

# Chapter 5

## Pathogen Genomics and the Potential for Understanding Diseases in the Developing World

Brenda Kwambana and Martin Antonio

### Introduction

Approximately 46% and 32% of deaths among children under age five globally occur in sub-Saharan Africa and South Asia, respectively. Over 80% of the 4.2 million child deaths in Africa are caused by infectious diseases, sharply contrasted to Europe where 39% of the 0.15 million child deaths are attributable to infectious diseases (Fig. 5.1) (Black et al. 2010). Hence, despite the remarkable public health advancements in hygiene, sanitation, antimicrobial drugs and vaccine strategies of the twenty-first century, the burden of infectious diseases remains unacceptably high in the developing world.

Morbidity and mortality due to infectious diseases have remained high in the developing world because several infections such as malaria have never been controlled and continue to persist, while other infectious diseases have emerged or re-emerged in the last few decades (Jones et al. 2008). Recently, e.g., there have been outbreaks of chikungunya (in South East Asia), dengue (in South America) and Ebola haemorrhagic fever (in Africa). The emergence and re-emergence of disease may be driven by several factors which include climatic changes that affect the survival of microbes and their vectors, modern agricultural practices, increased rates of travel and migration, high population densities and improved surveillance and detection of new threats.

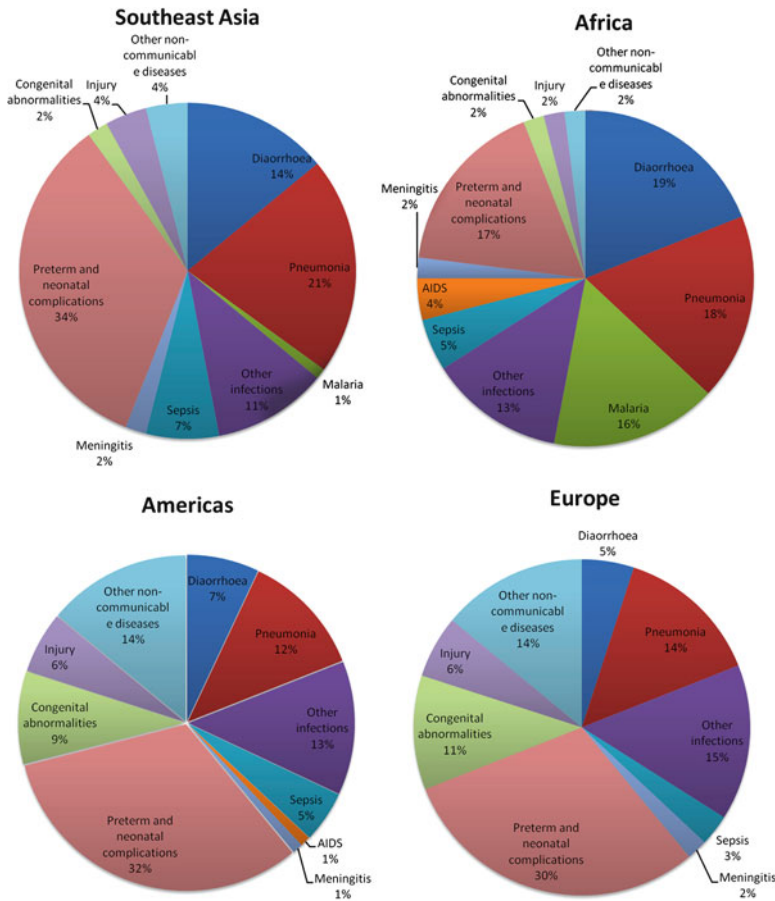
In addition, links between pathogens and diseases previously attributed to genetic, environmental and organic factors are being discovered, for instance, the link between *Helicobacter pylori* and peptic ulcers as well as gastric cancer (Dorer et al. 2009). The importance of disruptions of the normal microbial ecology in several autoimmune diseases such as necrotizing enterocolitis (NEC), atopic eczema,

---

B. Kwambana • M. Antonio (✉)

Bacterail Diseases Programme, Medical Research Council Laboratories (UK), Banjul, The Gambia

e-mail: mantonio@mrc.gm



**Fig. 5.1** Major causes of morbidity among children under five in South East Asia, the Americas, Africa and Europe (Black et al. 2010)

Crohn’s disease (CD) and chronic obstructive pulmonary disease (COPD) has also been demonstrated (Proal et al. 2009). Furthermore, it appears that some infections are polymicrobial (in that more than one microbe acts synergistically or sequentially to cause an infection) (Bakaletz 2004). Given the current prevalence and significance of infectious diseases caused by bacteria, fungi, protists and viruses in the developing world, the comprehensive characterization of the evolution, life cycle, physiological versatility and metabolism of pathogens is imperative.

Genomics (the study of genomes) facilitates the global analysis of pathogens and has huge potential to provide invaluable insights into various facets of pathogenesis. Genomics and genomics-derived applications have been used to study virulence, transmission, immune evasion, host-microbial interactions and microbial ecology in some of the most clinically significant pathogens in the developing world. Virulence is the degree of pathogenicity of an infectious agent, indicated by case mortality

rates and its ability to invade and damage host tissue. Virulence is mediated by several genetic factors which code for proteins involved in adhesion, colonization, toxin production and invasion. Whole-genome sequencing can be used in conjunction with microarray-based technologies, targeted genetic screens and proteomic analysis to discover virulence genes.

More recently, genomics has been used in developing rapid diagnostics tools as well as identifying vaccine and drug targets in several bacterial, protozoan and viral pathogens which could be adapted in resource-limited settings. The expansion of genomics has been fuelled by rapid advancements in DNA-sequencing technologies. Several faster, cheaper and higher throughput nucleic acid sequencing and analysis platforms have become available in the last few years and are more accessible than ever before, even in the developing countries (many are described in accompanying chapters throughout this book). Genomics tools have the advantage that they are not culture dependent and pathogen nucleic acids can be analysed long after the cells have lost viability. Hence, these tools can be invaluable in resource-limited settings whereby the reagents and equipment to support culture are not easily accessible (e.g. media, incubators and  $-20$ – $70^{\circ}\text{C}$  freezers).

Better understanding and characterization of the causes of morbidity will be essential in the achievement of the Millennium Development Goal 4 (MDG4) (Mittelmark 2009) to reduce childhood mortality by 35% between 1990 and 2015. This chapter focuses on how genomics tools have been applied to understand various aspects of human infectious diseases in the developing world and how these tools could be applied in future studies as part of the efforts to control and treat infectious diseases. Initially, the focus will be on the genomics application in the major infectious diseases in the developing world, i.e. pneumonia, meningitis, malaria, tuberculosis, the human immunodeficiency virus (HIV) and diarrhoeal diseases (Fig. 5.1). Thereafter, pathogen genomics of neglected tropical diseases (NTDs) will be discussed focusing on a few examples, Buruli ulcer, trachoma and trypanosomiasis.

## **Genomics for Major Infectious Diseases in the Developing World**

### ***Pneumonia and Meningitis***

Nearly a fifth of the eight million childhood deaths (<5 years) which occurred worldwide in 2008 were attributed to pneumonia, making it the most common cause of death among children (Rajaratnam et al. 2010). Approximately 50% of the severe pneumonia cases in developing countries are attributed to *Streptococcus pneumoniae* and *Haemophilus influenzae type b* (Hib) in areas where the vaccine is not widely available (Scott and English 2008). Likewise, these two pathogens are the major causes of non-epidemic meningitis among children in the developing world,

while *Neisseria meningitidis* is associated with epidemics, particularly in the Africa meningitis belt (Peltola 2001). Several pneumococcal virulence factors have been known for a long time and are fairly well characterized: the polysaccharide capsule, the toxin pneumolysin (PLY), choline binding proteins (CBPs) and the pilus, among many others (Mitchell and Mitchell 2010). In recent years, whole-genome analysis demonstrated the *S. pneumoniae* harbours several pathogenicity islands in its ~2-MB genome. Comparative genomics has shown that virulence may be associated with single-nucleotide polymorphisms (SNPs) or the absence of genetic islands, which encode for the virulence factors such as the pili and the polysaccharide capsule.

Comparative genomics of non-invasive, intermediately virulent and highly virulent *S. pneumoniae* serotype 1 showed that there are eight regions in the genome larger than 1 Kb unique to the highly virulent strains. Among them was identified a pathogenicity island encoding adherence factors, transporters and metabolic enzymes preferentially expressed in murine lungs and blood compared to nasopharyngeal tissue (Harvey et al. 2011). Whole-genome sequencing of a hypervirulent serotype 19A isolate showed the presence of a novel 20 Kb region which possibly encodes bacteriophage genes which enhance virulence (Thomas et al. 2011). Although *Streptococcus mitis* is the closest relative of *S. pneumoniae*, a comparison of the genomes showed that *S. mitis* lacks the CBPs, the hyaluronidase gene and ply which are important virulence factors in *S. pneumoniae* (Denapaite et al. 2010).

A strong association between the HLA-DR8 chromosomal loci and increased susceptibility to Hib disease among Alaskan Eskimo children has been demonstrated (Petersen et al. 1987). This suggests that host genomics could be an important determinant of the geographic distribution of Hib invasive disease. Gram-negative bacilli are present in the nasopharyngeal mucosa of more than half of Brazilian and Angolan children, in sharp contrast to 4% of Dutch children. Not surprisingly, Gram-negative bacilli are a common cause of pneumonia in Brazil and Angola as in many developing countries (Wolf et al. 1999; Obaro and Adegbola 2002). More specifically, genetic background or ethnicity could be directly associated with vulnerability to pneumonia, meningitis and other invasive bacterial diseases. For instance, in an arid region of Western Australia, *S. pneumoniae*, *Moraxella catarrhalis* and *Haemophilus influenzae* carriage is several-fold higher among Aboriginal than non-Aboriginal children, and this is reflected in the large disparity in the burden of pneumonia and meningitis in the two groups (Mackenzie et al. 2010). Nasopharyngeal carriage of respiratory pathogens among Chinese and Vietnamese children <5 years in Hong Kong was significantly different after adjusting for age, smoking and socio-economic conditions. Hence, it is plausible that host-genomics studies could reveal more interesting associations between host genetics and vulnerability to various pathogens, and this warrants more attention.

Pathogens can be transmitted across individuals through several mechanisms which include direct physical contact; ingestion of contaminated material, infected body fluids and secretions; inhalation of infected air droplets; and through a vector organism. Zoonotic infections are transmitted by direct and indirect mechanisms from animal hosts such as birds, pigs and bats to human hosts. Genomics applications can be used to track how pathogens are transmitted across individuals and identify potential

reservoirs within populations. For instance, multi-locus sequence typing (MLST) based on seven housekeeping genes of *S. pneumoniae* isolates from 158 West African villagers from 19 households showed strong evidence of non-random distribution of *S. pneumoniae* sequence types among households and that transmission most often occurs from children to other household members (Hill et al. 2010).

For survival and proliferation, pathogens have to establish themselves in a community competing for resources, nutrients and other growth factors (Brook 2007; Lysenko et al. 2005). Within an environment of polymicrobial immune stimulation, persisting microbes must also contend with the host's multilayered defences which constitute a barrage of bacteriostatic proteins, bactericidal factors, secretory immunoglobulins and a mucociliary clearance system. On the other hand, microorganisms, particularly those that have co-evolved with the human host, have developed several strategies to evade the mucosal immune responses. Mucosal immune responses to two closely related obligate colonizers of the URT mucosa, *Neisseria lactamica* and *N. meningitidis* appear distinct. In contrast to its pathogenic relative, *N. lactamica* does not induce the development of mucosal T- and B-cells in young individuals, maintaining immunological ignorance in the host which may facilitate the longer and more frequent colonization observed for *N. lactamica* strains. Host and pathogen genomics studies could be applied to elucidate why *N. lactamica* does not prime the mucosal immune system (Vaughan et al. 2009). This could be explained by findings from a comparative genome hybridization study of various unrelated *N. lactamica* strains to a pan-*Neisseria* microarray which showed that ~40% of *Neisseria* virulence genes are absent in the *N. lactamica* genome (Snyder and Saunders 2006).

Genomics-derived molecular tools can be used to accurately distinguish phenotypically, morphologically and biochemically identical pathogens at the bedside to effectively identify and treat infections. For instance, an infant was admitted to a teaching hospital in The Gambia from a primary health-care facility with clinical signs indicative of meningitis. *S. pneumoniae* serotype 14 was isolated from the cerebrospinal fluid examination (CSF) using standard laboratory culture and latex agglutination-based serotyping. The infection was treated with ceftriaxone and there was no evidence of neurological sequelae. However, within 28 days, the same infant was admitted to the Medical Research Council Unit with *S. pneumoniae* serotype 14 meningitis, and despite the antibiotic treatment administered, the infant suffered neurological sequelae from the second episode. Initially it appeared that this was a case of endogenous reactivation due to treatment failure during the first episode of meningitis. However, MLST analysis based on seven *S. pneumoniae* housekeeping genes was conducted on the two serotype 14 isolates revealing that they belonged to different and unrelated sequence types (ST), i.e. ST915 common in The Gambia from the first episode and a novel ST which belongs to a different clonal complex from the second episode. Hence, MLST analysis showed that this was not a relapse or recurrent infection by the same microbe, but rather an exogenous (different strain) infection (Antonio et al. 2009).

Molecular tools have the advantage that they are not culture dependent and pathogen nucleic acids can be analysed long after the cells have lost viability. In the

Pneumococcal Disease Surveillance in the West Africa Region (PneumoWAR) consortium, CSF samples from suspected meningitis cases that are culture negative or could not be cultured within the specified time are characterized by both species-specific and serotype-specific PCR. MLST and other genotyping techniques may also be used for further characterization of the pathogens directly from the CSF or purified nucleic acids. This has facilitated the identification of *S. pneumoniae*, *Haemophilus influenzae* type B and *N. meningitidis* in 20% of the culture-negative CSF samples received from the Paediatric Bacterial Meningitis Network (PBM). The data generated serves as an evidence base to support the implementation of lifesaving vaccines among governments, policymakers, pharmaceuticals and donor organizations.

## ***Tuberculosis***

Tuberculosis (TB) was previously considered a controlled infectious disease with the universal use of the Bacillus Calmette-Guérin (BCG) vaccine and antibiotic treatment. However, TB currently represents the most significant re-emerging infectious disease infecting a third of the world's population and causing more than two million deaths annually. TB is caused by *Mycobacterium tuberculosis* complex including *M. africanum* which causes up to 50% of TB in West Africa. The re-emergence of TB is partly attributable to the increasing incidence of multi-drug-resistant (MDR) TB and extensively drug-resistant (XDR) TB. Other important factors contributing to the re-emergence of TB are poor protection of the BCG vaccine in most regions and HIV co-infection. Mycobacteria genomics research has been crucial in the understanding of the evolution, adaptation and pathogenesis among pathogens causing TB and the development of unambiguous typing tools. The Mycobacteria genome is characterized by a lack of horizontal gene exchange and the presence of large sequence polymorphisms (LSPs), SNPs and several irreversible genomic deletions which are robust phylogenetically informative markers and useful for defining lineages within mycobacterial clusters. LSP- and SNP-based phylogenetic studies have shown that there are distinct geographical patterns in the distribution of *M. tuberculosis* globally. A few lineages are found in many different regions across the globe while others are restricted to certain geographic regions (de Jong et al. 2010).

The lack of horizontal gene transfer by mobile genetic elements such as plasmids suggests that drug resistance in Mycobacteria occurs from spontaneous chromosomal mutations. Molecular tools have been used to identify several of these mutations, for instance, streptomycin resistance is associated with mutations in the genes encoding ribosomal proteins, 16S rRNA and a 7-methylguanosine methyltransferase (*gidB*). Genotyping analysis has shown that specific mutations are associated with specific lineages, for instance, in one study, the *gidB*<sup>16</sup> polymorphism (16G allele) was exclusively found in the Latin American Mediterranean *M. tuberculosis* genotype. With the increasing burden of MDR and XDR TB in the developing world, the need for genomics research on the mechanisms of drug resistance and transmission has never been greater (Spies et al. 2011).

Molecular tools can also be adapted for the rapid diagnosis of infections (faster than by standard culture), even in resource-limited settings. Several rapid diagnostic systems have been developed, i.e. AccuProbe developed by Gen-Probe which uses chemiluminescent probes to detect nucleic acids of specific pathogens; the ResPlex system developed by Qiagen which uses microarray technology and detects amplified species-specific DNA sequences for several organisms; and the T500 biosensor (PCR/ESI-MS) which uses an MLST approach to simultaneously detect numerous pathogens, determine their clonality (relatedness) to other strains and identify resistance genotypes and virulence factors within a few hours. One of the most widely used rapid diagnosis systems is the GeneXpert system (Cepheid) which uses a real-time PCR system to identify infections in less than an hour. The GeneXpert system has been adapted for the identification of several pathogens including methicillin-resistant *Staphylococcus aureus* (MRSA), Group B *Streptococcus*, *Clostridium difficile* and *M. tuberculosis*. The Xpert MTB/RIF identifies *M. tuberculosis* using the *rpoB* gene ( $\beta$  subunit of the bacterial RNA polymerase) and rifampicin resistance which is a marker of MDR TB. The feasibility and effectiveness of this system were tested on sputum samples from individuals with persistent coughs (>2 weeks) at multiple centres in six developing countries, i.e. South Africa, Azerbaijan, Peru, the Philippines, India and Uganda, and *M. tuberculosis* could be detected in less than a day, compared to 16 and 30 days required for liquid and solid cultures, respectively. Likewise, the time to treatment for smear-negative tuberculosis was reduced from a median of 56 days to 5 days with the use of the MTB/RIF assay. Interestingly, the specificity and sensitivity of the assay were comparable among centres and countries with different capacities, indicating that the MTB/RIF system can be used effectively in resource-limited settings (Boehme et al. 2011). Genomics research in the developing world is useful in the rapid identification, characterization (virulence and antibiotic resistance) and diagnosis of emerging infectious diseases which is necessary for effective management and control of epidemics.

## **Malaria**

There is an estimated quarter of a billion cases of malaria annually in developing countries responsible for just under a million deaths annually. Approximately 10% of the cases occur in South East Asia, with the greatest burden of disease in Africa (85%). Malaria is caused by a protist parasite *Plasmodium* which invades and destroys erythrocytes (red blood cells) with detrimental effects to the host. This parasite is spread by the mosquito vectors *Anopheles gambiae* and *An. funestus*. *Plasmodium falciparum* causes the most deadly form of disease and about 90% of disease in sub-Saharan Africa. Another clinically significant species is *P. vivax* which causes up to 40% of malaria globally and is dominant in Asia as well as South and Central America. *P. malariae* and *P. ovale* are also important causes of malaria (Wells et al. 2009).

The emergence of *Plasmodium* resistance to older drugs such as chloroquine and newer artemisinin derivatives has necessitated research into new drug targets.

Whole genome sequencing of *P. falciparum* and *P. vivax* has fueled the rapid identification and characterization of a wide array of novel potential drug targets. Genes that encode essential pathways in the parasite which are considerably divergent from human homologues have presented good targets. A potential drug identified through genome analysis is fosmidomycin, which is currently in Phase II Clinical Trials for the treatment of malaria. Fosmidomycin blocks the activity of deoxyxylulose-5-phosphate reductoisomerase involved in the non-mevalonate isoprenoid biosynthesis pathway utilized by *Plasmodium* but is absent in humans and other mammals. Isoprenoids are essential multifunctional metabolites involved in cellular signalling and respiration; hence, inhibition of the synthesis pathway kills *P. falciparum* (Zhang et al. 2011). Other potential drug targets that have been identified through genomics applications and hold promise are the nucleoside biosynthesis pathway, dihydroorotate dehydrogenase and genes that occur in the apicoplast which is absent in humans (Wells et al. 2009).

Despite the massive disease control interventions implemented in many countries, the incidence of malaria has increased or remained static in some regions of Africa and Amazonia. One of the major challenges in malaria eradication is the large pool and persistence of the human reservoir of infection in endemic regions. Genome-wide SNP and microsatellite analyses have shown that *P. falciparum* chronic infections are characterized by numerous genetically diverse haploid parasite genomes established through a single mosquito bite or multiple bites (superinfection). High genetic diversity within *var*, a gene that encodes a major surface antigen of *P. falciparum*, is associated with immune evasion, establishment of chronic superinfections and human-to-mosquito transmission. These infections can persist for 6 months which allows parasites to persist during seasons when vectors are scarce (Chen et al. 2011). Comprehensive understanding of the mechanisms of immune evasion and virulence is important in the control of malaria and the development of more effective therapeutic agents and vaccine strategies.

The rapid and accurate diagnosis of malaria is critical in saving lives in malaria endemic areas where patients may present at health facilities with advanced, complicated or severe disease. Rapid and accurate diagnosis prevents unnecessary administration of antimalarial drugs associated with the emergence of drug resistance. Conventional malaria diagnosis in the developing world is conducted by light microscopy-based examination of thick and thin peripheral blood smears. However, this technique is technically demanding, and low-level parasitemia could be missed and mixed infections could be misdiagnosed. PCR-based parasite detection holds great promise as they can provide information about parasite species (including multiple infection), load and drug resistance fairly quickly. Before the *Plasmodium* genome sequence was available, the highly conserved 18S ribosomal RNA (18S rRNA) gene was used for the molecular detection and characterization of malaria infections using nested PCR. However, the 18S rRNA gene has few divergent copies which may be non-tandem in some species which affects the sensitivity and specificity of this gene target. As several genomes of different *Plasmodium* species have been sequenced, bioinformatics mining of the *Plasmodium* genomes has been used to identify novel PCR targets for the detection and speciation of malaria which have shown great promise (Demas et al. 2011).



## ***HIV/AIDS***

The human immunodeficiency virus (HIV) which causes acquired immune deficiency syndrome (AIDS) is one of the deadliest and most important threats to public health and economies globally. More than 20 million people have already died from AIDS in the last 25 years, and over 30 million people are living with HIV infection. Genomics-based evolution studies of HIV-1 suggest that this RNA lentivirus evolved from a simian immunodeficiency virus (SIV) which was transmitted from chimpanzees in West and Central Africa to humans at the beginning of the twentieth century. Similarly, HIV-2, the less aggressive relative of HIV-1, is thought to have evolved from SIV that infects sooty mangabeys in West Africa (Haagmans et al. 2009).

Although HIV-1 infections are widely distributed globally, HIV-2 infections show a pattern of restriction to West Africa perhaps due to lower transmissibility. HIV-2 infections are associated with lower plasma viral loads, decreased rates of mother-to-child transmission and reduced AIDS-related mortality (Gottlieb et al. 2006). There are numerous fully sequenced HIV genomes, and there is great potential that comparative genomics could elucidate the molecular basis for the differences observed in pathogenesis between the two types of HIV. More focus needs to be placed on linking the viral genome and differences observed in virulence among HIV subtypes. Few studies of HIV-1 (subtypes B and A) and SIV in macaque models have shown that the replication capacity of the lentiviruses increases over the course of infection characterized by more efficient utilization of the chemokine receptor 5 (CCR5). The increased replication capacity in viruses from late infection was conferred by sequence modifications in the V1–V5 envelope segments (Etemad et al. 2009).

In HIV research, significant focus has been on host-genomic factors that influence disease progression as opposed to pathogen-genomic factors. Geographic and genetic background-associated disparities in HIV-1 transmission and disease progression have been shown. For instance, nearly three quarters of the 30 million HIV-infected people are in sub-Saharan Africa and 4.2 million are in South Asia (Kaur and Mehra 2009a). Primary infection with HIV-1 may progress rapidly to AIDS within 3 years among fast progressors; however, among long-term non-progressors (LTNPs), disease progression is much slower. LTNPs maintain very low viral loads and normal immune function for decades. Another interesting group of individuals are those that are repeatedly and highly exposed to HIV-1 but remain seronegative and appear to harbour natural resistance.

These findings put together suggest that host factors that modulate immune evasion, viral entry and replication may be important determinants of disease outcome. The first HIV restriction factor to be identified is a 32-bp deletion allele in the cytoplasmic tail of the chemokine receptor 5 (CCR5 $\Delta$ 32) protein which is a secondary receptor on CD4 (+) T lymphocytes for HIV-1 and effectively blocks infection in homozygous individuals. The human leukocyte antigen (HLA) which has been shown to play an important role in modulating other infectious disease such as hepatitis B is another important factor in HIV infection (Vannberg et al. 2011). Specifically, certain SNPs in HLA-B and HLA-C are associated with lower viral

load setpoint and reduced CD4<sup>+</sup> T-cell decline. Most recently, genome-wide association studies of AIDS have shown a link between a SNP in the HLA region (*HCP5 rs2395029*) and viral load control among LTNPs (Limou et al. 2010). Other genome-wide association studies of several large cohorts have revealed at least 14 AIDS restriction genes which are involved in the regulation of viral cell entry, cytokine defences and other acquired and innate immune functions (Kaur and Mehra 2009b; O'Brien and Nelson 2004). The characterization of host factors that influence susceptibility to infection has been used as a platform for identifying novel antiretroviral drugs against HIV, which are much needed in the face of the frequent occurrence of resistant mutants. Recently, haematopoietic stem cells from a naturally resistant donor were transplanted in an infected patient. The patient has not required antiretroviral treatment and has maintained undetectable viral loads for 4 years which is a remarkable achievement that will impact the future of HIV control and treatment (Hutter and Thiel 2011).

## ***Diarrhoeal Diseases***

The World Health Organization (WHO) defines diarrhoea as the passage of loose or watery stools three or more times daily, and blood in stool could be indicative of acute diarrhoea or dysentery. Approximately 1.3 million deaths among children less than 5 years old are attributable to diarrhoea annually, and the greatest burden of disease is in South Asia and Africa where mortality rates can exceed 500 per 100,000 children. In the developing world, where the major infectious diarrhoea pathogens are viral, bacterial and protozoan, the primary modes of transmission are contaminated drinking water and food. The most common causes of infectious diarrhoea are rotavirus and enterotoxigenic *Escherichia coli*; however, *Shigella*, *Salmonella*, *Campylobacter* and *Yersinia* are important causes of infectious diarrhoea globally. *Vibrio cholera* causes highly infectious epidemic diarrhoea in regions where sanitation is poor (Thapar and Sanderson 2004). Here, the focus will be on *Shigella* (a bacterial pathogen) and rotavirus (a viral pathogen).

Rotavirus belongs to the Reoviridae family of RNA viruses and is the cause of approximately 60% of diarrhoeal episodes in developing countries annually. Most of the 400,000–600,000 deaths which occur yearly among children are in Africa and Asia. Genomics studies have shown that human rotavirus has huge genetic diversity conferred by the accumulation of mutations and gene rearrangements, frequent reassortment of the 11 genome segments. Zoonosis (transmission from animal reservoirs) of rotavirus also may increase antigenic diversity among human strains through the above mechanisms. Molecular epidemiology studies have shown that the circulating genotypes within a particular region can change over time and genotype diversity is uneven globally. For instance, in Europe, 70% of rotavirus isolated from humans were of the P[8]G1 genotype while numerous strains were found circulating in India (Greenberg and Estes 2009). Recently, whole genome sequencing of two Kenyan G2P[4] rotavirus strains showed that several genes were

distantly related to other human G2P[4] strains and the prototype DS-1 strain. These divergent genes from the Kenyan strains were instead more closely related to rotavirus genes of ruminant and camelid origin. These findings have significant implications for the application of the two currently licensed rotavirus vaccines (Rotateq from Merck and Rotarix from GlaxoSmithKline) which have excellent protective efficacy in high-income countries but have unequal efficacy in developing countries (Schaetti 2009). Future vaccine development will probably rely on extensive genomics studies to identify vaccine targets with broader specificities or the identification of region-specific protective targets.

The burden of bacillary dysentery which is primarily caused by *Shigella* is highest in the developing world among children under five. *Shigella* is caused by a Gram-negative bacterium, and the infections are highly contagious with less than 100 bacteria required to cause infections. Not surprisingly, there are an estimated 164 million cases and 1.1 million deaths associated with this pathogen annually. Invasion of the colonic mucosa by *Shigella* results in inflammation, ulceration and bleeding. In the pre-genomics era, non-motile *Shigella* which is an intracellular pathogen was regarded a distinct genus from *E. coli* which is an extracellular pathogen. However, phylogenetic studies based on typing of housekeeping genes, DNA hybridization and whole-genome sequencing showed that *Shigella* species and its various subtypes are taxonomically indistinguishable from *E. coli*. (Peng et al. 2009). *Shigella* species evolved from *E. coli* on multiple and distinct occasions which may explain the large genome diversity, antigenic variability and complex classification. Molecular analysis revealed the presence of a large virulence plasmid which encodes a type III secretion system (TTIS) and key regulatory genes associated with *Shigella* pathogenesis and niche (i.e. the human colonic mucosa) specific adaptation (Peng et al. 2006). Whole genome sequencing of several *Shigella* subtypes has shown that these bacteria have large sequence overlap with *E. coli* and that there are almost 1900 backbone open reading frames (ORFs). Genomics research provides the unprecedented capacity to develop novel vaccine strategies and identify specific drug targets from the backbone ORFs such as genes encoding essential signalling and metabolic and regulatory pathways.

Although the aetiology and pathogenesis of infectious diarrhoea is well characterized, molecular diagnostics remained a dark area until recently. The large number of possible aetiologic agents which represent very diverse organisms ranging from viral, bacterial to protozoan complicates the accurate identification of the causative pathogen and, thus, the administration of appropriate treatment. Furthermore, some of the pathogens such as *E. coli*, *Campylobacter* and *Salmonella* may be found in the gastrointestinal mucosa of asymptomatic individuals (Honey 2008). Diagnosis is particularly problematic in resource-limited settings where pathogen identification is primarily based on conventional laboratory culture of stool samples, which typically requires at least 48 h for bacterial pathogens and cannot discriminate pathogenic from non-pathogenic *E. coli* species.

In the last few years, a few diagnostic tests for infectious diarrhoea have become available, mostly based on the amplification of specific nucleotide sequences alone or in combination with microarray analysis and high-resolution melting analyses.

Multiple infections can be identified and accurate pathogen loads associated with disease can be determined. Molecular identification assays can be used to differentiate pathogens that are morphologically and biochemically indistinguishable such as Shiga toxin-producing and enteropathogenic strains of *E. coli* from avirulent strains (Bugarel et al. 2011). Important protozoan pathogens that are difficult to culture such as *Entamoeba histolytica*, *Cryptosporidium* and *Giardia* can be identified using molecular techniques targeting specific genes. The prevalence of these pathogens in the gastrointestinal mucosa was confirmed in large molecular epidemiology studies in Bangladesh (Haque et al. 2009). A recent study used metagenomic approaches to identify known diarrhoeal agents and to discover novel potential viral pathogens in stool samples from paediatric patients suffering from acute diarrhoea (Finkbeiner et al. 2008). The identification of new pathogens is particularly important because an aetiological agent is never found in as many as 40% of diarrhoea cases. Metagenomics is discussed in more detail in following sections. Although current metagenomic applications may still be inaccessible in many parts of the developing world with limited resources, data generated from these kinds of studies could be adapted to cheaper and more accessible applications for the diagnosis of infectious diarrhoea.

## Genomics Applications for Neglected Tropical Diseases (NTD)

Neglected tropical diseases (NTD) cover a wide spectrum of old and emerging diseases caused by helminths, protozoa, fungi, bacteria and viral pathogens that together infect an estimated one billion people but have received relatively little attention, warranting the name 'neglected'. About half of the people infected with NTDs are in sub-Saharan Africa. NTDs tend to occur in the poorest, most remote and disadvantaged regions of the world and are disabling, disfiguring and often times fatal. The pathogenesis, transmission and epidemiology of many of these pathogens are poorly understood, lacking appropriate diagnostic tools. Likewise, there are very few licensed vaccines against NTDs, and many available drugs are not optimally effective or ineffective against NTD infections. Hence, there is great potential for genomics to uncover the virulence factors, molecular epidemiology, genome organization and potential drug and vaccine targets of NTDs.

### *Trypanosomiasis*

Trypanosomiasis refers to several distinct diseases caused by the protozoan parasites trypanosomes. Human African trypanosomiasis (HAT) is a life-threatening infection caused by *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense* which infects 50,000–70,000 people in sub-Saharan Africa. In Latin America, trypanosomes cause Chagas diseases which is an important cause of

morbidity and mortality in that region. The sequencing of the trypanosomal genome has made remarkable contributions to the understanding of parasite pathogenesis, host-parasite interactions, mechanisms of genetic exchange, disease diagnosis, molecular epidemiology and drug mechanisms of action (Tait et al. 2011; Zucca and Savoia 2011). Current HAT treatment regimens are highly toxic and require prolonged periods of administration and monitoring which are problematic in resource-limited settings where the disease burden is greatest. Hence, there is a need for the identification of new drug targets which could be expedited by structural and comparative genomics. Promising targets and new drugs for the treatment of Chagas disease have already been identified using genomics and proteomic approaches, e.g. drugs that target the trypanosome purine salvage pathways, cruzipain (cysteine protease) and ergosterol and trypanothione biosynthesis are being investigated (McKerrow et al. 2009).

### ***Mycobacterial Infections***

Other than tuberculosis, mycobacteria also cause Buruli ulcer and leprosy which are disabling and disfiguring diseases infecting thousands of people in developing countries. Buruli ulcer is an emerging disease of the subcutaneous tissue that is caused by *Mycobacterium ulcerans* which is a slow-growing pathogen. *M. ulcerans* is closely related to *Mycobacterium marinum* which causes mild skin infections in humans. However, comparative genomics studies with its progenitor *M. marinum* have shown that *M. ulcerans* has large deletions of approximately 1.1 Mb in its genome and acquired a virulence plasmid that confers mycolactone production by horizontal gene transfer. The reduced genome size in *M. ulcerans* is associated with the loss of several virulence factors present in *M. marinum* and the acquisition of an immunosuppressive cytotoxin, making it a niche-adapted specialist (Demangel et al. 2009). The case of *M. ulcerans* highlights how evolutionary genomics can be used to understand pathogenesis in emerging infections, and the identification of key virulence factors sets the stage for future studies targeted at vaccine development.

### ***Trachoma***

The leading cause of infectious blindness globally is trachoma, with 63 million cases of active disease, and around 50% of the trachoma disease burden globally is concentrated in ten developing countries. Trachoma is caused by an obligate intracellular pathogen, *Chlamydia trachomatis*, which is also a leading cause of sexually transmitted infections and results in immense loss of productivity and represents a major public health and socio-economic burden. The need for an effective trachoma vaccine is greater than ever before because currently available antimicrobial treatments do not prevent re-infection and *C. trachomatis* is easily transmitted from

asymptomatic carriers to establish new infections. Chlamydial genomics research has revealed several protective antigens and epitopes that could be applied as effective vaccines (Eko et al. 2008).

## **Future Frontiers: Pathogen Genomics in the Developing World**

### ***Microbial Ecology and Disease***

Metagenomics refers to the collective culture-independent functional and/or sequence-based study of microbial genomes (termed the metagenome) contained in animal hosts, plants and environmental niches (Handelsman et al. 1998). Most recently, major advancements in next-generation DNA-sequencing tools with the capacity to handle mega metadata sets previously unimaginable and increased appreciation of the importance of complex microbial communities in health and disease have brought metagenomic approaches to the forefront of molecular microbiology and microbial ecology. Mucosal surfaces are the most common sites of microbial colonization and infection; these are the mucosae of the respiratory, digestive tract, urogenital tract, eye conjunctiva, the inner ear and exocrine gland ducts which together provide a surface area of 400 m<sup>2</sup> (Kunisawa et al. 2008). The term 'human microbiome' describes the microbial communities and their components that colonize the human body including the skin and mucosal surfaces (Petrosino et al. 2009). Host-microbe interactions are key components of normal human physiology, essential in the regulation of the inflammatory response, development of the immune system, metabolic function and nutrient processing, uptake and storage and inhibition of pathogens (Kelly and Conway 2005). Hence, the microbiome plays an important role not only in disease but also in the maintenance of health by enhancing or complementing host physiology and phenotype (Handelsman 2009).

Metagenomic tools have been used to show associations between the microbiome and several diseases such as NEC, atopic eczema, obesity, CD, COPD, cystic fibrosis, type II diabetes, oral cancer, bacterial vaginosis and periodontal disease among many others (Peterson et al. 2009). While some of these diseases are associated with disruptions or changes in normal microbial ecology, there is also evidence that some infections are polymicrobial with more than one microbe acting synergistically or sequentially to cause an infection. Endogenous microbial interactions in disease development are well illustrated in the pathogenesis of Burkitt's lymphoma and XDR-TB. Burkitt's lymphoma is a childhood cancer characterized by the proliferation of monoclonal B-cells; endogenous interactions between the malaria pathogen *P. falciparum* and the Epstein-Barr virus are believed to give rise to B-cell lymphoma and impaired immune surveillance (Rochford et al. 2005). XDR-TB is an emerging pathogen, primarily co-infecting people in the advanced stages of HIV infection (Goldman et al. 2007).

Although there is limited data, available reports suggest that the genomic and functional constitution of the pharyngeal microbiome markedly vary with geographic

location. The nasopharynx, in particular, is an important reservoir of commensal and pathogenic microbes which can migrate to and cause disease in other compartments such as the sinus, middle ear, lungs and blood (Buchanan et al. 1974). Invasive bacterial disease (IBD) such as pneumonia, meningitis and bacteraemia caused by pathogens that colonize the nasopharynx contribute to the disparity in childhood mortality between developing and developed countries. 16S rRNA gene-based terminal restriction fragment length polymorphism (T-RFLP) in conjunction with clone library sequencing was used to characterize the development of the nasopharyngeal microbiome among infants from a high-risk population in West Africa. This study showed that while most infants were co-colonized by multiple operational taxonomic units (OTUs) representing pathogenic species, *S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis* and *S. aureus*, a small proportion (<20%) of the infants did not or rarely harboured these OTUs. These findings were confirmed by PCR analysis. It was hypothesized that the structure and composition of the microbiome may be associated with the high rates of IBD caused by respiratory pathogens observed among infants from this population. Further studies comparing the infant nasopharyngeal microbiome in different geographic sites and among distinct populations are needed to understand the link between the microbiome and the risk of IBD.

The significance of the human microbiome in relation to health and disease has come to the forefront of microbiology, with several hundred millions poured into multi-centre international projects initiated in the last decade, most notably the National Institutes of Health (NIH) Human Microbiome Project (HMP). Metagenomic studies looking at different mucosal surfaces among various regions of the developing world will enhance our understanding of how to not only prevent disease but also to maintain health.

### ***New Vaccine Strategies***

There are still no licensed vaccines available against several important pathogens such as *S. aureus*, *Shigella*, *Moraxella catarrhalis*, HIV and *P. falciparum*. Some of the widely available vaccines such as BCG (against TB) are not optimally effective, and the licensed pneumococcal conjugate vaccines (PCVs) have limited valency, protecting against between 7 and 13 of the 94 serotypes, excluding many clinically significant serotypes in developing countries. Hence, there is great potential for vaccine and drug discovery at unprecedented speed in the developing world.

Novel bacterial antigens (subunit vaccines) could be identified by predicting the genes encoding potential surface-localized proteins from the sequenced genome, cloning the identified genes, expressing the proteins and testing their immunogenicity in vivo or in vitro. This has been applied to the identification of potential vaccine targets for *N. meningitidis* serogroup B, *Porphyromonas gingivalis*, *S. pneumoniae*, *Chlamydomphila pneumoniae* and *Pasteurella multocida*. Some of the antigens identified through whole genome sequencing have entered development or clinical phase. As numerous genomes of different strains of the same pathogenic species are rapidly

becoming available in public databases, highly conserved and universal antigens can be identified for inclusion in a vaccine formulation. Potential vaccine targets can also be identified by targeted searches of specific antigens within sequenced genomes. For instance, pilus-like structures were initially identified in *Corynebacterium diphtheriae* and have been shown to be good vaccine candidates in GBS; hence, researchers searched for similar genes in the closely related Group A Streptococcus (GAS) and *Streptococcus pneumoniae* genomes (Mora and Telford 2010). Formulations including multiple pilus-like structures have potential as vaccine candidates, with wide-coverage for these Gram-positive organisms. Although still in its infancy, genomics application in vaccine development may shorten the time to identifying novel antigens for new-generation vaccines compared to virulence-based approaches.

### *Novel Drug Targets*

Since the 1940s, numerous antimicrobials have been discovered from natural sources or have been developed synthetically, revolutionizing medicine and having massive impact on morbidity and mortality. However, in the last 20 years, pathogens have emerged with a barrage of defences against some of the most important antibiotics, raising serious public health concerns in both developed and developing countries. These pathogens include XDR/MDR TB, MRSA, vancomycin-resistant *Enterococcus*, *Pseudomonas aeruginosa* and extended-spectrum  $\beta$ -lactamase-producing *Enterobacteriaceae*. Hence, there is a need to develop new drug and treatment strategies that prevent colonization. Furthermore, there are still too few effective antimicrobial agents effective against viral pathogens (Brotz-Oesterhelt and Sass 2010).

The genomics approach can be used to discover new drugs by two major pathways: (1) screening the genomes of microbes which naturally produce antibiotics such as actinomycetes for bioactive compounds with therapeutic potential (Davies 2011) and (2) screening the pathogen genomes for highly conserved and universal pathways that could be potential drug targets. Potential drug targets also need to be essential for bacterial survival in the nutrient-rich environment in the host, and all the targeted pathogens must not have a mechanism to circumvent the targeted pathway. One such example of a potential drug discovered through large-scale genomic screening is the fatty acid biosynthesis inhibitor series, which targets FabI (enoyl-acyl carrier protein [ACP] reductase) and has entered clinical trials for treatment of *S. aureus* oral infections (Brotz-Oesterhelt and Sass 2010).

Although there has been relatively little success with novel antimicrobial drug discovery during the genomics era, genomics tools still hold great promise for future endeavours, particularly in the developing world. Theoretically, it could be possible to identify antimicrobials and antimicrobial targets from uncultured organisms. Metagenomic approaches targeting biosynthetic pathways could be used to screen complex microbial populations.



### ***Point of Care Diagnostics***

However, cutting-edge as well as simpler genomics-derived applications could and have been used at the point of care (bedside), particularly in diagnostics. Molecular diagnostics should be at the frontier of genomics in the developing world, with rapid and accessible assays that can be used with minimal resources.

Whole genome sequencing technologies are rapidly becoming cheaper, more accessible and faster. Hence, quite possibly, future diagnostics could employ whole genome sequencing, particularly with emerging infections and culture-negative infections for which molecular probes may not be available. Variations in genomes not represented by marker genes, oligonucleotides binding site polymorphisms and low depth of sampling may be overcome by whole genome sequencing. Sequencing of microbial genomes is a tool that provides comprehensive taxonomic and phylogenetic information not only about a pathogen but also its functional (biochemical and metabolic) properties (Petrosino et al. 2009). Whole genome amplification (WGA) based on ph29 polymerase-mediated multiple displacement amplification has been developed to generate billion-fold amplification of DNA from femtogram amounts, even from a single microbial cell (Siegl et al. 2011). However, annotation of genomes has huge computational demands, and analysis of taxonomic groups without previously sequenced genomes is a major challenge. Although still under development, WGA and single-cell genomics hold great promise for investigations for rapid and detailed diagnostic testing in the developing world. Comprehensive characterization of the ‘pan-genome’ (microbial evolution and population structure) has significant implications in the design of vaccines and treatment strategies against pathogens (Petrosino et al. 2009).

Rapid molecular diagnosis of infections is important for several reasons: (1) timely administration of the appropriate treatment and determination of drug resistance, (2) early management and control of highly infectious illnesses (prevent epidemics), (3) prevention of dropout and loss-to-follow-up with prolonged diagnostic periods and (4) high-throughput generation of molecular epidemiology data (detailed genotyping of the pathogen possible). Clinical specimens can be analysed in real time using automated systems requiring very little hands-on time. Hence, effective application of genomics applications has huge potential to reduce morbidity and mortality associated with misdiagnosis and delayed diagnosis. Nonetheless, the systems described above are not yet widely accessible in the developing world as the technology, resources and capacity are often not available. Hence, collaborations with developed countries, access to funding and information exchange need to be expanded to support the implementation of these systems in the developing world.

### ***Identification of New Threats***

Emerging infectious diseases are among the most important threats to public health and global economies. There has been a significant increase in the rates of emerging infectious diseases in the last 60 years, partly attributable to the HIV pandemic.

Approximately 60% are zoonotic and arise from wildlife (e.g. Ebola virus). Genomics tools have been developed for the rapid identification, classification, diagnosis and treatment of emerging and re-emerging microbial pathogens. The viral agent of the devastating SARs coronavirus was rapidly identified employing molecular virology/immunology and epidemiology techniques. The severe acute respiratory syndrome coronavirus (SARS-CoV) entire genome was sequenced shortly after its emergence in southern China in 2003, and comparative genomics were used to confirm its natural reservoir, the civet cat, which contributed to the effective control of the epidemic (Haagmans et al. 2009). Genomic analyses identified uncultured bacteria *Tropheryma whippelii* and *Bartonella henselae* as the causative agents of Whipple's disease and bacillary angiomatosis, respectively. Developing countries in the tropics have the highest relative risk of emerging infectious diseases from zoonotic pathogens from wildlife reservoirs and vector-borne pathogens. Hence, vigilant surveillance efforts should be strengthened in these regions.

When a pathogen crosses the species barrier, genomics studies are useful in the identification of genetic changes that enable the pathogen to interact with different receptors and establish infections in a new host species. Comparative genomics of SARs-CoV from civets and humans during the 2002–2004 outbreaks in China showed that SARs genomes were highly conserved between viruses from the two hosts. Genomic analyses also demonstrated that high genetic diversity in several critical genes, particularly the spike gene of SARs-CoV in civets, played a key role in increased civet-to-human and human-to-human transmission of the virus. Likewise, genomics studies have demonstrated that genetic variability of haemagglutinin molecules enables the influenza A virus to switch receptor and host specificity. For instance, mammalian influenzae A viruses preferentially bind oligosaccharides terminated with sialic acid- $\alpha$ -2,6-Gal, but avian H5N1 influenzae favour sialic acid- $\alpha$ -2,3-Gal disaccharide which are produced by human lower respiratory tract cells (Haagmans et al. 2009).

These examples highlight how pathogen-genomic studies can be utilized to understand the mechanisms that pathogens employ to increase transmissibility and cross species barriers. Most (60%) of emerging infectious diseases are zoonotic many of which originate in the developing world. Genomics can be applied to elucidate the genetic properties and systems that enhance transmissibility of a pathogen. For instance, the ability to persist in a population for relatively long periods of time may boost transmission of a pathogen from an individual to several contacts. Other ways of enhancing transmission include the capacity to cross species barriers and adapt to new hosts (e.g. migratory birds), survive in harsh environments (radiation, nutrient deprivation and high immune pressure) and adhere to a wide range of host cell types (Jones et al. 2008).

## Conclusions

Genomics is useful for epidemiologic characterization of known virulence factors and the discovery of novel virulence factors. Whole genome sequencing conducted on numerous microbial species has shown considerable variations in the size,

content and organization of the genomes of pathogens which may be associated with metabolic versatility and virulence which is invaluable for understanding disease in the developing world. Host and pathogen genomics can provide invaluable insights into the factors that influence the heterogeneity seen in transmission, vulnerability to infection, disease progression and disease outcomes among individuals from different age strata, ethnicities, socio-economic backgrounds and geographic regions exposed to a pathogenic microbe and inform better disease control and management strategies. As many genomics tools are not widely accessible in the developing world, fostering international collaborations will become increasingly important. Furthermore, funds to foster international collaborations, build capacity within resource-limited settings and the development of rapid, cheap and sustainable techniques that can be used in the developing world are urgently needed.

## References

- Antonio M, Oluwalana C, Secka O, Corrah T, Howie S, Adegbola RA (2009) Exogenous re-infection by a novel *Streptococcus pneumoniae* serotype 14 as a cause of recurrent meningitis in a child from The Gambia. *Ann Clin Microbiol Antimicrob* 8:3
- Bakaletz LO (2004) Developing animal models for polymicrobial diseases. *Nat Rev* 2(7):552–568
- Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG et al (2010) Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* 375(9730):1969–1987
- Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R et al (2011) Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet* 377(9776):1495–1505
- Brook I (2007) Current issues in the management of acute bacterial sinusitis in children. *Int J Pediatr Otorh* 71(11):1653–1661
- Brotz-Oesterhelt H, Sass P (2010) Postgenomic strategies in antibacterial drug discovery. *Future Microbiol* 5(10):1553–1579
- Buchanan RE, Gibbons NE, Bergey DH (1974) American society for microbiology. *Bergey's manual of determinative bacteriology*, 8th edn. Williams & Wilkins Co, Baltimore
- Bugarel M, Beutin L, Scheutz F, Loukiadis E, Fach P (2011) Identification of genetic markers for differentiation of Shiga toxin-producing, enteropathogenic, and avirulent strains of *Escherichia coli* O26. *Appl Environ Microbiol* 77(7):2275–2281
- Chen DS, Barry AE, Leliwa-Sytek A, Smith TA, Peterson I, Brown SM et al (2011) A molecular epidemiological study of var gene diversity to characterize the reservoir of *Plasmodium falciparum* in humans in Africa. *PLoS One* 6(2):e16629
- Davies J (2011) How to discover new antibiotics: harvesting the parvome. *Curr Opin Chem Biol* 15(1):5–10
- de Jong BC, Antonio M, Gagneux S (2010) *Mycobacterium africanum*—review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis* 4(9):e744
- Demangel C, Steiner TP, Cole ST (2009) Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*. *Nat Rev* 7(1):50–60
- Demas A, Oberstaller J, Debarry J, Lucchi NW, Srinivasamoorthy G, Sumari D et al (2011) Applied genomics: data mining reveals species-specific malaria diagnostic targets more sensitive than 18S rRNA. *J Clin Microbiol* 47(7):2411
- Denapaite D, Bruckner R, Nuhn M, Reichmann P, Henrich B, Maurer P et al (2010) The genome of *Streptococcus mitis* B6—what is a commensal? *PLoS One* 5(2):e9426

- Dorer MS, Talarico S, Salama NR (2009) *Helicobacter pylori*'s unconventional role in health and disease. *PLoS Pathog* 5(10):e1000544
- Eko FO, Talin BA, Lubitz W (2008) Development of a *Chlamydia trachomatis* bacterial ghost vaccine to fight human blindness. *Hum Vaccin* 4(3):176–183
- Etemad B, Fellows A, Kwambana B, Kamat A, Feng Y, Lee S et al (2009) Human immunodeficiency virus type 1V1-to-V5 envelope variants from the chronic phase of infection Use CCR5 and fuse more efficiently than those from early after infection. *J Virol* 83(19):9694–9708
- 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(2):e1000011
- Goldman RC, Plumley KV, Laughon BE (2007) The evolution of extensively drug resistant tuberculosis (XDR-TB): history, status and issues for global control. *Infect Disord Drug Targets* 7(2):73–91
- Gottlieb GS, Hawes SE, Agne HD, Stern JE, Critchlow CW, Kiviat NB et al (2006) Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: implications for differences in transmission. *AIDS (London)* 20(6):895–900
- Greenberg HB, Estes MK (2009) Rotaviruses: from pathogenesis to vaccination. *Gastroenterology* 136(6):1939–1951
- Haagmans BL, Andeweg AC, Osterhaus AD (2009) The application of genomics to emerging zoonotic viral diseases. *PLoS Pathog* 5(10):e1000557
- Handelsman J (2009) Microbial symbiosis: in sickness and in health. *DNA Cell Biol* 28(8):359–360
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5(10):R245–R249
- Haque R, Mondal D, Karim A, Molla IH, Rahim A, Faruque AS et al (2009) Prospective case–control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh. *Clin Infect Dis* 48(9):1191–1197
- Harvey RM, Stroehrer UH, Ogunniyi AD, Smith-Vaughan HC, Leach AJ, Paton JC (2011) A variable region within the genome of *Streptococcus pneumoniae* contributes to strain-strain variation in virulence. *PLoS One* 6(5):e19650
- Hill PC, Townend J, Antonio M, Akisanya B, Ebruke C, Lahai G et al (2010) Transmission of *Streptococcus pneumoniae* in rural Gambian villages: a longitudinal study. *Clin Infect Dis* 50(11):1468–1476
- Honey K (2008) Good bugs, bad bugs: learning what we can from the microorganisms that colonize our bodies. *J Clin Invest* 118(12):3817
- Hutter G, Thiel E (2011) Allogeneic transplantation of CCR5-deficient progenitor cells in a patient with HIV infection: an update after 3 years and the search for patient no. 2. *AIDS (London)* 25(2):273–274
- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL et al (2008) Global trends in emerging infectious diseases. *Nature* 451(7181):990–993
- Kaur G, Mehra N (2009a) Genetic determinants of HIV-1 infection and progression to AIDS: immune response genes. *Tissue Antigens* 74(5):373–385
- Kaur G, Mehra N (2009b) Genetic determinants of HIV-1 infection and progression to AIDS: susceptibility to HIV infection. *Tissue Antigens* 73(4):289–301
- Kelly D, Conway S (2005) Bacterial modulation of mucosal innate immunity. *Mol Immunol* 42(8):895–901
- Kunisawa J, Nochi T, Kiyono H (2008) Immunological commonalities and distinctions between airway and digestive immunity. *Trends Immunol* 29(11):505–513
- Limou S, Coulonges C, Herbeck JT, van Manen D, An P, Le Clerc S et al (2010) Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. *J Infect Dis* 202(6):908–915
- Lysenko ES, Ratner AJ, Nelson AL, Weiser JN (2005) The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. *PLoS Pathog* 1(1):e1
- Mackenzie GA, Leach AJ, Carapetis JR, Fisher J, Morris PS (2010) Epidemiology of nasopharyngeal carriage of respiratory bacterial pathogens in children and adults: cross-sectional surveys in a population with high rates of pneumococcal disease. *BMC Infect Dis* 10:304

- McKerrow JH, Doyle PS, Engel JC, Podust LM, Robertson SA, Ferreira R et al (2009) Two approaches to discovering and developing new drugs for Chagas disease. *Mem Inst Oswaldo Cruz* 104(Suppl 1):263–269
- Mitchell AM, Mitchell TJ (2010) *Streptococcus pneumoniae*: virulence factors and variation. *Clin Microbiol Infect* 16(5):411–418
- Mittelmark MB (2009) Millennium development goals. *Glob Health Promot* 16(4):3–4, 73–4/88–9
- Mora M, Telford JL (2010) Genome-based approaches to vaccine development. *J Mol Med* 88(2):143–147
- O'Brien SJ, Nelson GW (2004) Human genes that limit AIDS. *Nat Genet* 36(6):565–574
- Obaro S, Adegbola R (2002) The pneumococcus: carriage, disease and conjugate vaccines. *J Med Microbiol* 51(2):98–104
- Peltola H (2001) Burden of meningitis and other severe bacterial infections of children in Africa: implications for prevention. *Clin Infect Dis* 32(1):64–75
- Peng J, Zhang X, Yang J, Wang J, Yang E, Bin W et al (2006) The use of comparative genomic hybridization to characterize genome dynamics and diversity among the serotypes of *Shigella*. *BMC Genomics* 7:218
- Peng J, Yang J, Jin Q (2009) The molecular evolutionary history of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Infect Genet Evol* 9(1):147–152
- Petersen GM, Silimperi DR, Rotter JI, Terasaki PI, Schanfield MS, Park MS et al (1987) Genetic factors in *Haemophilus influenzae* type b disease susceptibility and antibody acquisition. *J Pediatr* 110(2):228–233
- Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA et al (2009) The NIH human Microbiome project. *Genome Res* 19(12):2317–2323
- Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. *Clin Chem* 55(5):856–866
- Proal AD, Albert PJ, Marshall T (2009) Autoimmune disease in the era of the metagenome. *Autoimmun Rev* 8(8):677–681
- Rajaratnam JK, Marcus JR, Flaxman AD, Wang H, Levin-Rector A, Dwyer L et al (2010) Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970–2010: a systematic analysis of progress towards millennium development goal 4. *Lancet* 375(9730):1988–2008
- Rochford R, Cannon MJ, Moormann AM (2005) Endemic Burkitt's lymphoma: a polymicrobial disease? *Nat Rev* 3(2):182–187
- Schaetti C (2009) Vaccines for enteric diseases: update on recent developments. *Expert Rev Vaccines* 8(12):1653–1655
- Scott JA, English M (2008) What are the implications for childhood pneumonia of successfully introducing Hib and pneumococcal vaccines in developing countries? *PLoS Med* 5(4):e86
- Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C et al (2011) Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. *ISME J* 5(1):61–70
- Snyder LA, Saunders NJ (2006) The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as 'virulence genes'. *BMC Genomics* 7:128
- Spies FS, Ribeiro AW, Ramos DF, Ribeiro MO, Martin A, Palomino JC et al (2011) Streptomycin resistance and lineage specific polymorphisms in *Mycobacterium tuberculosis* gidB gene. *J Clin Microbiol* 49(7):2625
- Tait A, Morrison LJ, Duffy CW, Cooper A, Turner CM, Macleod A (2011) Trypanosome genetics: populations, phenotypes and diversity. *Vet Parasitol* 18(1):61–68
- Thapar N, Sanderson IR (2004) Diarrhoea in children: an interface between developing and developed countries. *Lancet* 363(9409):641–653
- Thomas JC, Figueira M, Fennie KP, Laufer AS, Kong Y, Pichichero ME et al (2011) *Streptococcus pneumoniae* clonal complex 199: genetic diversity and tissue-specific virulence. *PLoS One* 6(4):e18649
- Vannberg FO, Chapman SJ, Hill AV (2011) Human genetic susceptibility to intracellular pathogens. *Immunol Rev* 240(1):105–116

- Vaughan AT, Gorringe A, Davenport V, Williams NA, Heyderman RS (2009) Absence of mucosal immunity in the human upper respiratory tract to the commensal bacteria *Neisseria lactamica* but not pathogenic *Neisseria meningitidis* during the peak age of nasopharyngeal carriage. *J Immunol* 182(4):2231–2240
- Wells TN, Alonso PL, Gutteridge WE (2009) New medicines to improve control and contribute to the eradication of malaria. *Nat Rev Drug Discov* 8(11):879–891
- Wolf B, Gama A, Rey L, Fonseca W, Roord J, Fleer A et al (1999) Striking differences in the nasopharyngeal flora of healthy Angolan, Brazilian and Dutch children less than 5 years old. *Ann Trop Paediatr* 19(3):287–292
- Zhang B, Watts KM, Hodge D, Kemp LM, Hunstad DA, Hicks LM et al (2011) A second target of the antimalarial and antibacterial agent fosmidomycin revealed by cellular metabolic profiling. *Biochemistry* 50(17):3570–3577
- Zucca M, Savoia D (2011) Current developments in the therapy of protozoan infections. *Open Med Chem J* 5:4–10