Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

10 Infectious Disease Genomics

*Yu-Tsueng Liu**

Division of Infectious Diseases, Department of Medicine, Moores Cancer Center, University of California San Diego, La Jolla, CA, USA

10.1 Introduction

The history and development of infectious disease genomics are tightly associated with the Human Genome Project (HGP) (Watson, 1990). A series of important discussions about the HGP were made in 1985 and 1986 (Dulbecco, 1986; Watson, 1990), which led to the appointment of a special National Research Council (NRC) committee by the National Academy of Sciences to address the needs and concerns, such as its impact, leadership, and funding sources. The committee recommended that the United States begin the HGP in 1988 (NRC, 1988). They emphasized the need for technological improvements in the efficiency of gene mapping, sequencing, and data analysis capabilities. In order to understand potential functions of human genes through comparative sequence analyses, they also advised that the HGP must not be restricted to the human genome and should include model organisms including mouse, bacteria, yeast, fruit fly, and worm. In the meantime, the Office of Technology Assessment (OTA) of the US Congress also issued a similar report to support the HGP (OTA, 1988). In 1990, the Department of Energy (DOE) and the National Institutes of Health (NIH) jointly presented an initial 5-year plan for the HGP (DHHS and DOE, 1990). In October 1993, the Sanger Center/Institute (Hinxton, UK) was officially open to join the HGP. The cost of DNA sequencing was about $2–5 per base in 1990, and the initial aim was to reduce the costs to less than $0.50 per base before large-scale sequencing (DHHS and DOE, 1990). The sequencing cost gradually declined during the subsequent years. In 2004, the National Human Genome Research Institute (NHGRI) challenged scientists to achieve a $100,000 human genome (3 Gb/haploid genome) by 2009 and a $1000 genome by 2014 to meet the need of genomic medicine.

The first complete genome to be sequenced was the phiX174 bacteriophage (5.4 kb) by Sanger's group in 1977 (Sanger et al., 1977). The complete genome sequence of SV40 polyomavirus (5.2 kb) was published in 1978 (Fiers et al., 1978; Reddy et al., 1978). The human Epstein–Barr virus (170 kb) genome was determined in 1984 (Baer et al., 1984). The first completed free-living organism genome was

*E-mail: ytliu@ucsd.edu

Genetics and Evolution of Infectious Diseases. DOI: 10.1016/B978-0-12-384890-1.00010-8
© 2011 Elsevier Inc. All rights reserved.

Haemophilus influenza (1.8 Mb), sequenced through a whole-genome shotgun approach in 1995 (Fleischmann et al., 1995). The second sequenced bacterial genome, *Mycoplasma genitalium* (600 kb), was completed in less than a month in the same year using the same approach (Smith, 2004). The DOE was the first to start a microbial genome program (MGP) as a companion to its HGP in 1994 (DOE, 2009). The initial focus was on nonpathogenic microbes. Along with the development of the HGP, there was exponential growth of the number of completely sequenced free-living organism genomes. The Fungal Genome Initiative (FGI) (FGI, 2010) was established in 2000 to accelerate the slow pace of fungal genome sequencing since the report of the genome of *Saccharomyces cerevisiae* in 1996 (Goffeau et al., 1996). One of the major interests was to sequence organisms that are important in human health and commercial activities. As of September 2009, 1100 completed genome projects, a 1.7-fold increase from 2 years ago, were documented (Liolios et al., 2010). These include 914 bacterial, 68 archaeal, and 118 eukaryotic genomes. In addition, more than 4000 other ongoing sequencing projects were reported.

The genomes of human malaria parasite *Plasmodium falciparum* and its major mosquito vector *Anopheles gambiae* were published in 2002 (Gardner et al., 2002; Holt et al., 2002). The effort to sequence the malaria genome began in 1996 by taking advantage of a clone derived from laboratory-adapted strain (Hoffman et al., 1997). Many parasites have complex life cycles that involve both vertebrate and invertebrate hosts and are difficult to maintain in the laboratory. Currently, a few other important human pathogenic parasites, such as Trypansomes (Berriman et al., 2005; El-Sayed et al., 2005), Leishmania (Ivens et al., 2005), and Schistosomas (Berriman et al., 2009; Consortium, 2009), have been either completely or partially sequenced (Brindley et al., 2009; Aurrecoechea et al., 2010). In the meantime, the genome sequence of *Aedes aegypti*, the primary vector for yellow fever and dengue fever, was published in 2007 (Nene et al., 2007). The genome size (1376 Mb) of this mosquito vector is about 5 times larger than the previously sequenced genome of the malaria vector *Anopheles gambiae*. Approximately 50% of the genome consists of transposable elements. In 2010, the genome sequence of the body louse (*Pediculus humanus humanus*), an obligatory parasite of humans and the main vector of epidemic typhus (*Rickettsia prowazekii*), relapsing fever (*Borrelia recurrentis*), and trench fever (*Bartonella quintana*), was reported (Kirkness et al., 2010). Its 108 Mb genome is the smallest among the known insect genomes. Genome sequencing projects for other important human disease vectors are in progress (Lawson et al., 2009; Megy et al., 2009). These include *Culex pipiens* (mosquito vector of West Nile virus), *Ixodes scapularis* (tick vector of Lyme disease, babesia, and anaplasma), and *Glossina morsitans* (tsetse fly vector of African trypanosomiasis). The challenge to sequence the genome of an insect vector is much greater than a microbe. For example, the genomes of ticks were estimated to be between 1 and 7 Gb and may have a significant proportion of repetitive DNA sequences, which may be a problem for genome assembly (Pagel Van Zee et al., 2007). Furthermore, the evolutionary distances among insect species may also affect homology-based gene predictions.

It is as important to understand the sequence diversity within a species as to perform a de novo sequencing of a reference genome from the perspective of human

health. This is true for both hosts and pathogens (Feero et al., 2008; Alcais et al., 2009). The goal of the 1000 Genomes Project is to find most genetic variants that have frequencies of at least 1% in the human populations studied (Kaiser, 2008). One of the similar efforts for human pathogens is the NIH Influenza Genome Sequencing Project. When this project began in November 2004, only seven human influenza H3N2 isolates had been completely sequenced and deposited in the GenBank database (Fauci, 2005; Ghedin et al., 2005). As of May 2010, more than 5000 human and avian isolates have been completely sequenced, including the 1918 "Spanish" influenza virus (Taubenberger et al., 2005). Databases for human immunodeficiency virus (HIV) and hepatitis C virus have also been established.

While most human studies of microbes have focused on the disease-causing organisms, interest in resident microorganisms has also been growing. In fact, it has been estimated that the human body is colonized by at least 10 times more prokaryotic and eukaryotic microorganisms than the number of human cells (Savage, 1977). It was suggested to have "the second human genome project" to sequence human microbiome (Relman and Falkow, 2001). Highly variable intestinal microbial flora among normal individuals has been well documented (Eckburg et al., 2005; Costello et al., 2009; Turnbaugh et al., 2009). Therefore, the Human Microbiome Project (HMP) was initiated by the NIH to study samples from multiple body sites from each of at least 250 "normal" volunteers to determine whether there are associations between changes in the microbiome and several different medical conditions, and to provide both standardized data resources and new technological approaches (Peterson et al., 2009).

The completed or ongoing genome projects (Table 10.1) will provide enormous opportunities for the discovery of novel vaccines and drug targets against human pathogens as well as the improvement of diagnosis and discovery of infectious agents and the development of new strategies for invertebrate vector control. Specific examples will be provided to illustrate how the information provided by various genome projects may help achieve the goal of promoting human health.

10.2 Vaccine Target

Meningococcal isolates produce 1 of 13 antigenically distinct capsular polysaccharides, but only 5 (A, B, C, W135, and Y) are commonly associated with disease (Lo et al., 2009). The polysaccharide capsule is important for meningococci to escape from complement-mediated killing. While conventional vaccines consisting of the conjugation of capsular polysaccharides to carrier proteins for meningococcus serogroups A, C, Y, and W-135 have been clinically successful, the same approach failed to produce clinically useful vaccine for serogroup B (MenB). The capsule polysaccharide (α2-8 *N*-acetylneuraminic acid) of MenB is identical to human polysialic acid and therefore is poorly immunogenic (Finne et al., 1987). Alternatively, vaccines consisting of outer membrane vesicles (OMV) have been successfully developed to control MenB outbreaks in areas where epidemics are dominated by one particular strain (Bjune et al., 1991; Sierra et al., 1991; Boslego

Table 10.1 The Completed or Ongoing Genome Projects

General
NCBI (Sayers et al., 2010) (http://www.ncbi.nlm.nih.gov/sites/genome)
ENSEMBL (Kersey et al., 2010) (http://www.ensemblgenomes.org/)
JCVI (Davidsen et al., 2010) (http://cmr.jcvi.org/)
GOLD (Liolios et al., 2010) (http://www.genomesonline.org)
Sanger Pathogen Genomics (http://www.sanger.ac.uk/Projects/Pathogens/)
GeMInA (Genomic Metadata for Infectious Agents) (Ecker et al., 2005; Schriml et al., 2010) (http://gemina.igs.umaryland.edu)
Bacteria
HMP (Nelson et al., 2010) (http://www.hmpdacc.org/)
Fungi
FGI (http://www.broadinstitute.org/science/projects/fungal-genome-initiative)
Parasites
Eukaryotic pathogens (Aurrecoechea et al., 2010) (http://EuPathDB.org)
Parasite genome projects (http://www.pasteur.fr/recherche/unites/tcruzi/minoprio/genomics/parasites.htm)
Invertebrate vectors
VectorBase (Lawson et al., 2009; Megy et al., 2009) (http://www.vectorbase.org)
Viruses
Influenza virus (Bao et al., 2008) (http://www.ncbi.nlm.nih.gov/genomes/FLU/)
HIV (http://www.hiv.lanl.gov/)
HCV (http://hcv.lanl.gov/)

et al., 1995; Jackson et al., 2009). The most significant limitation of this type of vaccine is that the immune response is strain-specific, mostly directed against the porin protein, PorA, which varies substantially in both expression level and sequence across strains (Martin et al., 2000; Pizza et al., 2000).

With the completion of the genome sequence of a virulent MenB strain, a "reverse vaccinology" approach was applied for the development of a universal MenB vaccine by Novartis (Pizza et al., 2000; Tettelin et al., 2000; Giuliani et al., 2006). Through bioinformatic searching for surface-exposed antigens, which may be the most suitable vaccine candidates due to their potential to be readily recognized by the immune system, 570 open reading frames (ORFs) were selected from a total of 2158 ORFs of the MC58 genome. Eventually, five antigens were chosen as the vaccine components based on a series of criteria including the ability of candidates to be expressed in *Escherichia coli* as recombinant proteins (350 candidates), the confirmation of surface exposure by immunological analyses, the ability of induced protective antibodies in experimental animals (28 candidates), and the conservation of antigens within a panel of diverse meningococcal strains, primarily the disease-associated MenB strains (Pizza et al., 2000; Giuliani et al., 2006; Rinaudo et al., 2009). The vaccine formulation consists of an fHBP-GNA2091 fusion protein, a GNA2132-GNA1030 fusion protein, NadA, and OMV from the New Zealand MeNZB vaccine strain, which contains the immunogenic PorA. Initial phase II clinical results in adults and infants showed that this vaccine could induce a protective immune response against three diverse MenB strains in

89–96% of subjects following three vaccinations and 93–100% after four vaccinations (Rinaudo et al., 2009). In 2010, a phase III trial for this vaccine (4CMenB) has met primary endpoint.

10.3 Drug Target

Targeting an essential pathway is a necessary but not sufficient requirement for an effective antimicrobial agent (Brinster et al., 2009). Identification of essential genes in a completely sequenced genome has been actively pursued with various approaches (Hutchison et al., 1999; Ji et al., 2001). The indispensable fatty acid synthase (FAS) pathway in bacteria has been regarded as a promising target for the development of antimicrobial agents (Wright and Reynolds, 2007). The subcellular organization of the fatty acid biosynthesis components is different between mammals (type I FAS) and bacteria (dissociated type II FAS), which raises the likelihood of host specificity of the targeting drugs. Comparison of the available genome sequences of various species of prokaryotes reveals highly conserved FAS II systems suggesting that the antimicrobial agent can be broad spectrum (Zhang et al., 2003). In addition, through computational analyses, new members of the FAS II system have been discovered in different bacterial species (Heath and Rock, 2000; Marrakchi et al., 2002). One of the protein components in this system, FabI, is the target of an anti-tuberculosis drug isoniazid and a general antibacterial and antifungal agent, triclosan (Banerjee et al., 1994; Levy et al., 1999; Zhang et al., 2006).

Through a systematic screening of 250,000 natural product extracts, a Merck team identified a potent and broad-spectrum antibiotic, platensimycin, which is derived from *Streptomyces platensis* and a selective FabF/B inhibitor in FAS II system (Wang et al., 2006). Treatment with platensimycin eradicated *Staphylococcus aureus* infection in mice. Platensimycin did not have cross-resistance to other antibiotic-resistant strains in vitro, including methicillin-resistant *S. aureus*, vancomycin-intermediate *S. aureus*, and vancomycin-resistant enterococci. No toxicity was observed using a cultured human cell line. The activity of platensimycin was not affected by the presence of human serum in this study. However, the FAS II system appears to be dispensable for another Gram-positive bacterium, *Streptococcus agalactiae*, when exogenous fatty acids are available, such as in human serum (Brinster et al., 2009; Balemans et al., 2010). The susceptibility to inhibitors targeting the FAS II system indicates heterogeneity in fatty acid synthesis or in acquiring exogenous fatty acids among Gram-positive pathogens (Balemans et al., 2010). Comparative genomic approaches may be useful to identify and develop a strategy to target the salvage pathway for *Streptococcus agalactiae*. Alternatively, similar approaches as described earlier for MenB vaccine may also be applied for *Streptococcus agalactiae* (Group B streptococcus) (Maione et al., 2005).

10.4 Vector Control

An early mathematical model for malaria control suggested that the most vulnerable element in the malaria cycle was survivorship of adult female mosquitoes

(Macdonald, 1957; Enayati and Hemingway, 2010). Therefore, insect control is an important part of reducing transmission. The use of DDT as an indoor residual spray in the global malaria eradication program from 1957 to 1969 reduced the population at risk of malaria to ~50% by 1975 compared with 77% in 1900 (Hay et al., 2004; Enayati and Hemingway, 2010). Engineering genetically modified mosquitoes refractory to malaria infection appeared to be an alternative approach (Curtis, 1968) given the environmental impact of DDT and the emergence of insecticide-resistant insects. The Vector Biology Network (VBN) was formed in 1989 and proposed a 20-year plan with the World Health Organization (WHO) in 2001 to achieve three major goals: (1) to develop basic tools for the stable transformation of anopheline mosquitoes by the year 2000; (2) to engineer a mosquito incapable of carrying the malaria parasite by 2005; and (3) to run controlled experiments to test how to drive the engineered genotype into wild mosquito populations by 2010 (Alphey et al., 2002; Morel et al., 2002; Beaty et al., 2009). While some proof-of-concept experiments were achieved for the first two aims in 2002 when the *Anopheles gambiae* genome was completely sequenced (Catteruccia et al., 2000; Ito et al., 2002), the progress has been relatively slow (Marshall and Taylor, 2009).

Genomic loci of the *Anopheles gambiae* responsible for *Plasmodium falciparum* resistance have been identified through surveying a mosquito population in a West African malaria transmission zone (Riehle et al., 2006). A candidate gene, Anopheles Plasmodium-responsive leucine-rich repeat 1 (*APL1*), was discovered. Subsequently, other resistant genes have also been identified (Blandin et al., 2009; Povelones et al., 2009). Studying the genetic basis of resistance to malaria parasites and immunity of the mosquito vector will be important to control malaria transmission.

10.5 Diagnostic Target and Pathogen Discovery

Perhaps the most immediate impact of a completely sequenced pathogen genome is for infectious disease diagnosis. The information may be of great importance to the public health when a newly emerged or re-emerged pathogen is discovered. The 2009 swine-origin influenza A virus (S-OIV) (Dawood et al., 2009) and 2003 SARS (severe acute respiratory syndrome) coronavirus (Ksiazek et al., 2003; Rota et al., 2003) are the two most recent examples.

S-OIV emerged in the spring of 2009 in Mexico and was also discovered in specimens from two unrelated children in the San Diego area in April 2009 (CDC, 2009; Dawood et al., 2009). Those samples were positive for influenza A but negative for both human H1 and H3 subtypes. The complete genome sequence and a real-time PCR-based diagnostic assay were released to the public in late April. The outbreak evolved rapidly and the WHO declared the highest Phase 6 worldwide pandemic alert on June 11, 2009. S-OIV has three genome segments (HA, NP, NS) from the classic North American swine (H1N1) lineage, two segments (PB2, PA) from the North American avian lineage, one segment (PB1) from the seasonal H3N2, and most notably, two segments (NA, M) from the Eurasian swine (H1N1) lineage (Dawood et al., 2009). With the available influenza genome database,

diagnostic assays to distinguish previous seasonal H1N1, H3N2, and S-OIV can be easily accomplished (Lu et al., 2009).

A comprehensive pathogen genome database is not only useful for infectious disease diagnosis but also for novel pathogen discovery (Liu, 2008). Homologous sequences within the same family or among different family members are important for new pathogen identification even with the advent of third-generation sequencing technology (Munroe and Harris, 2010). De novo pathogen discovery may be also complicated by coexisting microorganisms, such as commensal bacteria in the human body. Without prior knowledge of these microorganisms, one may be misled.

In 2003, a microarray-based assay, designated Virochip, was used to help discover the SARS coronavirus (Wang et al., 2003). The Virochip contained the most highly conserved 70mer sequences from every fully sequenced reference viral genome in GenBank. The computational search for conservation was performed across all known viral families. A microarray hybridized with a reaction derived from a viral isolate cultivated from a SARS patient revealed that the strongest hybridizing array elements belong to families Astroviridae and Coronaviridae. Alignment of the oligonucleotide probes having the highest signals showed that all four hybridizing oligonucleotides from the Astroviridae and one oligonucleotide from avian infectious bronchitis virus, an avian coronavirus, shared a core consensus motif spanning 33 nucleotides. Interestingly, it had been known previously through bioinformatic analyses that this sequence is present in the 3′ UTR of all astroviruses, avian infectious bronchitis virus, and an equine rhinovirus (Jonassen et al., 1998). Therefore, a new member of the coronavirus was identified through the unique hybridizing pattern and subsequent confirmations.

The finding of the seventh human oncogenic virus, Merkel cell polyomavirus (MCV) (Feng et al., 2008) in 2008 is another example of why conserved sequences are important for novel pathogen discovery. MCV is the etiological agent of Merkel cell carcinoma (MCC), which is a rare but aggressive skin cancer of neuroendocrine origin. Two cDNA libraries derived from MCC tumors were subjected to high-throughput sequencing by a next-generation Roche/454 sequencer. Nearly 400,000 sequence reads were generated. The majority (99.4%) of the sequences derived from human origin were removed from further analyses. Only one of the remaining 2395 cDNA was homologous to the T antigen of two known polyomaviruses. One additional cDNA was subsequently identified to be part of the MCV sequence when the complete viral sequence was known. Later analyses showed that 80% (8/10) of the MCC had integrated MCV in the human genome. Monoclonal viral integration was revealed by the patterns of Southern blot analysis. Only 8−16% of control tissues had low copy number of MCV infection.

10.6 Conclusion

While we can expect that the efforts of a variety of genome projects may improve human health, the socioeconomic issues that are not discussed in this chapter may be substantial. In addition, the tremendous amount of information derived from

these projects will also be a challenge for scientists as well nonscientists to follow and understand.

References

Alcais, A., Abel, L., Casanova, J.L., 2009. Human genetics of infectious diseases: between proof of principle and paradigm. J. Clin. Invest. 119, 2506−2514.

Alphey, L., Beard, C.B., Billingsley, P., Coetzee, M., Crisanti, A., Curtis, C., et al., 2002. Malaria control with genetically manipulated insect vectors. Science 298, 119−121.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Fischer, S., Gajria, B., Gao, X., et al., 2010. EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res. 38, D415−419.

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., et al., 1984. DNA sequence and expression of the B95-8 Epstein−Barr virus genome. Nature 310, 207−211.

Balemans, W., Lounis, N., Gilissen, R., Guillemont, J., Simmen, K., Andries, K., et al., 2010. Essentiality of FASII pathway for *Staphylococcus aureus*. Nature 463, E3, discussion E4.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K.S., Wilson, T., et al., 1994. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science 263, 227−230.

Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., et al., 2008. The influenza virus resource at the National Center for Biotechnology Information. J. Virol. 82, 596−601.

Beaty, B.J., Prager, D.J., James, A.A., Jacobs-Lorena, M., Miller, L.H., Law, J.H., et al., 2009. From Tucson to genomics and transgenics: the vector biology network and the emergence of modern vector biology. PLoS Negl. Trop. Dis. 3, e343.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., et al., 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science 309, 416−422.

Berriman, M., Haas, B.J., LoVerde, P.T., Wilson, R.A., Dillon, G.P., Cerqueira, G.C., et al., 2009. The genome of the blood fluke *Schistosoma mansoni*. Nature 460, 352−358.

Bjune, G., Hoiby, E.A., Gronnesby, J.K., Arnesen, O., Fredriksen, J.H., Halstensen, A., et al., 1991. Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. Lancet 338, 1093−1096.

Blandin, S.A., Wang-Sattler, R., Lamacchia, M., Gagneur, J., Lycett, G., Ning, Y., et al., 2009. Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. Science 326, 147−150.

Boslego, J., Garcia, J., Cruz, C., Zollinger, W., Brandt, B., Ruiz, S., et al., 1995. Efficacy, safety, and immunogenicity of a meningococcal group B (15:P1.3) outer membrane protein vaccine in Iquique, Chile. Chilean National Committee for Meningococcal Disease. Vaccine 13, 821−829.

Brindley, P.J., Mitreva, M., Ghedin, E., Lustigman, S., 2009. Helminth genomics: the implications for human health. PLoS Negl. Trop. Dis. 3, e538.

Brinster, S., Lamberet, G., Staels, B., Trieu-Cuot, P., Gruss, A., Poyart, C., 2009. Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. Nature 458, 83−86.

Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., et al., 2000. Stable germline transformation of the malaria mosquito *Anopheles stephensi*. Nature 405, 959–962.

CDC, 2009. Swine influenza A (H1N1) infection in two children—Southern California, March–April 2009. MMWR Morb. Mortal. Wkly Rep. 58, 400–402.

Consortium, 2009. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. Nature 460, 345–351.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., Knight, R., 2009. Bacterial community variation in human body habitats across space and time. Science 326, 1694–1697.

Curtis, C.F., 1968. Possible use of translocations to fix desirable genes in insect pest populations. Nature 218, 368–369.

Davidsen, T., Beck, E., Ganapathy, A., Montgomery, R., Zafar, N., Yang, Q., et al., 2010. The comprehensive microbial resource. Nucleic Acids Res. 38, D340–D345.

Dawood, F.S., Jain, S., Finelli, L., Shaw, M.W., Lindstrom, S., Garten, R.J., et al., 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N. Engl. J. Med. 360, 2605–2615.

DHHS and DOE, 1990. Understanding our genetic inheritance, the U.S. Human Genome Project: the first five years: fiscal years 1991–1995. <http://www.ornl.gov/sci/techresources/Human_Genome/project/5yrplan/summary.shtml>.

DOE, 2009. Microbial Genome Program. <http://microbialgenomics.energy.gov/mgp.shtml>.

Dulbecco, R., 1986. A turning point in cancer research: sequencing the human genome. Science 231, 1055–1056.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., et al., 2005. Diversity of the human intestinal microbial flora. Science 308, 1635–1638.

Ecker, D.J., Sampath, R., Willett, P., Wyatt, J.R., Samant, V., Massire, C., et al., 2005. The microbial rosetta stone database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. BMC Microbiol. 5, 19.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.N., et al., 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309, 409–415.

Enayati, A., Hemingway, J., 2010. Malaria management: past, present, and future. Annu. Rev. Entomol. 55, 569–591.

Fauci, A.S., 2005. Race against time. Nature 435, 423–424.

Feero, W.G., Guttmacher, A.E., Collins, F.S., 2008. The genome gets personal—almost. JAMA 299, 1351–1352.

Feng, H., Shuda, M., Chang, Y., Moore, P.S., 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319, 1096–1100.

FGI (2010). Fungal Genome Initiative. <http://www.broadinstitute.org/science/projects/fungal-genome-initiative>.

Fiers, W., Contreras, R., Haegemann, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., et al., 1978. Complete nucleotide sequence of SV40 DNA. Nature 273, 113–120.

Finne, J., Bitter-Suermann, D., Goridis, C., Finne, U., 1987. An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. J. Immunol. 138, 4402–4407.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., et al., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496–512.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., et al., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419, 498−511.

Ghedin, E., Sengamalay, N.A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., et al., 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 437, 1162−1166.

Giuliani, M.M., Adu-Bobie, J., Comanducci, M., Arico, B., Savino, S., Santini, L., et al., 2006. A universal vaccine for serogroup B meningococcus. Proc. Natl. Acad. Sci. U.S.A. 103, 10834−10839.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., et al., 1996. Life with 6000 genes. Science 274, 546, 563−567.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M., Snow, R.W., 2004. The global distribution and population at risk of malaria: past, present, and future. Lancet Infect. Dis. 4, 327−336.

Heath, R.J., Rock, C.O., 2000. A triclosan-resistant bacterial enzyme. Nature 406, 145−146.

Hoffman, S.L., Bancroft, W.H., Gottlieb, M., James, S.L., Burroughs, E.C., Stephenson, J.R., et al., 1997. Funding for malaria genome sequencing. Nature 387, 647.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., et al., 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298, 129−149.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., et al., 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. Science 286, 2165−2169.

Ito, J., Ghosh, A., Moreira, L.A., Wimmer, E.A., Jacobs-Lorena, M., 2002. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature 417, 452−455.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., et al., 2005. The genome of the kinetoplastid parasite, *Leishmania major*. Science 309, 436−442.

Jackson, C., Lennon, D.R., Sotutu, V.T., Yan, J., Stewart, J.M., Reid, S., et al., 2009. Phase II meningococcal B vesicle vaccine trial in New Zealand infants. Arch. Dis. Child. 94, 745−751.

Ji, Y., Zhang, B., Van, S.F., Horn, W.P., Woodnutt, G., Burnham, M.K., et al., 2001. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science 293, 2266−2269.

Jonassen, C.M., Jonassen, T.O., Grinde, B., 1998. A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. J. Gen. Virol. 79 (Pt 4), 715−718.

Kaiser, J., 2008. DNA sequencing. A plan to capture human diversity in 1000 genomes. Science 319, 395.

Kersey, P.J., Lawson, D., Birney, E., Derwent, P.S., Haimel, M., Herrero, J., et al., 2010. Ensembl genomes: extending Ensembl across the taxonomic space. Nucleic Acids Res. 38, D563−D569.

Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., et al., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. U.S.A. 107, 12168−12173.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., et al., 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1953−1966.

Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., Butler, R., et al., 2009. VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res. 37, D583−587.

Levy, C.W., Roujeinikova, A., Sedelnikova, S., Baker, P.J., Stuitje, A.R., Slabas, A.R., et al., 1999. Molecular basis of triclosan activity. Nature 398, 383−384.

Liolios, K., Chen, I.M., Mavromatis, K., Tavernarakis, N., Hugenholtz, P., Markowitz, V.M., et al., 2010. The Genomes OnLine Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 38, D346−354.

Liu, Y.T., 2008. A technological update of molecular diagnostics for infectious diseases. Infect. Disord. Drug Targets 8, 183−188.

Lo, H., Tang, C.M., Exley, R.M., 2009. Mechanisms of avoidance of host immunity by *Neisseria meningitidis* and its effect on vaccine development. Lancet Infect. Dis. 9, 418−427.

Lu, Q., Zhang, X.Q., Pond, S.L., Reed, S., Schooley, R.T., Liu, Y.T., 2009. Detection in 2009 of the swine origin influenza A (H1N1) virus by a subtyping microarray. J. Clin. Microbiol. 47, 3060−3061.

Macdonald, G., 1957. The Epidemiology and Control of Malaria. Oxford University Press, Oxford.

Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., et al., 2005. Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309, 148−150.

Marrakchi, H., Choi, K.H., Rock, C.O., 2002. A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*. J. Biol. Chem. 277, 44809−44816.

Marshall, J.M., Taylor, C.E., 2009. Malaria control with transgenic mosquitoes. PLoS Med. 6, e20.

Martin, S.L., Borrow, R., van der Ley, P., Dawson, M., Fox, A.J., Cartwright, K.A., 2000. Effect of sequence variation in meningococcal PorA outer membrane protein on the effectiveness of a hexavalent PorA outer membrane vesicle vaccine. Vaccine 18, 2476−2481.

Megy, K., Hammond, M., Lawson, D., Bruggner, R.V., Birney, E., Collins, F.H., 2009. Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. Infect. Genet. Evol. 9, 308−313.

Morel, C.M., Toure, Y.T., Dobrokhotov, B., Oduola, A.M., 2002. The mosquito genome—a breakthrough for public health. Science 298, 79.

Munroe, D.J., Harris, T.J., 2010. Third-generation sequencing fireworks at Marco Island. Nat. Biotechnol. 28, 426−428.

Nelson, K.E., Weinstock, G.M., Highlander, S.K., Worley, K.C., Creasy, H.H., Wortman, J.R., et al., 2010. A catalog of reference genomes from the human microbiome. Science 328, 994−999.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., et al., 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316, 1718−1723.

NRC, 1988. Mapping and sequencing the human genome. <http://www.nap.edu/catalog.php?record_id=1097>.

OTA, 1988. Mapping our genes—genome projects: how big? how fast? <http://www.ornl.gov/sci/techresources/Human_Genome/publicat/OTAreport.pdf>.

Pagel Van Zee, J., Geraci, N.S., Guerrero, F.D., Wikel, S.K., Stuart, J.J., Nene, V.M., et al., 2007. Tick genomics: the *Ixodes* genome project and beyond. Int. J. Parasitol. 37, 1297−1305.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A., et al., 2009. The NIH Human Microbiome Project. Genome Res. 19, 2317–2323.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., et al., 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287, 1816–1820.

Povelones, M., Waterhouse, R.M., Kafatos, F.C., Christophides, G.K., 2009. Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. Science 324, 258–261.

Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., et al., 1978. The genome of simian virus 40. Science 200, 494–502.

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

Riehle, M.M., Markianos, K., Niare, O., Xu, J., Li, J., Toure, A.M., et al., 2006. Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. Science 312, 577–579.

Rinaudo, C.D., Telford, J.L., Rappuoli, R., Seib, K.L., 2009. Vaccinology in the genome era. J. Clin. Invest. 119, 2515–2525.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., et al., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300, 1394–1399.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., et al., 1977. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265, 687–695.

Savage, D.C., 1977. Microbial ecology of the gastrointestinal tract. Annu. Rev. Microbiol. 31, 107–133.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., et al., 2010. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 38, D5–D16.

Schriml, L.M., Arze, C., Nadendla, S., Ganapathy, A., Felix, V., Mahurkar, A., et al., 2010. GeMInA, Genomic Metadata for Infectious Agents, a geospatial surveillance pathogen database. Nucleic Acids Res. 38, D754–D764.

Sierra, G.V., Campa, H.C., Varcacel, N.M., Garcia, I.L., Izquierdo, P.L., Sotolongo, P.F., et al., 1991. Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba. NIPH Ann. 14, 195–207, discussion 208–210.

Smith, H.O., 2004. History of microbial genomics. In: Fraser, C.M., Read, T.D., Nelson, K.E. (Eds.), Microbial Genomes. Humana, Totowa, NJ, pp. 3–16.

Taubenberger, J.K., Reid, A.H., Lourens, R.M., Wang, R., Jin, G., Fanning, T.G., 2005. Characterization of the 1918 influenza virus polymerase genes. Nature 437, 889–893.

Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., et al., 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science 287, 1809–1815.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al., 2009. A core gut microbiome in obese and lean twins. Nature 457, 480–484.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., et al., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 1, E2.

Wang, J., Soisson, S.M., Young, K., Shoop, W., Kodali, S., Galgoci, A., et al., 2006. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. Nature 441, 358–361.

Watson, J.D., 1990. The Human Genome Project: past, present, and future. Science 248, 44–49.

Wright, H.T., Reynolds, K.A., 2007. Antibacterial targets in fatty acid biosynthesis. Curr. Opin. Microbiol. 10, 447−453.

Zhang, Y.M., Marrakchi, H., White, S.W., Rock, C.O., 2003. The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. J. Lipid Res. 44, 1−10.

Zhang, Y.M., White, S.W., Rock, C.O., 2006. Inhibiting bacterial fatty acid synthesis. J. Biol. Chem. 281, 17541−17544.