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.
1. Introduction

The history and development of infectious disease genomics are closely associated with the Human Genome Project (HGP)\(^1\). A series of important discussions about the HGP were made from 1985 to 1986\(^1,2\), 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\(^3\). 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\(^4\). In 1990, the Department of Energy (DOE) and the National Institutes of Health (NIH) jointly presented an initial 5-year plan for the HGP\(^5\). In October 1993, the Sanger Center/Institute (Hinxton, UK) was officially open to join the HGP. The cost of DNA sequencing was about $2 to $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\(^5\). 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. In early 2014, Illumina announced that the company would begin producing a new system to deliver full coverage human genomes for less than $1,000\(^6\).

The first complete genome to be sequenced was the phiX174 bacteriophage (5.4 kb) by Sanger’s group in 1977\(^7\). The complete genome sequence of SV40 polyomavirus (5.2 kb) was published in 1978\(^8,9\). The human Epstein–Barr virus (170 kb) genome was determined in 1984\(^10\). The first completed free-living organism genome was *Hae-mophilus influenza* (1.8 Mb), sequenced through a whole-genome shotgun approach in 1995\(^11\). The second sequenced bacterial genome, *Mycoplasma genitalium* (600 kb), was completed in less than 1 month in the same year using the same approach\(^12\). The DOE was the first to start a microbial genome program (MGP) as a companion to its HGP in 1994\(^13\). 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)\textsuperscript{14} was established in 2000 to accelerate the slow pace of fungal genome sequencing since the report of the genome of \textit{Saccharomyces cerevisiae} in 1996.\textsuperscript{15} One of the major interests was to sequence organisms that are important in human-health and commercial activities. With the explosion in the number of sequenced genomes, thanks to the development of next generation—sequencing methods, many genome-based studies have become popular. Compared to 6 years ago when only 1100 completed genome projects were documented, the GOLD (Genomes OnLine Database) contains information for 67,879 genome-sequencing projects, of which 7210 were completed, as of August 2015.\textsuperscript{16,17}

The genomes of human malaria parasite \textit{Plasmodium falciparum} and its major mosquito vector \textit{Anopheles gambiae} were published in 2002.\textsuperscript{18,19} Historically, the effort to sequence the malaria genome began in 1996 by taking advantage of a clone derived from laboratory-adapted strain.\textsuperscript{20} Notably, many parasites have complex life cycles that involve both vertebrate and invertebrate hosts and are difficult to maintain in the laboratory. Few other important human pathogenic parasites, such as trypanosomes, \textit{Leishmania}, and schistosomes,\textsuperscript{24,25} have been either completely or partially sequenced.\textsuperscript{26,27} In the meantime, the genome sequence of \textit{Aedes aegypti}, the primary vector for yellow fever and dengue fever, was published in 2007.\textsuperscript{28} The genome size (1376 Mb) of this mosquito vector is about 5 times larger than the previously sequenced genome of the malaria vector \textit{A. gambiae}. About 50% of the genome consists of transposable elements. In 2010, the genome sequence of the body louse (\textit{Pediculus humanus humanus}), an obligatory parasite of humans and the main vector of epidemic typhus (\textit{Rickettsia prowazekii}), relapsing fever (\textit{Borrelia recurrentis}), and trench fever (\textit{Bartonella quintana}), was reported.\textsuperscript{29} Its 108 Mb genome is the smallest among the known insect genomes. Subsequently, more vector genomes have been published.\textsuperscript{30–32} Genome-sequencing projects for other important human disease vectors are in progress.\textsuperscript{33,34} These include \textit{Culex pipiens} (mosquito vector of West Nile virus), and \textit{Ixodes scapularis} (tick vector of Lyme disease, \textit{Babesia} and \textit{Anaplasma}). The challenge to sequence the genome of an insect vector is much greater than a microbe. For example, the genome of ticks was 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.\textsuperscript{35} 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.\textsuperscript{36,37} 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.\textsuperscript{38} 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.\textsuperscript{39,40} As of May 2010, more than 5000 human and avian isolates had been completely sequenced, including the 1918 “Spanish” influenza virus.\textsuperscript{41} 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.\textsuperscript{42} It was suggested to have “the 2nd human genome project” to sequence the human microbiome.\textsuperscript{43} Highly variable intestinal microbial flora among normal individuals has been well documented.\textsuperscript{44–46} Therefore, the Human Microbiome Project (HMP) was initiated by the NIH in late 2007.\textsuperscript{47} The analysis and data of 242 healthy adults at 15 (for males) or 18 (for females) body sites over 22 months were published in 2012.\textsuperscript{48}

The completed or ongoing genome projects (Table 10.1) 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 are

| **Table 10.1 Completed or Ongoing Genome Projects** |
|---------------------------------|
| **General**                     |
| NCBI\textsuperscript{106} (http://www.ncbi.nlm.nih.gov/sitesgenome) |
| ENSEMBL\textsuperscript{107} (http://www.ensemblgenomes.org/) |
| JCVI\textsuperscript{108} (http://cmr.jcvi.org/) |
| GOLD\textsuperscript{16} (http://www.genomesonline.org) |
| Sanger Pathogen genomics (http://www.sanger.ac.uk/Projects/Pathogens/) |
| GeMInA (genomic metadata for infectious Agents)\textsuperscript{109,110} (http://gemina.igs.umaryland.edu) |
| **Bacteria**                    |
| Human Microbiome Project\textsuperscript{111} (http://www.hmpdacc.org/) |
| **Fungi**                      |
| Fungal Genome Initiative (FGI) (http://www.broadinstitute.org/science/projects/fungal-genome-initiative) |
| **Parasites**                  |
| Eukaryotic pathogens\textsuperscript{27} (http://EuPathDB.org) |
| Parasite genome projects (http://www.pasteur.fr/recherche/unites/tcruzi/minoprio/genomics/parasites.htm) |
| **Invertebrate vectors**       |
| VectorBase\textsuperscript{33,34} (http://www.vectorbase.org) |
| **Viruses**                    |
| Influenza virus\textsuperscript{112} (http://www.ncbi.nlm.nih.gov/genomes/FLU/) |
| HIV (http://www.hiv.lanl.gov/) |
| HCV (http://hcv.lanl.gov/) |
provided to illustrate how the information provided by various genome projects may help achieve the goal of promoting human health.

2. Vaccine Target

Meningococcal isolates produce one of 13 antigenically distinct capsular polysaccharides, but only five (A, B, C, W135, and Y) are commonly associated with disease. 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 ($\alpha_2-8\text{-N-acetylneuraminic acid}$) of MenB is identical to human polysialic acid, therefore is poorly immunogenic. Alternatively, vaccines consisting of outer-membrane vesicles (OMVs) have been successfully developed to control MenB outbreaks in areas where epidemics are dominated by one particular strain. 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.

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. 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. The vaccine formulation consists of an fHBP-GNA2091 fusion protein, a GNA2132-GNA1030 fusion protein, NadA, and OMVs 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. This vaccine (Bexsero) has been approved in the USA and in more than 30 other countries.

3. New Drug Discovery

Natural products, especially microbial secondary metabolites, are important source of bioactive compounds. Actinomycetes have been a main source of natural-product
discovery in bacteria. Consequently, the high rediscovery rate of known compounds and scaffolds were inevitable with activity-based screening. Genome mining of gene clusters that produce secondary metabolites have been a new approach to overcome this problem. For example, an antibiotic, clostrubin, was discovered through searching novel compounds from *Clostridium beijerinckii* due to the presence of several cryptic gene clusters for secondary metabolite biosynthesis.  

Genome mining starts with a genome-wide search for highly conserved members of the required biosynthesis gene cluster. Computational programs that support the prediction of operons help to assign boundaries of newly identified biosynthesis gene clusters. A large-scale, high-throughput genome mining for the genetic potential for producing phosphonic acids by screening more than 10,000 actinomycetes has been achieved in 2015. It was believed that phosphonates would have greater potential to become pharmaceuticals, with a past commercialization rate of 15% (3/20), such as fosfomycin, compared to the 0.1% average for natural products as a whole. In addition, bioinformatical discovery of phosphonate biosynthetic loci has been well established, as all but two previously characterized phosphonate biosynthetic pathways start with phosphoenolpyruvate (PEP) mutase that is encoded by *pepM*. Among 10,000 actinomycetes, only 278 strains were confirmed to have *pepM* by polymerase chain reaction (PCR) screening and genome sequencing. A diverse collection of phosphonate biosynthetic gene clusters were identified within these strains. Remarkably, 55 out of the 64 distinct clusters would direct the synthesis of unknown compounds. Characterization of strains within five of these groups resulted in discovery of argolaphos, and other interesting compounds, including valinophos, and phosphonocystoximate. Argolaphos showed broad-spectrum antibacterial activity against *Salmonella typhimurium*, *E. coli*, and *Staphylococcus aureus*.

4. **Drug Target**

Targeting an essential pathway is a necessary but not sufficient requirement for an effective antimicrobial agent. Identification of essential genes in a completely sequenced genome has been actively pursued with various approaches. The indispensable fatty acid synthase (FAS) pathway in bacteria has been regarded as a promising target for the development of antimicrobial agents. 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. In addition, through computational analyses, new members of the FAS II system have been discovered in different bacterial species. One of the protein components in this system, FabI, is the target
of an antituberculosis drug isonizid and a general antibacterial and antifungal agent, triclosan.72–74

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.75 Treatment with platensimycin eradicated S. 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 and 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.65,76 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.76 Comparative genomic approaches may be useful to identify and develop a strategy to target the salvage pathway for S. agalactiae. Alternatively, similar approaches as described earlier for MenB vaccine may also be applied for S. agalactiae (Group B Streptococcus).77

5. Therapeutic Response and Drug Resistance

Emergence of drug-resistant malaria to chloroquine in 1950s and sulfadoxine–pyrimethamine in 1960s occurred from western Cambodia to the Greater Mekong subregion (GMSR, including Cambodia, Lao, Myanmar, Thailand, and Vietnam) and to Africa. The finding of artemisinin-resistant malaria in Cambodia and GMSR raised a concern regarding the global spread of these parasites. While a number of studies, including population genetics and laboratory-based investigations were conducted, no reliable molecular marker was identified until the major breakthrough reported in early 2014.78 Clinical artemisinin resistance has been defined as a reduction of parasite-clearance rate, which is expressed as an increase of parasite-clearance half-life, or a persistence of microscopically detectable parasites 3 days after artemisinin-based combination therapy (ACT). Although artemisinin was thought to have broad-stage specificity against malaria throughout the life cycle, it was showed that artemisinin-resistant parasites only had decrease of artemisinin susceptibility at ring stages, which was demonstrated by the ring-stage survival assay (RSA0–3 h).79

An in vitro laboratory-based approach was conducted at a time when population-based genome-wide association studies (GWAS) did not clearly identify the genes responsible for artemisinin resistance.78 For 5 years, an artemisinin-resistant F32-ART5 parasite line was selected by culturing an artemisinin-sensitive F32-Tanzania clone under a dose-escalating, 125-cycle regimen of artemisinin. Eight mutations in seven genes were eventually selected from the result based on
whole-genome sequence analysis F32-ART5 and F32-TEM (its sibling clone cultured without artemisinin) at 460× and 500× average nucleotide coverage, respectively. To examine whether these in vitro selected mutations were associated with artemisinin resistance in Cambodia, sequence polymorphism in all seven genes were analyzed from 49 culture-adapted clinical isolates related to their RSA0–3 h. Only polymorphisms of a gene, K13-propeller, showed a significant association with RSA0–3 h survival rates. In total, four mutant alleles, each harboring a single nonsynonymous SNP (Y493H, R539T, I543T, and C580Y) within a kelch repeat of the C-terminal K13-propeller domain were identified. To confirm that K13-propeller polymorphism is a molecular marker of clinical artemisinin resistance, parasite-clearance half-lives in patients were correlated with their K13 alleles. Of the 150 patients, 72 carried parasites with a wild-type allele and the others carried parasites with only one of the three single nonsynonymous SNPs in the K13-propeller: C580Y (n = 51), R539T (n = 6), and Y493H (n = 21). The parasite-clearance half-life in patients with wild-type parasites is significantly shorter (median 3.30 h) than those with these three mutant alleles (median 6.28–7.19 h). Subsequently, clinical studies have validated the association between K13 propeller mutations and artemisinin resistance.80–82

6. Vector Control

Early mathematical model for malaria control suggested that the most vulnerable element in the malaria cycle was survivorship of adult female mosquitoes.83,84 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 has reduced the population at risk of malaria to about 50% by 1975 compared with 77% in 1900.83,85 Engineering genetically modified mosquitoes refractory to malaria infection appeared to be an alternative approach,86 given the environmental impact of DDT and the emergence of insecticide-resistant insects. The Vector Biology Network (VBN) was formed in 1989 and had proposed a 20-year plan with the 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 2010.87–89 While some proof-of-concept experiments have been achieved for the first two aims in 2002 when the A. gambiae genome was completely sequenced,90,91 the progress has been relatively slow.92

Genomic loci of the A. gambiae responsible for P. falciparum resistance have been identified through surveying a mosquito population in a West African malaria transmission zone.93 A candidate gene, Anopheles Plasmodium-responsive leucine-rich repeat 1 (APL1) was discovered. Subsequently, other resistant genes have also been identified.94,95 Studying the genetic basis of resistance to malaria parasites and immunity of the mosquito vector will be important to control malaria transmission.96
7. Clinical Application

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 reemerged pathogen is discovered. A few examples will be described.

A novel swine-origin influenza A virus (S-OIV) emerged in the spring of 2009 in Mexico and subsequently was discovered in specimens from two unrelated children in the San Diego area in mid-April 2009.97,98 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 WHO declared the highest Phase 6 worldwide pandemic alert on June 11, 2009. S-OIV has three genome segments (HA, NP, and NS) from the classic North American swine (H1N1) lineage, two segments (PB2 and PA) from the North American avian lineage, one segment (PB1) from the seasonal H3N2, and most notably, two segments (NA and M) from the Eurasian swine (H1N1) lineage.98 With the available influenza genome database, diagnostic assays to distinguish previous seasonal H1N1, H3N2, and S-OIV can be easily accomplished.99

A comprehensive pathogen genome database is not only useful for infectious disease diagnosis but also for novel pathogen discovery.100 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.101 De novo pathogen discovery may also be 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 conoronavirus.102 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 (severe acute respiratory syndrome) 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 bioinformatics analyses that this sequence is present in the 3′ UTR of all astroviruses, avian infectious bronchitis virus, and an equine rhinovirus.103 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)104 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.

In 2015, an interesting and unexpected discovery of the malignant transformation of *Hymenolepis nana*, a human tape worm, in a human host has been reported by conventional and next generation—sequencing approaches. Initially, examination of a 41-year-old HIV-infected man revealed extensive lymphadenopathy. *H. nana* eggs and *Blastocystis hominis* cysts were found in stool. The disease progressed to death despite antiparasitic and antiretroviral treatment. Histological examination of biopsied lymph nodes revealed proliferative cells with overt malignant features. They were monomorphic with morphologic features characteristic of stem cells (a high nucleus-to-cytoplasm ratio). However, the small cell size (<10) suggested infection with an unfamiliar, possibly unicellular, eukaryotic organism. Infection with a plasmodial slime mold rather than *H. nana* was considered because of the prominent syncytia formation and the primitive appearance of the atypical cells but lack of architecture identifiable as tapeworm tissue. PCR screening suggested that these cells were *H. nana*. Next generation—genome sequencing and comparative analysis revealed *H. nana* variants harboring mutations typically found in cancer.

As of 2016, next generation—sequencing technologies are gradually being applied for diagnosis and monitoring of infectious diseases, including genotypic resistance testing, direct detection of unknown disease-associated pathogens without culture, investigation of microbial population diversity in the host, and strain typing. However, promising, next generation—sequencing approaches for clinical diagnosis require further improvements for automation, standardization of technical and bioinformatic procedures, and other practical issues, such as costs and turnaround time.

### 8. 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 pose a challenge for scientists as well nonscientists to follow and understand.
References

1. Watson JD. The human genome project: past, present, and future. *Science* 1990;248:44–9.
2. Dulbecco R. A turning point in cancer research: sequencing the human genome. *Science* 1986;231:1055–6.
3. NRC. *Mapping and sequencing the human genome*. 1988. http://www.nap.edu/catalog.php?record_id=1097.
4. OTA. *Mapping our genes—genome projects: how big? How fast?*. 1988. http://www.ornl.gov/sci/techresources/Human_Genome/publicat/OTAreport.pdf.
5. DHHS and DOE. *Understanding our genetic inheritance, the U.S. Human genome project: the first five years: fiscal years 1991–1995*. 1990. http://www.ornl.gov/sci/techresources/Human_Genome/project/5yrplan/summary.shtml.
6. Hayden EC. The $1,000 genome. *Nature* 2014;507:294–5.
7. Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977;265:687–95.
8. Reddy VB, Thimmappaya B, Dhar R, et al. The genome of simian virus 40. *Science* 1978;200:494–502.
9. Fiers W, Contreras R, Haegemann G, et al. Complete nucleotide sequence of SV40 DNA. *Nature* 1978;273:113–20.
10. Baer R, Bankier AT, Biggin MD, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* 1984;310:207–11.
11. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 1995;269:496–512.
12. Smith HO. Hisotry of microbial genomics. In: Fraser CM, Read TD, Nelson KE, editors. *Microbial genomes*. Totowa, NJ: Humana; 2004. p. 3–16.
13. DOE. *Microbial genome program*. 2009. http://microbialgenomics.energy.gov/mgp.shtml.
14. Campbell IG, Russell SE, Choong DY, et al. Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res* 2004;64:7678–81.
15. Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. *Science* 1996;274(546):63–7.
16. Liolios K, Chen IM, Mavromatis K, et al. The Genomes on Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010;38:D346–54.
17. Reddy TB, Thomas AD, Stamatis D, et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res* 2015;43:D1099–106.
18. Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 2002;419:498–511.
19. Holt RA, Subramanian GM, Halpern A, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002;298:129–49.
20. Hoffman SL, Bancroft WH, Gottlieb M, et al. Funding for malaria genome sequencing. *Nature* 1997;387:647.
21. El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 2005;309:409–15.
22. Berriman M, Ghedin E, Hertz-Fowler C, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005;309:416–22.
23. Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, Leishmania major. Science 2005;309:436–42.
24. Berriman M, Haas BJ, LoVerde PT, et al. The genome of the blood fluke Schistosoma mansoni. Nature 2009;460:352–8.
25. Consortium. The Schistosoma japonicum genome reveals features of host-parasite interplay. Nature 2009;460:345–51.
26. Brindley PJ, Mitreva M, Ghedin E, Lustigman S. Helminth genomics: the implications for human health. PLoS Negl Trop Dis 2009;3:e538.
27. Aurrecoechea C, Brestelli J, Brunk BP, et al. EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res 2010;38:D415–9.
28. Nene V, Wortman JR, Lawson D, et al. Genome sequence of Aedes aegypti, a major arbovirus vector. Science 2007;316:1718–23.
29. Kirkness EF, Haas BJ, Sun W, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci USA 2010;107(27):12168–73.
30. Neafsey DE, Waterhouse RM, Abai MR, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science 2015;347:1258522.
31. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, et al. Genome of Rhodnius prolixus, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc Natl Acad Sci USA 2015;112:14936–41.
32. International Glossina Genome I. Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis. Science 2014;344:380–6.
33. Lawson D, Arensburger P, Atkinson P, et al. VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res 2009;37:D583–7.
34. Megy K, Hammond M, Lawson D, Bruggner RV, Birney E, Collins FH. Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. Infect Genet Evol 2009;9:308–13.
35. Pagel Van Zee J, Geraci NS, Guerrero FD, et al. Tick genomics: the Ixodes genome project and beyond. Int J Parasitol 2007;37:1297–305.
36. Feero WG, Guttmacher AE, Collins FS. The genome gets personal—almost. JAMA 2008;299:1351–2.
37. Alcais A, Abel L, Casanova JL. Human genetics of infectious diseases: between proof of principle and paradigm. J Clin Invest 2009;119:2506–14.
38. Kaiser J. DNA sequencing. A plan to capture human diversity in 1000 genomes. Science 2008;319:395.
39. Fauci AS. Race against time. Nature 2005;435:423–4.
40. Ghedin E, Sengamalay NA, Shumway M, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 2005;437:1162–6.
41. Taubenberger JK, Reed AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. Nature 2005;437:889–93.
42. Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 1977;31:107–33.
43. Relman DA, Falkow S. The meaning and impact of the human genome sequence for microbiology. Trends Microbiol 2001;9:206–8.
44. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science 2009;326:1694–7.
45. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. Science 2005;308:1635–8.
46. Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. *Nature* 2009;457:480–4.

47. Relman DA. Microbiology learning about who we are. *Nature* 2012;486:194–5.

48. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.

49. Lo H, Tang CM, Exley RM. Mechanisms of avoidance of host immunity by Neisseria meningitidis and its effect on vaccine development. *Lancet Infect Dis* 2009;9:418–27.

50. Finne J, Bitter-Suermann D, Goridis C, Finne U. 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* 1987;138:4402–7.

51. Bjune G, Hoiby EA, Gronnesby JK, et al. Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. *Lancet* 1991;338:1093–6.

52. Sierra GV, Campa HC, Varacel NM, et al. Vaccine against group B Neisseria meningitidis: protection trial and mass vaccination results in Cuba. *NIPH Ann* 1991;14:195–207. discussion 8–10.

53. Jackson C, Lennon DR, Sotutu VT, et al. Phase II meningococcal B vesicle vaccine trial in New Zealand infants. *Arch Dis Child* 2009;94:745–51.

54. Boslego J, Garcia J, Cruz C, et al. 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* 1995;13:821–9.

55. Pizza M, Scarlato V, Masignani V, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 2000;287:1816–20.

56. Martin SL, Borrow R, van der Ley P, Dawson M, Fox AJ, Cartwright KA. Effect of sequence variation in meningococcal PorA outer membrane protein on the effectiveness of a hexavalent PorA outer membrane vesicle vaccine. *Vaccine* 2000;18:2476–81.

57. Tettelin H, Saunders NJ, Heidelberg J, et al. Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. *Science* 2000;287:1809–15.

58. Giuliani MM, Adu-Bobie J, Comanducci M, et al. A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA* 2006;103:10834–9.

59. Rinaudo CD, Telford JL, Rappuoli R, Seib KL. Vaccinology in the genome era. *J Clin Invest* 2009;119:2515–25.

60. Bruno L, Cortese M, Rappuoli R, Merola M. Lessons from reverse vaccinology for viral vaccine design. *Curr Opin Virol* 2015;11:89–97.

61. Bidot S, Ishida K, Cyrullies M, Hertweck C. Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. *Angew Chem Int Ed Engl* 2014;53:7856–9.

62. Ju KS, Gao J, Doroghazi JR, et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc Natl Acad Sci USA* 2015;112:12175–80.

63. Berdy J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiott* 2012;65:385–95.

64. Metcalf WW, van der Donk WA. Biosynthesis of phosphonic and phosphinic acid natural products. *Annu Rev Biochem* 2009;78:65–94.

65. Brinster S, Lamberet G, Staels B, Trieu-Cuot P, Gruss A, Poyart C. Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. *Nature* 2009;458:83–6.

66. Ji Y, Zhang B, Van SF, et al. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001;293:2266–9.
67. Hutchison CA, Peterson SN, Gill SR, et al. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 1999;**286**:2165–9.
68. Wright HT, Reynolds KA. Antibacterial targets in fatty acid biosynthesis. *Curr Opin Microbiol* 2007;**10**:447–53.
69. Zhang YM, Marrakchi H, White SW, Rock CO. The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. *J Lipid Res* 2003;**44**:1–10.
70. Heath RJ, Rock CO. A triclosan-resistant bacterial enzyme. *Nature* 2000;**406**:145–6.
71. Marrakchi H, Choi KH, Rock CO. A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*. *J Biol Chem* 2002;**277**:44809–16.
72. Levy CW, Roujeinikova A, Sedelnikova S, et al. Molecular basis of triclosan activity. *Nature* 1999;**398**:383–4.
73. Zhang YM, White SW, Rock CO. Inhibiting bacterial fatty acid synthesis. *J Biol Chem* 2006;**281**:17541–4.
74. Banerjee A, Dubnau E, Quemard A, et al. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* 1994;**263**:227–30.
75. Wang J, Soisson SM, Young K, et al. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* 2006;**441**:358–61.
76. Balemans W, Lounis N, Gilissen R, et al. Essentiality of FASII pathway for *Staphylococcus aureus*. *Nature* 2010;**463**:E3. discussion E4.
77. Maione D, Margarit I, Rinaudo CD, et al. Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 2005;**309**:148–50.
78. Ariey F, Witkowski B, Amaratunga C, et al. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* 2014;**505**:50–5.
79. Witkowski B, Khim N, Chim P, et al. Reduced artemisinin susceptibility of *Plasmodium falciparum* ring stages in western Cambodia. *Antimicrob Agents Chemother* 2013;**57**:914–23.
80. Ashley EA, Dhorda M, Fairhurst RM, et al. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 2014;**371**:411–23.
81. Miotto O, Amato R, Ashley EA, et al. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 2015;**47**:226–34.
82. Tun KM, Imwong M, Lwin KM, et al. Spread of artemisinin-resistant *Plasmodium falciparum* in Myanmar: a cross-sectional survey of the K13 molecular marker. *Lancet Infect Dis* 2015;**15**:415–21.
83. Enayati A, Hemingway J. Malaria management: past, present, and future. *Annu Rev Entomol* 2010;**55**:569–91.
84. Macdonald G. *The epidemiology and control of malaria*. Oxford: Oxford Univ. Press; 1957.
85. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis* 2004;**4**:327–36.
86. Curtis CF. Possible use of translocations to fix desirable genes in insect pest populations. *Nature* 1968;**218**:368–9.
87. Beaty BJ, Prager DJ, James AA, et al. From Tucson to genomics and transgenics: the vector biology network and the emergence of modern vector biology. *PLoS Negl Trop Dis* 2009;**3**:e343.
88. Morel CM, Toure YT, Dobrokhotov B, Oduola AM. The mosquito genome—a breakthrough for public health. *Science* 2002;**298**:79.
89. Alphey L, Beard CB, Billingsley P, et al. Malaria control with genetically manipulated insect vectors. *Science* 2002;**298**:119–21.
90. Catteruccia F, Nolan T, Loukeris TG, et al. Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* 2000;**405**:959–62.
91. Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* 2002;**417**:452–5.
92. Marshall JM, Taylor CE. Malaria control with transgenic mosquitoes. *PLoS Med* 2009;**6**:e20.
93. Riehle MM, Markianos K, Niare O, et al. Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. *Science* 2006;**312**:577–9.
94. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK. Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. *Science* 2009;**324**:258–61.
95. Blandin SA, Wang-Sattler R, Lamacchia M, et al. Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. *Science* 2009;**326**:147–50.
96. Severo MS, Levashina EA. Mosquito defenses against *Plasmodium* parasites. *Curr Opin Insect Sci* 2014;3:30–6.
97. CDC. Swine influenza A (H1N1) infection in two children—Southern California, March–April 2009. *MMWR Morb Mortal Wkly Rep* 2009;**58**:400–2.
98. Dawood FS, Jain S, Finelli L, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med* 2009;**360**:2605–15.
99. Lu Q, Zhang XQ, Pond SL, Reed S, Schooley RT, Liu YT. Detection in 2009 of the swine origin influenza A (H1N1) virus by a subtyping microarray. *J Clin Microbiol* 2009;**47**:3060–1.
100. Liu YT. A technological update of molecular diagnostics for infectious diseases. *Infect Disord Drug Targets* 2008;**8**:183–8.
101. Munroe DJ, Harris TJ. Third-generation sequencing fireworks at Marco Island. *Nat Biotechnol* 2010;**28**:426–8.
102. Wang D, Urisman A, Liu YT, et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 2003;**1**:E2.
103. Jonassen CM, Jonassen TO, Grinde B. A common RNA motif in the 3’ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol* 1998;**79**(Pt 4):715–8.
104. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008. http://dx.doi.org/10.1126/science.1152586.
104a. Muehlenbachs A, Bhatnagar J, Agudelo CA, et al. Malignant transformation of hyme-nolepis nana in a human host. *N Engl J Med* 2015;**373**(19):1845–52. http://dx.doi.org/10.1056/NEJMoa1505892.
105. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-generation sequencing for infectious disease diagnosis and management: a report of the association for molecular pathology. *J Mol Diagn* 2015;**17**:623–34.
106. Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology information. *Nucleic Acids Res* 2010;**38**:D5–16.
107. Kersey PJ, Lawson D, Birney E, et al. Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Res* 2010;**38**:D563–9.
108. Davidsen T, Beck E, Ganapathy A, et al. The comprehensive microbial resource. *Nucleic Acids Res* 2010;**38**:D340–5.
109. Ecker DJ, Sampath R, Willett P, et al. The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. *BMC Microbiol* 2005;**5**:19.
110. Schriml LM, Arze C, Nadendla S, et al. GeMInA, genomic metadata for infectious agents, a geospatial surveillance pathogen database. *Nucleic Acids Res* 2010;38:D754–64.

111. Nelson KE, Weinstock GM, Highlander SK, et al. A catalog of reference genomes from the human microbiome. *Science* 2010;328:994–9.

112. Bao Y, Bolotov P, Dernovoy D, et al. The influenza virus resource at the National Center for Biotechnology information. *J Virol* 2008;82:596–601.