustl.edu/news/page/normal/13840.html?emailID=23653).

In addition to that award and the anticipated outcome, there have been a significant number of plant and microbial genome projects initiated and conducted in the developing world. The range of microbial species that have been sequenced includes human, plant, and animal pathogens, as well as organisms that have potential benefit to the environment. Some of these species include *Actinobacillus pleuropneumoniae* serovar 3 str. JL03 that causes fibrinous and necrotizing pleuropneumonia in swine, and that was sequenced by the Huazhong Agricultural University in China and *Haemophilus parasuis* SH0165 also sequenced by the Huazhong Agricultural University/Institute of Pathogen Biology/Chinese Academy of Medical Sciences/Peking Union Medical College. *Chromobacterium violaceum* ATCC 12472 was sequenced by the LNCC (National Laboratory of Scientific Computing in Rio de Janiero, Brazil); this bacterium carries the bacteriocidal

pigment violacein and can also cause diarrhea and septicemia in humans. *Ehrlichia ruminantium* str. Welgevonden was sequenced at the University of Pretoria, South Africa. *Leifsonia xyli* subsp. xyli str. CTCB07, the causative agent of ratoon stunting disease in sugar cane, was sequenced by the Sao Paulo state (Brazil) Consortium and *Leptospira interrogans* serovar Copenhageni str. Fiocruz L1-130 and *Xylella fastidiosa* were also sequenced by the same group. *Leptospira interrogans* serovar Lai str. 56601 sequenced by the Chinese National HGC, Shanghai, and *Lysinibacillus sphaericus* C3-41 sequenced by the Chinese Academy of Sciences/Beijing Genomics Institute, Chinese Academy of Sciences.

The developing world has also been involved in the sequencing and analysis of some of the major eukaryotic parasites such as *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, and *Theileria parva* (Nene et al., 2000; Berriman et al., 2005; Bishop et al., 2005; El-Sayed et al., 2005; Gardner et al., 2005). There have also been several initiatives that have looked at host genotyping in many developing countries. According to Coloma and Harris (2009), Thailand, South Africa, Indonesia, Brazil, Mexico, and India have all devoted resources to studies on human genetics and variation in human populations.

As a result of many of these initiatives in developing countries, a limited capacity of tool development for genomics and bioinformatics approaches has occurred. However, much more remains to be achieved in technology and knowledge transfer, particularly in sub-Saharan Africa and Latin America. The main focus should be on developing genomics platforms leveraging on the next-generation sequencing approaches that remain to be established in much of the developing world.

Monitoring of Zoonotic Infections and Global Surveillance of Emerging and Reemerging Infectious Diseases

Events of emerging and reemerging infectious diseases in the human population remain constant occurrences in sub-Saharan Africa, Southeast Asia, and South America. Emerging infectious disease events such as SARS (Field, 2009) and the most recent pandemic of H1N1 (Gibbs et al., 2009) illustrate and confirm the constant flow of pathogens from wild and domesticated animals, and other reservoirs into the human population. Chikungunya fever, which is an arboviral infection, reemerged in Asia in 2005–2006 after a long period of quiescence (Bhatia and Narain, 2009). It is thought that factors including microbial, climatic, social, and economic aspects influenced the reemergence of the disease and the pace at which it spread, eventually resulting in high death rates (Bhatia and Narain, 2009).

Indeed, there are many microorganisms (viral, bacterial, and eukaryotes) that have moved into the human population and remain part of the human microbiota, which, when able to effectively survive, can cause either new diseases or disease with a much higher severity. Such cases of unknown and potentially pathogenic microorganisms in circulation in the human population are usually favored by factors related to (1) the ability of microorganisms to adapt in new hosts, (2) human actions (interactions with wild animals, hunting and effects on the environment

leading to ecological disturbances), and (3) human movements as a result of global world travel (Field, 2009). Consequently, at any particular time, there could be a set of known and unknown microorganisms present in a given individual or a population as a result of their presence in and interaction with a specific environment or organisms therein such as animal reservoirs (i.e., bats, mice, and rats) of known and unknown microorganisms. The main challenge in the context of forecasting, and better yet preventing emerging and reemerging infectious diseases has been early detection and genetic identification of such known and unknown microorganisms. Global human microbiome studies using metagenomics analysis of known and unknown microorganisms provide unique but powerful opportunities to uncover the near-complete composition of the microbial content of an individual or a population at any given time, thus setting the stage for a comprehensive inventory of the genetic characteristics of potential human pathogens.

Studies of the human microbiome in the developing world will likely contribute significantly to the discovery of emerging pathogens (viruses, bacteria, and others) in circulation in humans. In fact, in both developed and developing countries the issues of early identification of emerging pathogens have been an impediment for the prevention of emerging and reemerging infections. Based on recent studies, human metagenomics coupled with the next-generation DNA sequencing provides an opportunity for early detection of microbial organisms even when there is significantly low abundance (Relman, 2002).

Because of the extreme importance of monitoring zoonotic infections, metagenomics studies should in principle be extended from humans to domesticated and nondomesticated animals. In fact, based on the technologies already available for human metagenomics studies, there has been increasing interest in launching animal metagenomics initiatives. Such initiatives will not only provide insights into the resident and transient microbial populations but also, in the case of natural reservoirs of given microorganisms, provide an opportunity to pinpoint the genetic changes that must occur for their adaptation to a new host – the human body. This is applicable in particular to invertebrate vectors and bats that are known to be host to a number of highly pathogenic viruses that pose significant public health problems to humans.

The Case of Selected Emerging Infectious Diseases: Tuberculosis (TB) and Leptospirosis

Developing countries are particularly affected by the impact of *Mycobacterium tuberculosis* virulence and TB drug resistance. This has been primarily because of genome plasticity in the causative agent. Unfortunately, available microarray-based platforms to identify strain diversity have not been fully implemented with the greatest TB incidence largely due the HIV/AIDS epidemic. The renewed interest and funding for top infectious diseases has recently revamped efforts to accelerate TB research, with a particular focus on the use of integrated approaches to find better control measures. In this context, it is proposed and highly anticipated that key

aspects such as the integration of large-scale "omics," datasets focusing on parasite genetic determinants, host genetics, and host–parasite interactions will be crucial for this quest for better control measures. In addition, and given recent reports, the human microbiome would be a great addition to this integration of data in the context of systems biology. To this end, the evaluation of the human microbiome in cases of latent, nonlatent TB, and drug-resistant TB infections will provide insights into the role of the human (resident and nonresident) flora in various aspects of TB infections. Such information would most likely contribute to improving diagnosis, control, and spread of TB infection.

Another example of the potential to come from using human metagenomic research and approaches in the developing world relates to another emerging infectious pathogen that causes *Leptospirosis*. The Leptospires cause an infection that is associated with very high levels of mortality annually, but have received relatively little attention, probably because the infection is concentrated in the tropical regions and in the developing world. More than half a million cases are reported annually, and majority of these cases are associated with human exposure to pathogenic *Leptospira* species in the environment. Mortality rates as high as 25% have been recorded.

The genus *Leptospira* is serologically divided into two species: *L. interrogans*, which is pathogenic to humans and animals, and *L. biflexa*, a free-living nonpathogenic species found in water and wet soil. More than 16 pathogenic and saprophytic species have been recognized. Many animals including rodents and dogs are known to be reservoirs of *Leptospira*, and humans are considered to be the accidental hosts of this pathogen. Transmission of the pathogen is primarily from soil and water to mammalian tissues (often noticed following on large-scale flooding), with the infection occurring via penetrating leptospires through mucosa or open skin. Symptoms of leptospirosis include meningitis, pneumonitis, hepatitis, nephritis, pancreatitis, erythema nodosum, and death. No human vaccine against leptospirosis is available, and mild leptospirosis is treated with doxycycline, ampicillin, or amoxicillin. For severe leptospirosis, the primary therapy is penicillin G. The molecular diagnosis of Leptospirosis has been with traditional approaches such as restriction enzyme analysis, nucleic acid probes and hybridization, pulse field gel electrophoresis (PFGE), and varying ribotyping approaches.

Genome sequences from at least six *Leptospiras* have become available in the past few years, and these genomes are providing insight on the diversity of these species. In addition, the availability of these genomes is allowing for the identification of novel virulence factors, and ultimately will facilitate vaccine development. Recently, the genome sequence of the free-living *L. biflexa* was completed (Picardeau et al., 2008) and shown to contain 3,590 protein-coding genes distributed across three circular replicons. In the current study, it has been estimated that 2,052 genes (61%) represent a progenitor genome that existed before divergence of pathogenic and saprophytic *Leptospira* species. Basically, nearly one-third of the *L. biflexa* genes are absent in pathogenic *Leptospira.* In addition, 1,431 pathogen specific genes that are found in the pathogenic Leptospires are not present in *L. biflexa.* Of these, 893 genes have no known function suggesting that there are

mechanisms that are unique to *Leptospira* and that the pathogenic specific genes need further study. The resulting genome studies suggest that there is still a significant amount of information that is not understood about the *Leptospiras*, particularly as it relates to how the species adapts to new environments and how the genomes mutate. Metagenomic studies of samples derived from infected populations will present an opportunity to study the pathogen without repeated passage where it has been shown to have genome rearrangements. In parallel, the pathogen can be studied directly in the environment when it is in transition from its natural host to humans (the accidental host). Interestingly, there is a large NIAID-funded project underway at the JCVI to sequence the genomes of an additional 400 *Leptospira* isolates (Joe Vinetz, personal communication; http://gsc.jcvi.org/).

Leptospirosis is another example of an emerging infectious disease that is prevalent in tropical environments and has not received as much attention as the major diseases in the developed world although the causative organisms result in a high mortality rate. Genomics and metagenomics approaches have the potential to increase the understanding of these species and their impact on human health.

Potential for Understanding and Control of Diarrheal Diseases

Diarrheal diseases remain one of the leading causes of deaths worldwide (Culligan et al., 2009). Specifically, diarrheal diseases are the second most common cause of child deaths worldwide, and more than 80% of child deaths due to diarrhea occur in Africa and south Asia. Worldwide, 88% of deaths from diarrhea are due to unsafe water and poor sanitation or hygiene. Three-quarters of all deaths from diarrhea in children younger than 5 years occur in 15 countries. There are about 2.5 billion cases of diarrhea among children each year, in addition to those who die from the disease. The UN reports that vaccines and better hygiene could decrease the number of deaths from diarrhea among children.

Since the 1970s, oral rehydration therapy has been the cornerstone of treatment programs. This therapy prevents dehydration that is associated with diarrhea. Giving zinc supplements with oral rehydration salts has also been shown to reduce the length of the illness and also the risk of more diarrhea episodes. Sixty percent (60%) of children in developing countries do not get the recommended treatment for diarrhea, which is vaccination against rotavirus, the leading cause of the disease. In fact, rotavirus causes about 40% of hospital admissions of children below 5 years suffering from diarrhea. Current therapies are focused on rehydration therapies but the studies from a human microbiome approach, coupled with the development of novel antibiotics and/or probiotics holds significant potential (Culligan et al., 2009).

Many diarrhoeal diseases have been associated with viruses (Ramani and Kang, 2009). Recent results suggest that viruses are present in as much as 43% of diarrheal samples in the developing world (Ramani and Kang, 2009). There are however a significant number of cases of diarrhea without obvious causes, thus making it difficult to control them. In addition and specifically in the case of rotaviruses, because of their high genetic diversity, the emergence of new genotypes, and the reassortment

between different genotypes (Matthijnssens et al., 2009), there is constant need for surveillance of circulating strains. Human metagenomics studies hold the promise for increasing our understanding of the diversity of rotavirus and other etiological agents of diarrheal diseases. Based on previous studies, gastrointestinal tract metagenomics studies in both healthy and diarrheal patients in developing countries may lead to the identification and association of additional microorganisms (bacteria, viruses, and eukaryotes) with various cases of diarrheal diseases (Finkbeiner et al., 2008).

As an example, recent human microbiome studies have led to the discovery of a novel virus of the Cosavirus genus and its association with acute diarrhea in a child in Australia (Holtz et al., 2008). Regular and comprehensive metagenomics analyses focusing on acute and difficult-to-cure cases of diarrhea and diarrhea cases with known and unknown causes primarily in developing countries may provide opportunities for (1) a constant assessment of the diversity of known causative agents of diarrhea and (2) identification of new microorganisms as they relate to cases of diarrheal diseases.

Potential for Understanding Sexually Transmitted Diseases

Sexually transmitted diseases (STDs) are common infections throughout the developed and the developing world. STDs can result in premature birth, stillbirth, and neonatal infections (De Schryver and Meheus, 1990).

Many ongoing studies on BV aim to understanding the microbial populations that are present in the vaginal ecosystem and how they vary under health and disease conditions. Recent studies that are focused on 16S rRNA gene analysis have suggested that the extent of microbial diversity in the vaginal tract is not fully understood, which in turn has implications for current treatment regimes. This has potentially significant implications for asymptomatic disease conditions for example. Additional results show a lack of homogeneity within the vaginal tract, highlighting a complex ecosystem (Kim et al., 2009). Metagenomic approaches to studying this environment promise to give additional insights into the extent of diversity within this niche.

Ongoing studies in several parts in sub-Saharan Africa reveal that there is some relationship between the population of microbes that exists in the vaginal tract and STDs. Recently, van de Wijgert et al. (2008) described a study in which they investigated the relationships among BV, vaginal yeast, and vaginal practices, mucosal inflammation, and HIV acquisition. From a cohort of 4,531 HIV-negative women, they observed that women who were positive for BV or vaginal yeast had a higher likelihood to acquire HIV, and they concluded that BV and yeast may contribute more to the HIV epidemic than previously appreciated (van de Wijgert et al., 2008). Similar observations have been made in a review of all available literature on the extent to which BV may increase the risk of HIV acquisition (Atashili et al., 2008).

Earlier, in 2000, van De Wijgert et al. (2000) studied 169 Zimbabwean women to determine if intravaginal practices could be associated with changes in the vaginal

flora and acquisition of STDs. In this study, they found that some disturbances of the flora could be associated with increased likelihood of STDs and HIV; the absence of *Lactobacilli* from the vaginal flora was associated with being positive for HIV (van De Wijgert et al., 2000). Martin et al. (1999) had similarly looked at a cohort of sex workers in Kenya and demonstrated that although only 26% of these women were colonized with *Lactobacillus* species at baseline, follow-up studies showed that the absence of culturable vaginal lactobacilli could be associated with the increased likelihood of acquiring HIV-1. Abnormal vaginal flora on Grams-stain was associated with increased risk of both HIV-1 acquisitions. This group proposed that the treatment of BV and the use of lactobacilli to colonize the vaginal cavity should be evaluated for reduce risk of acquiring HIV-1, gonorrhea, and trichomoniasis (Martin et al., 1999). How the microbial populations in the vaginal cavity can contribute to reduce chances of HIV infection is one of those major areas that need attention, and that will undoubtedly benefit from human microbiome research.

Potential for Enhancement of Malaria Treatment Regimens

According to The World Health Organisation (WHO, http://www.who.int/mediacentre/factsheets/fs094/en/), every 30 seconds a child dies of malaria, a disease that can be prevented and cured. In 2006 there were 247 million cases of malaria, and these resulted in nearly 1 million deaths mostly among African children. In fact, 90% of all malaria deaths occur in sub-Saharan Africa. People who live in lower-income communities, i.e., approximately half of the world's population, are at risk of the disease. The WHO reports that in 2006 malaria was present in 109 countries and territories.

The disease, however, can be eradicated, says Bill Gates. In an interview with the BBC World Services World Today program in January 2010, Gates said "*we have a vaccine that's in the last trial phase – called phase three.*" He added that "a partially effective vaccine could be available within 3 years." A vaccine that is fully effective against malaria would take 5–10 years to develop.

Although most cases of malaria are found in sub-Saharan Africa, there are other countries, including in Asia, Latin America, the Middle East, and parts of Europe, that are also affected. Key interventions include prompt and effective treatment with artemisin-based combination therapies; people at risk using insecticide nets; and indoor residual spraying with insecticide to control the vector mosquitoes. Genomics approaches have already been used to elucidate the genomes of several of the *Plasmodium* species (Gardner et al., 2002; Carlton et al., 2008; Pain et al., 2008; Mitsui et al., 2009), but new metagenomics approaches present opportunities to monitor the impact of the parasite of the microbial communities that reside on and in the human body, with a longer-term potential to develop novel probiotic approaches to supplement nutrition of infected individuals while the parasite runs its course.

In countries that have a high rate of malaria, economic growth rates may be cut by as much as 1.3%. In addition, genomic studies on the environments, in which the

mosquitoes reside and breed, are being and will continue to allow for an increased understanding of the communities that they require for their survival (Ponnusamy et al., 2008a, b, 2009). This is particularly relevant since mosquitoes breed in areas where there are wet conditions, and the transmission of the disease can differ according to local factors such as rainfall, proximity of mosquito breeding sites to people, and the mosquito species in the area. A November 2009 report from Susan Anyangu in Nairobi, Kenya, carried by Inter Press Service (IPS) states that the RTS.S vaccine being developed is to be used specifically in Africa. It will be for infants and children aged less than 5 years (the most vulnerable to malaria). The vaccine could be ready for use in 5 years time.

Supplementing nutrition of people with malaria with probiotic solutions that have been derived from a metagenomic approach to understand the human microbiome holds significant promise. The FAO/WHO defines probiotics as "Live microorganisms which, when administered in adequate amounts confer a health benefit on the host." Probiotics have become more and more valuable over the past few years and are available in a number of food sources, including yogurts and other milk products, fermented and unfermented milk, and some juices. These live microorganisms are in most cases bacteria that are similar to beneficial microorganisms found in the gastrointestinal tract. Each species that is present in the gut environment would seem to hold some potential for use as a probiotic and therefore in human health.

Probiotics have been shown to be effective in treating irritable bowel syndrome (IBS), childhood and traveler's diarrhea, prevention and treatment of vaginal yeast infection and urinary tract infection, preventing and treating inflammation of the colon after surgery, reduction of the recurrence of bladder cancer, shortening the time of intestinal infections, and preventing eczema in children. Although the benefits of probiotics are evident, they have yet to be adapted extensively in the developing world (Reid and Devillard, 2004). Other ideas on the use of probiotics for reducing the morbidity and mortality associated with HIV/AIDS have been explored and proposed (Reid et al., 2005) where it has been proposed that lactic acid bacteria could play a role in maintaining the health of the human gut. We can only hope that as a result of the initiatives of the human microbiome project, new probiotics for a range of human health conditions may be developed based on baselines for people in different geographic locales.

Challenges: Funding and Technology Transfer

The efficient implementation of human microbiome research relies on the advanced instrumentation necessary for the processing of collected clinical samples, preparation and amplification of nucleic acid, and DNA sequencing. In addition, DNA sequence analysis also requires advanced bioinformatics resources. All genomics-related technologies developed over the past 10 years remain very expensive to be acquired by developing countries. This is usually justified by low-use volume and

high costs of equipment and maintenance (Coloma and Harris, 2009). Therefore, as suggested by these authors, involvement of laboratories and institutions in developing countries should take advantage of "North–South" and "South–South" collaborations. Previous examples of successful "North–South" collaborations could be leveraged to initiate new ones in the context of human microbiome studies.

For the past several years, there have been numerous initiatives in developing countries to reduce the technological divide and hence begin to actively contribute to genomics research. In this context, activities have included training and capacity building in genomics and bioinformatics. In addition, there has also been an emphasis on the development of "Centers for Excellence" to provide resources and a critical scientific mass at regional levels. Four such regional "Centers of Excellence" are currently being established in eastern and central Africa, southern Africa, west Africa, and north Africa.

One of the most advanced "Centers for Excellence," Biosciences for Eastern and Central Africa (BecA), located at the International Livestock Research Institute (ILRI) in Nairobi, Kenya, has established facilities (with advanced genomics and bioinformatics resources) to support and accelerate research in a wide range of disciplines, including plant/crop sciences and animal sciences. Such infrastructure would ideally be poised for use as a focal point for the implementation of a regional initiative on the human microbiome. The existence of such facilities would normally be used to engage various African institutions in South–South collaborations. The "South–South" collaborations indeed provide opportunities to strengthen the scientific capacity of institutions in developing countries, which would be translated into their effective participation in North–South initiatives.

Genomics and metagenomics initiatives are usually quite expensive, and obviously, most institutions in the developing world would not be able to fund such activities independently. However, given the existence of several human microbiome projects in the United States, Canada, Europe, China, Japan, Singapore, and Australia, components in developing countries could easily be justified. For example, an African component of the human microbiome would provide elements to answering important outstanding microbiome questions, among which are included: (1) Is there a core human microbiome? (2) Does the composition of the human microbiome vary from one geographical region to another?

Given the anticipation of such interesting outcomes, existing initiatives could further provide seeds to launch other initiatives in the developing world. Furthermore, in the context of the use of biosciences for Africa's development, a strong case should be made to various stakeholders such as The African Union and other regional organizations to fund the African component of the human microbiome. This next wave of genomics research will not be without its own set of challenges. Recent studies, for example, show that many diseases present with similar observations, and as such initial surveys into the human microbiome under health and disease may give unexpected outcomes (Yazdanbakhsh and Kremsner, 2009).

Acknowledgement The authors wish to acknowledge the invaluable information found on the World Health Organisation (WHO) website and on the Mayo Clinic website.

References

Atashili J, Poole C, Ndumbe PM, Adimora AA and Smith JS (2008) Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS 22:1493–1501

Bentley S (2009) Sequencing the species pan-genome. Nat Rev Microbiol 7:258–259

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE and El-Sayed NM (2005) The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416–422

Bhatia R, Narain JP (2009) Re-emerging chikungunya fever: some lessons from Asia. Trop Med Int Health 14:940–946

Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, Perez-Perez G, Blaser MJ, Relman DA (2006) Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci USA 103:732–737

Bishop R, Shah T, Pelle R, Hoyle D, Pearson T, Haines L, Brass A, Hulme H, Graham SP, Taracha EL, Kanga S, Lu C, Hass B, Wortman J, White O, Gardner MJ, Nene V, de Villiers EP (2005) Analysis of the transcriptome of the protozoan *Theileria parva* using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. Nucleic Acids Res 33: 5503–5511

Callister SJ, McCue LA, Turse JE, Monroe ME, Auberry KJ, Smith RD, Adkins JN, Lipton MS (2008) Comparative bacterial proteomics: analysis of the core genome concept. PLoS One 3:e1542

Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, Del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczky M, Meyer EV, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, Salzberg SL, Stoeckert CJ, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 455:757–763

Chain PS, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Detter JC (2009) Genomics. Genome project standards in a new era of sequencing. Science 326:236–237

Coloma J, Harris E (2009) Molecular genomic approaches to infectious diseases in resource-limited settings. PLoS Med 6:e1000142

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI and Knight R (2009) Bacterial community variation in human body habitats across space and time. Science 326: 1694–1697

Culligan EP, Hill C, Sleator RD (2009) Probiotics and gastrointestinal disease: successes, problems and future prospects. Gut Pathog 1:19

De Schryver A, Meheus A (1990) Epidemiology of sexually transmitted diseases: the global picture. Bull World Health Organ 68:639–654

Dorer MS, Talarico S, Salama NR (2009) *Helicobacter pylori's* unconventional role in health and disease. PLoS Pathog 5:e1000544

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) Diversity of the human intestinal microbial flora. Science 308:1635–1638

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N (2005) Comparative genomics of trypanosomatid parasitic protozoa. Science 309:404–409

Field HE (2009) Bats and emerging zoonoses: henipaviruses and SARS. Zoonoses Public Health 2009-May 28th

Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D (2008) Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog 4:e1000011

Foster JA, Krone SM, Forney LJ (2008) Application of ecological network theory to the human microbiome. Interdiscip Perspect Infect Dis 2008:839501

Gao Z, Tseng CH, Pei Z, Blaser MJ (2007) Molecular analysis of human forearm superficial skin bacterial biota. Proc Natl Acad Sci USA 104:2927–2932

Gao Z, Tseng CH, Strober BE, Pei Z, Blaser MJ (2008) Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. PLoS One 3:e2719

Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJ, Sato S, Ralph SA, Mann DJ, Xiong Z, Shallom SJ, Weidman J, Jiang L, Lynn J, Weaver B, Shoaibi A, Domingo AR, Wasawo D, Crabtree J, Wortman JR, Haas B, Angiuoli SV, Creasy TH, Lu C, Suh B, Silva JC, Utterback TR, Feldblyum TV, Pertea M, Allen J, Nierman WC, Taracha EL, Salzberg SL, White OR, Fitzhugh HA, Morzaria S, Venter JC, Fraser CM, Nene V (2005) Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. Science 309:134–137

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419:498–511

Gibbs AJ, Armstrong JS, Downie JC (2009) From where did the 2009 'swine-origin' influenza A virus (H1N1) emerge? Virol J 6:207

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. Science 312:1355–1359

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA (2009) Topographical and temporal diversity of the human skin microbiome. Science 324:1190–1192

Holtz LR, Finkbeiner SR, Kirkwood CD, Wang D (2008) Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. Virol J 5:159

Kim TK, Thomas SM, Ho M, Sharma S, Reich CI, Frank JA, Yeater KM, Biggs DR, Nakamura N, Stumpf R, Leigh SR, Tapping RI, Blanke SR, Slauch JM, Gaskins HR, Weisbaum JS, Olsen GJ, Hoyer LL, Wilson BA (2009) Heterogeneity of vaginal microbial communities within individuals. J Clin Microbiol 47:1181–1189

Kudva IT, Evans PS, Perna NT, Barrett TJ, Ausubel FM, Blattner FR, Calderwood SB (2002) Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. J Bacteriol 184:1873–1879

Lasken RS (2009) Genomic DNA amplification by the multiple displacement amplification (MDA) method. Biochem Soc Trans 37:450–453

Mahmoud A, Zerhouni E (2009) Neglected tropical diseases: moving beyond mass drug treatment to understanding the science. Health Aff (Millwood) 28:1726–1733

Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, Roca J, Dore J (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut 55:205–211

Martin HL, Richardson BA, Nyange PM, Lavreys L, Hillier SL, Chohan B, Mandaliya K, Ndinya-Achola JO, Bwayo J, Kreiss J (1999) Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J Infect Dis 180:1863–1868

Matthijnssens J, Bilcke J, Ciarlet M, Martella V, Banyai K, Rahman M, Zeller M, Beutels P, Van Damme P, Van Ranst M (2009) Rotavirus disease and vaccination: impact on genotype diversity. Future Microbiol 4:1303–1316

Mitsui H, Arisue N, Sakihama N, Inagaki Y, Horii T, Hasegawa M, Tanabe K, Hashimoto T (2009) Phylogeny of Asian primate malaria parasites inferred from apicoplast genome-encoded genes with special emphasis on the positions of *Plasmodium vivax* and *P. fragile*. Gene 2010, Gene 450:32–38

Mullard A (2008) Microbiology: the inside story. Nature 453:578–580

Nealson KH, Venter JC (2007) Metagenomics and the global ocean survey: what's in it for us, and why should we care? ISME J 1:185–187

Nene V, Bishop R, Morzaria S, Gardner MJ, Sugimoto C, ole-MoiYoi OK, Fraser CM, Irvin A (2000) *Theileria parva* genomics reveals an atypical apicomplexan genome. Int J Parasitol 30:465–474

Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, Pasini EM, Aslett MA, Balasubrammaniam S, Borgwardt K, Brooks K, Carret C, Carver TJ, Cherevach I, Chillingworth T, Clark TG, Galinski MR, Hall N, Harper D, Harris D, Hauser H, Ivens A, Janssen CS, Keane T, Larke N, Lapp S, Marti M, Moule S, Meyer IM, Ormond D, Peters N, Sanders M, Sanders S, Sargeant TJ, Simmonds M, Smith F, Squares R, Thurston S, Tivey AR, Walker D, White B, Zuiderwijk E, Churcher C, Quail MA, Cowman AF, Turner CM, Rajandream MA, Kocken CH, Thomas AW, Newbold CI, Barrell BG, Berriman M (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. Nature 455: 799–803

Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M (2009) The NIH human microbiome project. Genome Res 19(12):2317–2323

Picardeau M, Bulach DM, Bouchier C, Zuerner RL, Zidane N, Wilson PJ, Creno S, Kuczek ES, Bommezzadri S, Davis JC, McGrath A, Johnson MJ, Boursaux-Eude C, Seemann T, Rouy Z, Coppel RL, Rood JI, Lajus A, Davies JK, Medigue C, Adler B (2008) Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis. PLoS One 3:e1607

Ponnusamy L, Wesson DM, Arellano C, Schal C, Apperson CS (2009) Species composition of bacterial communities influences attraction of mosquitoes to experimental plant infusions. Microb Ecol 2010 59:158–73

Ponnusamy L, Xu N, Nojima S, Wesson DM, Schal C, Apperson CS (2008a) Identification of bacteria and bacteria-associated chemical cues that mediate oviposition site preferences by *Aedes aegypti*. Proc Natl Acad Sci USA 105:9262–9267

Ponnusamy L, Xu N, Stav G, Wesson DM, Schal C, Apperson CS (2008b) Diversity of bacterial communities in container habitats of mosquitoes. Microb Ecol 56:593–603

Ramani S, Kang G (2009) Viruses causing childhood diarrhoea in the developing world. Curr Opin Infect Dis 22:477–482

Reid G, Devillard E (2004) Probiotics for mother and child. J Clin Gastroenterol 38:S94–101

Reid G, Anand S, Bingham MO, Mbugua G, Wadstrom T, Fuller R, Anukam K, Katsivo M (2005) Probiotics for the developing world. J Clin Gastroenterol 39:485–488

Relman DA (2002) New technologies, human–microbe interactions, and the search for previously unrecognized pathogens. J Infect Dis 186(Suppl 2):S254–258

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC (2007) The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77

Seksik P, Sokol H, Lepage P, Vasquez N, Manichanh C, Mangin I, Pochart P, Doré J, Marteau P (2006) The role of bacteria in onset and perpetuation of inflammatory bowel disease. Aliment Pharmacol Ther 24 Suppl 3:11–18

Seksik P, Cosnes J, Sokol H, Nion-Larmurier I, Gendre JP, Beaugerie L. (2009) Incidence of benign upper respiratory tract infections, HSV and HPV cutaneous infections in inflammatory bowel disease patients treated with azathioprine. Aliment Pharmacol Ther 29:1106–1113

Sokol H, Lepage P, Seksik P, Doré J, Marteau P (2006) Temperature gradient gel electrophoresis of fecal 16S rRNA reveals active Escherichia coli in the microbiota of patients with ulcerative colitis. J Clin Microbiol 44: 3172–3177

Sokol H, Lepage P, Seksik P, Doré J, Marteau P (2007) Molecular comparison of dominant microbiota associated with injured versus healthy mucosa in ulcerative colitis. Gut 56:152–154

Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, Blugeon S, Bridonneau C, Furet JP, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas G, Blottiere HM, Dore J, Marteau P, Seksik P, Langella P (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci USA 105:16731–16736

Turnbaugh PJ, Backhed F, Fulton L, Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe 3:213–223

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2009) A core gut microbiome in obese and lean twins. Nature 457:480–484

van De Wijgert JH, Mason PR, Gwanzura L, Mbizvo MT, Chirenje ZM, Iliff V, Shiboski S, Padian NS (2000) Intravaginal practices, vaginal flora disturbances, and acquisition of sexually transmitted diseases in Zimbabwean women. J Infect Dis 181:587–594

van de Wijgert JH, Morrison CS, Cornelisse PG, Munjoma M, Moncada J, Awio P, Wang J, Van der Pol B, Chipato T, Salata RA, Padian NS (2008) Bacterial vaginosis and vaginal yeast, but not vaginal cleansing, increase HIV-1 acquisition in African women. J Acquir Immune Defic Syndr 48:203–210

Yazdanbakhsh M, Kremsner PG (2009) Influenza in Africa. PLoS Med 6:e1000182

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H, Mashiyama ST, Joachimiak

MP, van Belle C, Chandonia JM, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC (2007) The sorcerer II global ocean sampling expedition: expanding the universe of protein families. PLoS Biol 5:e16

Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB, Beja O (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the global ocean sampling expedition metagenomes. Environ Microbiol 9:1464–1475