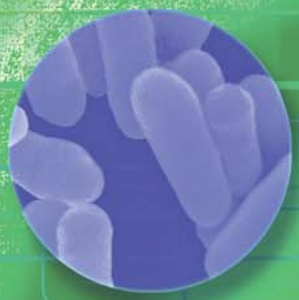
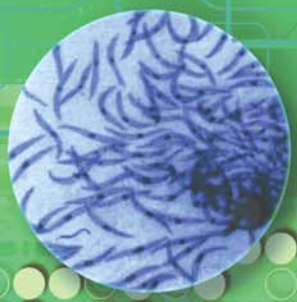
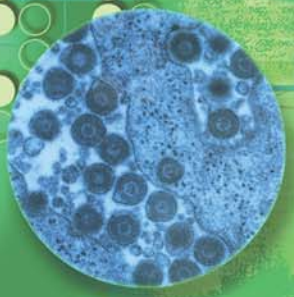
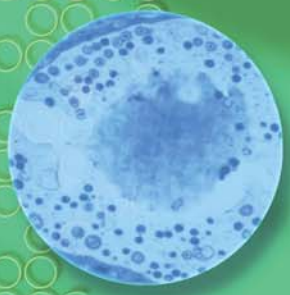
Deciphering Pathogens

Blueprints for New Medical Tools









U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

National Institute of Allergy and Infectious Diseases

Tuberculosis Bacteria

Malaria Sporozoites

Herpes Virus

Chlamydia pneumoniae Bacteria

Deciphering Pathogens

Blueprints for New Medical Tools

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
NATIONAL INSTITUTES OF HEALTH

National Institute of Allergy and Infectious Diseases

Division of Microbiology and Infectious Diseases

NIH Publication No. 02-4987
September 2002
www.niaid.nih.gov

Credits

cover insets TOP TO BOTTOM:
Tuberculosis bacteria
© Dennis Kunkel Microscopy, Inc.

Malaria sporozoite
National Institute of Allergy and Infectious Diseases

Herpes virus
© Dennis Kunkel Microscopy, Inc.

Chlamydia pneumoniae
National Institute of Allergy and Infectious Diseases

2 TOP: **Anthrax spores**
National Institute of Allergy and Infectious Diseases

BOTTOM: **Malaria-infected cells**
Ga. Biomedical Partnership www.gabio.org/overview.asp

3 Mycobacterium tuberculosis
National Institute of Allergy and Infectious Diseases

BACKGROUND: **Tuberculosis bacteria**
© Dennis Kunkel Microscopy, Inc.

4 TOP LEFT: **Tuberculosis bacteria**
© Dennis Kunkel Microscopy, Inc.

TOP CENTER: **Malaria sporozoite**
National Institute of Allergy and Infectious Diseases

TOP RIGHT: **Human immunodeficiency virus**
Centers for Disease Control and Prevention

MIDDLE LEFT: **Pneumococcus bacteria**
Janice Carr, Centers for Disease Control and Prevention

MIDDLE RIGHT: **Influenza virus**
National Institute of Allergy and Infectious Diseases, NIH, Gupta Murti, University of Tennessee

BOTTOM: **Ebola virus**
National Institute of Allergy and Infectious Diseases

Prion
Southwest Immunology, Inc.

6 DNA Illustration
Marie Dauenheimer, MA, CMI Courtesy of the National Institute of General Medical Sciences, NIH

7 INSET AND BACKGROUND:
Malaria sporozoite
National Institute of Allergy and Infectious Diseases

8 Micrograph
The Institute for Genomic Research www.tigr.org

11 LEFT: **Gonorrhea bacteria**
National Institute of Allergy and Infectious Diseases

RIGHT: **Human papillomavirus**
National Institute of Allergy and Infectious Diseases

BACKGROUND: **Treponema pallidum, bacteria that cause syphilis**
National Institute of Allergy and Infectious Diseases

12 TOP: **Escherichia coli**
National Institute of Allergy and Infectious Diseases

BOTTOM: **Vibrio cholerae**
National Institute of Allergy and Infectious Diseases

15 INSET AND BACKGROUND:
Chlamydia pneumoniae
National Institute of Allergy and Infectious Diseases

17 INSET: **Microarray**
Genomes to Life U.S. Department of Energy http://doegenomestolife.org

18 Anthrax spores
National Institute of Allergy and Infectious Diseases

19 TOP: **Ebola virus**
National Institute of Allergy and Infectious Diseases

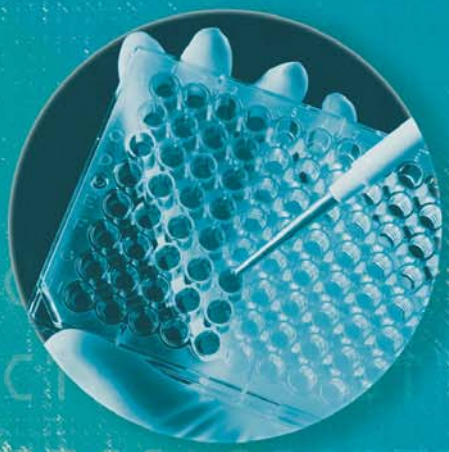
BOTTOM: **Ebola virus**
School of Veterinary Medicine, University of California, Davis.

20 LEFT: **Candida albicans**
© Dennis Kunkel Microscopy, Inc.

CENTER: **Anthrax spores**
National Institute of Allergy and Infectious Diseases

RIGHT: **Prion**
Researched Diagnostics Inc

Contents

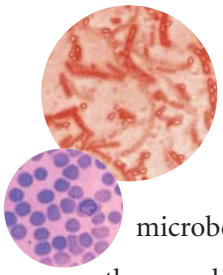


Understanding the Inner Workings of Disease-Causing Microbes

The human species has co-existed with microbes—tiny organisms like bacteria, viruses, fungi, and parasites—since we first appeared on Earth. During the hundreds of millennia since then, it has been a complex and sometimes uneasy relationship. Indeed, humans cannot live without certain microbes that help maintain proper physiological conditions inside our bodies, nourish our soil, or keep our environment in balance. At the same time, these tiny creatures have also swiftly devastated human populations around the world.

Scientists who study microbes that cause human illness, commonly called pathogens, have continually developed newer and better strategies to understand how the germs cause disease and how to stop them. That knowledge has led to the control of many diseases, like measles, diphtheria, tetanus, pertussis, mumps, and polio, and the complete eradication of natural disease from one-time killers like smallpox.

At the end of the 20th century, the new research approach known as "genomics" emerged as a powerful investigative tool for understanding biology. Now, investigators in every area of the life sciences use the vast and orderly collections of an organism's genetic information, known as its "genome," to understand each and every process an organism carries out, all at once, as a complete system.



The Burden of Infectious Diseases

Probably nowhere in science is genomics making more inroads into understanding human diseases than in the field of microbiology—the study of microbes and their interactions with human beings and the environment. With the complete genomes of dozens of pathogenic microbes in hand—and many more on the way—scientists have begun a powerful new assault on some of humankind's oldest enemies. Illnesses caused by pathogenic microbes are today the second leading cause of death worldwide and third leading cause of death in the United States. Throughout the world, infectious diseases are also responsible for countless hospitalizations, sick days, missed work, and other lost opportunities. Those losses cost hundreds of billions of dollars each year.

The Leading Infectious Causes of Death Worldwide, 2000

Lower respiratory infections	3.9 million
HIV/AIDS	2.9 million
Diarrheal diseases	2.1 million
Tuberculosis	1.7 million
Malaria	1.1 million
Measles	777,000

Source: World Health Report 2001, World Health Organization, p. 144.

1546

Girolamo Fracastoro
Suggested that invisible organisms cause disease

1676

Antony van Leeuwenhoek
Observed living microbes, dubbed "animalcules," through a microscope

A CLOSER LOOK: Tuberculosis

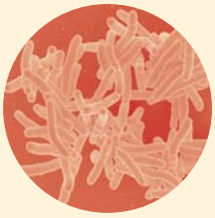
Genomics Yield Clues for Vaccines, New Treatments for Tuberculosis

Tuberculosis (TB) is a contagious disease that has exacted a devastating toll throughout human history. TB kills two million people each year, and experts expect 500 million new infections to occur globally over the next decade. To make matters worse, worrisome new multidrug-resistant strains of the TB bacterium have emerged, making treatment more difficult.

NIAID and its collaborators sponsored the genomic analysis of the TB bacterium, *Mycobacterium tuberculosis*, as well as that of the closely related *Mycobacterium leprae*, which causes leprosy. Although leprosy is a much different disease, the two pathogens nonetheless have a great deal in common. For example, both are able to sidestep critical defensive responses of an infected person's immune system. Detailed comparisons of their genomes therefore will help researchers learn how they and other microbes can circumvent immune responses.

Researchers studying the *M. tuberculosis* genome are developing other noteworthy insights into the germ's biochemical makeup and disease-causing properties. For instance, about 10 percent of that pathogen's genes are devoted to making lipids—fatty substances used in its outer coat that enable it to withstand human immune responses and to ward off the actions of several antibiotics. In addition, both *M. tuberculosis* and *M. leprae* are missing genes that enable other microorganisms to grow more efficiently, which helps explain a frustrating feature of these microorganisms—they are extraordinarily difficult to grow in laboratories. That characteristic has repeatedly thwarted scientists and clinical microbiologists in their efforts to examine and learn more about these pathogens.

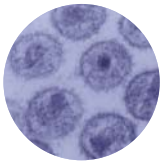
NIAID has also teamed with the National Institute of General Medical Sciences (NIGMS) to co-fund the *Mycobacterium tuberculosis* Structural Genomics Consortium. The consortium is a collaboration of scientists from 12 different countries who are analyzing the molecular structures of more than 400 proteins identified from the microbe's genome. Obtaining those structures should help researchers develop new or improved drugs and vaccines against tuberculosis.



Tuberculosis Bacteria

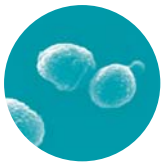
Tuberculosis Bacteria

Malaria Sporozoites

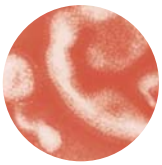


Human Immuno-deficiency Virus

Three infectious diseases alone—TB, malaria, and HIV/AIDS—illustrate the tremendous human toll taken by microbes each year. Though an ancient disease, TB remains a leading case of death worldwide, claiming 2 million lives every year. One-third of the world's population is infected with the TB bacterium, *Mycobacterium tuberculosis*, and someone is newly infected every second. Malaria, which is caused by a single-celled parasite spread to people by mosquitoes, kills up to 2.7 million people annually, most of them African children. Between 400 million and 900 million acute malaria cases occur each year in African children alone. HIV/AIDS first emerged in the early 1980s and has since killed more than 20 million people around the world. As of the end of 2001, 40 million people were living with HIV/AIDS.

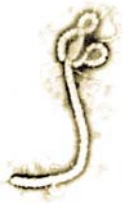


Pneumococcus Bacteria

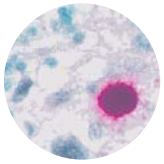


Influenza Virus

Those three diseases, however, are not the only threats to human health. Infectious diarrheal diseases kill almost as many people each year as HIV/AIDS, and respiratory infections kill far more. In the United States, pneumococcal pneumonia and influenza account for thousands of deaths during an ordinary year. That death toll surged to more than 20 million worldwide during the influenza pandemic that occurred during the 1918 flu season. While many important, infectious killers are held in check by vaccines and antibiotics, others cannot be stopped because there is currently no way to prevent or treat them. Other infections, such as West Nile fever, dengue, mad cow disease, and those caused by Ebola and hanta viruses, can emerge suddenly out of nowhere.



Ebola Virus



Prion

1861

Louis Pasteur
Ended the controversy over the origin of microorganisms; proved that microbes arise from pre-existing microbes, not spontaneously

Understanding the Genomes of Microbes Will Provide New Insights into Health

By rigorously evaluating and funding laboratory, clinical, and field investigations that analyze the genomes of pathogenic microbes, the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, a component of the U.S. Department of Health and Human Services, supports a diverse research program to define exactly how a microbe's genetic instructions drive it to make people sick. NIAID genomics research focuses on these areas

- Obtaining the complete genome sequence of organisms that are pathogenic to humans
- Analyzing pathogen genomes to understand how the microbe causes disease
- Using knowledge of microbe gene function to identify targets for new drugs or vaccines that can interrupt the infectious or disease-causing process
- Identifying mechanisms of drug resistance in microbes
- Studying the human genome to identify genetic variations that may protect or predispose people to infection or make them resistant to the benefits of vaccines or antibiotics
- Making the data freely available for use by all researchers

1876

Robert Koch
Showed that anthrax was caused by the bacterium *Bacillus anthracis*; first to link a specific microbe with a disease

1881

Robert Koch
Developed methods for studying bacteria in pure culture

Reading a Microbe's Genetic Language

In an earlier time, the vast range in size and complexity of microbes and their genetic instructions presented huge technical difficulties for scientists trying to parse out the molecular instructions that give germs the upper hand over human beings. Although researchers have long known a pathogen's sinister behaviors are encoded in the genetic language of DNA (deoxyribonucleic acid), or its related molecule, RNA (ribonucleic acid), deciphering the code in the laboratory once took months to years for even the smallest organisms. But now, thanks to the push to advance DNA sequencing of much larger genomes, newer automated technology allows researchers to read relatively small microbe genomes in days or even hours.

Illustration of DNA bases organized into genes.

The instructions in a genome are contained in sets of DNA bases represented by the letters A, C, T, and G, and organized into genes. (Some virus genomes consist of RNA, a type of chemical photocopy of DNA.) The sequence of bases in DNA or RNA spells out the recipe for a pathogen's full set of biochemical ingredients—typically, anywhere from several hundred to several thousand kinds of proteins. Those proteins in turn are used to build cell structures and to carry out most of its activities. Using genome sequencing, scientists studying pathogenic microbes can for the first time look at complete collections of genes and the proteins they encode. Genomes therefore enable researchers to identify previously hidden molecules and biological processes that may give way to new drugs and vaccines aimed at interrupting them.

1884

Robert Koch
Published Koch's postulates for determining the cause of infectious diseases

A CLOSER LOOK: Malaria

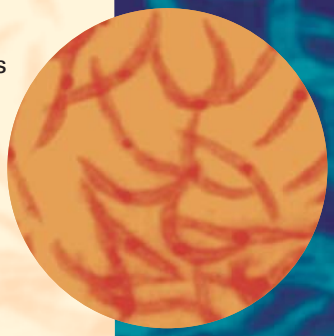
Using Genomics to Solve the Complexities of a Notorious Killer

As many as 300 to 500 million new cases of malaria develop each year, and the disease is responsible for up to 2.7 million deaths yearly, many among young children. Although potent drugs and aggressive mosquito control measures once appeared to have malaria on the retreat in some areas of the world, widening drug resistance and concerns about unsafe uses of insecticides are among the reasons why malaria is now resurgent.

The parasites that cause malaria in humans, *Plasmodium falciparum* and three closely related species, are considerably more complex than bacteria. Not only does the parasite spend part of its life in a mosquito, but once it infects humans it undergoes distinct changes in form and size as it invades different cell types, including red blood cells and liver cells. As might be expected, the complex life cycle of *Plasmodium* species reflects its similarly complex underlying genetics. It's genome is nearly 10-times the size of a typical bacterium and is divided among several chromosomes. In the case of *P. falciparum*, 14 different chromosomes contain about 30 million base pairs of DNA.

During the mid-1990s, NIAID joined forces with other agencies and private foundations in the United States and United Kingdom to establish an international consortium for determining the genomic sequence of *P. falciparum.* Genomics experts working as part of that consortium have been steadily accelerating their efforts to determine the sequence of this parasite. By 1998, the DNA sequence for one of the 14 chromosomes had been completed, quickly leading to the identification of genes encoding a family of proteins that help this parasite elude the human immune system. The completed *P. falciparum* genome sequence was published in 2002, along with the genome of its mosquito host. Those sequences can now be combined with the recently completed human genome sequence to help paint a complete picture of the genetic mechanisms that control malaria infection, transmission, and immunity.

The accumulating genomic data have helped researchers determine how the parasite becomes resistant to malaria drugs and to identify promising targets for new antimalaria agents. For instance, a group of researchers in Germany scanned the public genome database to learn that the first fully sequenced *P. falciparum* chromosome contains genes encoding a relatively unusual means for making steroids. Further analysis suggested a point of vulnerability where drugs might inhibit some of this vital biochemical process within the parasite without affecting similar processes within infected people. Preliminary findings confirm that a drug called fosmidomycin and other closely related compounds are effective when used to treat malaria-like infections in mice—a sign that these drugs should be further evaluated for their safety and effectiveness in humans.

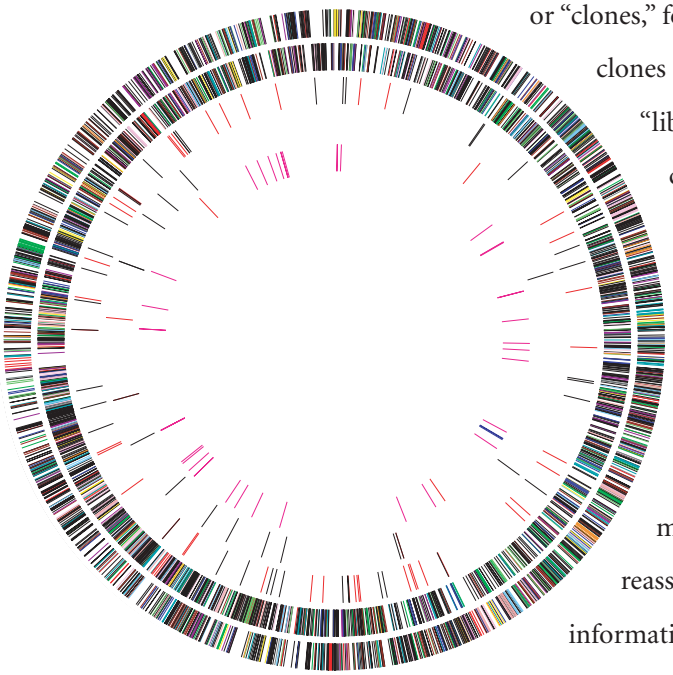


Malaria Sporozoites

Since the first genomic sequence of a bacterial pathogen was completed in 1995, pathogen genome sequencing efforts have continued to accelerate at an astonishing rate. The genomes of several dozen microbes known to cause disease in humans have now been sequenced, with more soon to be reported. An up-to-date list of completed and ongoing genome sequencing projects can be found on the NIAID Web site at www.niaid.nih.gov.

How a Genome Is Sequenced

Microbial genome projects start with laboratory procedures aimed at dividing large DNA molecules into smaller fragments for easier handling. Special enzymes are used to cut a microbe's single DNA-containing chromosome (or, in some cases, its several separate DNA molecules) randomly into hundreds of pieces, which are collected into separate entities, or "clones," for follow-up analytical procedures. The entire collection of clones from a particular microorganism often is referred to as a "library," symbolizing its role as a comprehensive and orderly collection of the organism's genetic information.



Diagrammatic representation of the genome for the infectious bacterium *Streptococcus pneumoniae*. The bacterium has a circular chromosome, and the dashes shown throughout the circle represent the locations of individual genes. The color of each dash corresponds to the proposed function of each gene.

Typically, specialized instruments determine the base-pair sequence of DNA in each clone, and researchers repeat a sequencing analysis several times to ensure the information is accurate. That raw information is fed into computer programs that collect the sequences of the separate fragments, determine overlaps between fragments, and eventually reassemble them. The completed process catalogs each bit of information in the exact order it was in the overall genome.

1889

Martinus Beijerinck
Described general concept of a virus and linked a virus with a disease of plants

Even with a pathogen's full base-pair sequence in hand, the genome still contains many mysteries that remain to be ferreted out. Important follow-up efforts known as "annotation" involve a series of steps, many of them involving sophisticated computer analysis, to determine the identity and function of a microbe's full set of genes. So far, researchers studying microbial genomic sequences can identify and categorize perhaps one-third to one-half of the genes in a newly sequenced microbe. But many of the other gene sequences are yet to be identified.

Genomic Studies Explain How Microbes Cause Disease

Many of the genes within a microbial cell perform what scientists call "housekeeping" functions by providing energy, structural components, and other mundane but essential features to meet life's basic demands at the cellular level. However, some of those genes are peculiar to specific microbial pathogens, helping to set them apart from benign microorganisms and also from other pathogens. For instance, some pathogens infect the respiratory tract and may cause pneumonia, others infect the skin or soft tissues, and others infect the genital tract and are sexually transmitted. Still others may be foodborne and cause gastrointestinal disorders.

Genomic approaches are used to identify key similarities and differences among different microbes. Comparative genomics is a field of study that examines variation between genomes. By identifying genetic differences among different microbe species or strains, researchers can identify the genes that enable the microbes to behave as they do.

1928
Frederick Griffith
Discovered that bacteria can exchange genetic material

1944
Oswald Avery, Colin Macleod, Maclyn McCarty
Expanded on Griffith's work to prove that DNA is the genetic material

Comparative genomics can also identify common genetic features that can be targeted for new drugs, vaccines, or diagnostic tests.

For example, researchers have used genome sequencing to glean insights about two similar pathogens— *Vibrio cholerae* and *Escherichia coli.* Both bacteria sometimes cause severe gastrointestinal diseases but at other times are benign.

Lessons from the Human Genome Project

Applying the full force of genomic analysis to disease-causing microbes is a direct offshoot from the Human Genome Project, an ambitious international research effort to characterize the 3 billion base-pair genome of the human being. The completed draft sequence of the human genome was published in February 2001 and marked a watershed moment in biomedical research. The Project also encompasses DNA sequencing of other organisms important to human health research, such as common laboratory animals and several microbes used for studying the basic processes of cell biology. In addition, information obtained about human genes can be combined with microbial genetic information to paint a comprehensive picture of the myriad interactions between microbe and host. The Human Genome Project and microbial genomics research are also linked in other ways. Both rely on collaborations and partnerships with other government agencies, as well as with commercial and non-profit organizations in the United States and abroad. Microbial genomics also uses much of the technology developed for human genome sequencing.

1953

James Watson, Francis Crick, Maurice Wilkins, Rosalind Franklin
Solved the molecular structure of DNA

A CLOSER LOOK: Sexually Transmitted Diseases

Seeking Common Threads Among Pathogens Causing STDs

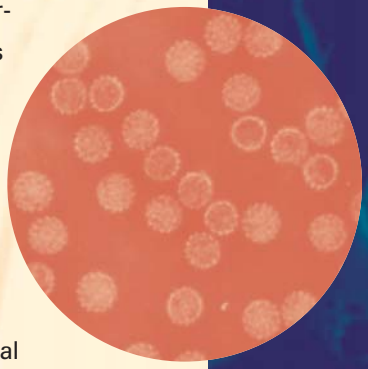
The microbes that cause sexually transmitted diseases (STDs) are a disparate lot—viruses, bacteria, and several more complex organisms, including yeasts and protozoa. They also cause a disparate set of diseases, some short-lived and little more than annoying, others invisible but potentially serious, and still others unambiguously harmful and life threatening. According to the Institute of Medicine, some 12 million new cases of STDs occur each year in the United States, about one-fourth of them among teenagers. The annual costs associated with these diseases are estimated at $10 billion, or $17 billion if sexually transmitted HIV is included in the tabulation.

Gonorrhea Bacteria

Despite the variety of microbes responsible for STDs, the germs that cause them share many similarities because of the specialized niche the pathogens occupy in their human hosts and because of the way in which most of them are transmitted.

To facilitate studies of STDs, NIAID began a collaborative effort with scientists at the Los Alamos National Laboratory in New Mexico to compile a special series of STD genomic databases. The initial focus of these efforts was the human papillomavirus (HPV), a virus that causes genital warts over the short term, and cervical and other cancers over the longer term. Because only certain genetic strains of HPV appear to cause disease, scientists developed an extensive genomic database to determine the genetic basis for the different strain properties.

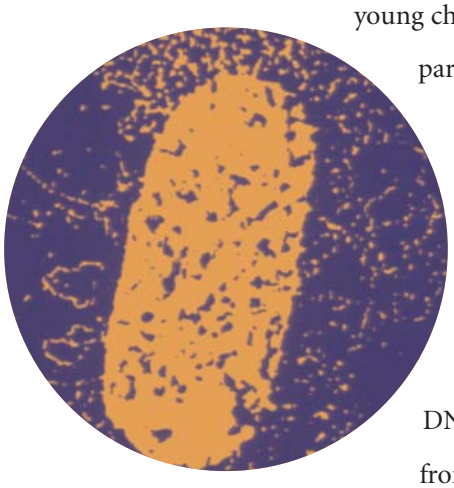
The modest effort to catalog and track HPV grew into a broader effort to catalog and analyze a series of STD pathogens using genomic and other database information. Eventually, researchers gathered the genomic sequences of human herpes virus, some of which cause genital lesions, and at least a half-dozen bacterial pathogens responsible for a variety of well-known STDs, including syphilis, gonorrhea, and chlamydial infections that can lead to pelvic inflammatory disease and sterility.



Human Papillomavirus

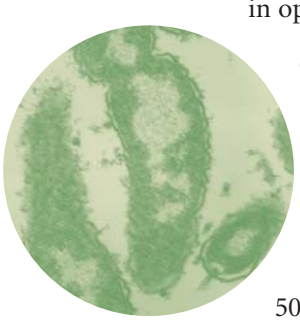
Researchers have begun to comb the databases to look for common genes and gene products the pathogens require for infection and survival in people. To date they have discovered at least four different STD bacteria deploy essentially the same biochemical ingredients in the early stages of establishing an infection. This and other information coming from continuing genomic analysis should lead to new targets for drugs or vaccines.

V. cholerae is responsible for causing outbreaks of cholera, a severe diarrheal disease that kills thousands of people each year. During pandemics, which typically occur several times per century, global tolls rise sharply, causing high death rates particularly among vulnerable young children living in developing countries where the bacteria are particularly common in water sources.



***Escherichia coli* Bacterium**

In 2001, a team of NIAID-supported cholera and genome-sequencing experts determined the full DNA sequence of the bacterium. Initial genomic analysis indicates that each comma-shaped *V. cholerae* bacterium contains two unevenly sized chromosomes that, taken together, comprise about 4 million DNA base pairs and encode nearly 4,000 genes. Information obtained from the bacterium's genomic sequence already is providing researchers with a better understanding of how this pathogen persists in open marine environments, often in a quiescent state while not causing harm to humans or wildlife between periods when it causes major cholera outbreaks.



***Vibrio cholerae* Bacterium**

Researchers have learned that most of *V. cholerae's* genes are much like those found in another more common bacteria, *E. coli,* which ordinarily is not nearly so harmful to humans. Despite the extensive genetic similarity between the two microbes, nearly 500 of the *V. cholerae* genes have no known counterpart in *E. coli.* Scientists are further analyzing the 500 "anonymous" *V. cholerae* genes to learn how they may play important roles in making it such a devastating pathogen.

Even different strains of the same microbe can carry different abilities to cause illness. *E. coli* O157:H7, for example, is a virulent strain that can cause bloody diarrhea, kidney troubles, and even death. Recognized since the early 1980s, this form of *E. coli* is responsible for nearly 75,000 cases of foodborne illness each year and as many as 60 deaths. Genomic sequence comparisons with benign strains of *E. coli* indicate that the two types share some 4,000 genes but the virulent *E. coli* O157:H7 strain carries 1,387 genes that the benign strain lacks.

1960s

Marshall Niremberg, Heinrich Mathaei, Severo Ochoa, and others
Cracked the genetic code, enabling researchers to "read" the DNA sequence of a gene and determine the protein it encodes

This initial analysis immediately provides insights into genetic features that likely contribute to the virulence of O157:H7. Some of its additional genes apparently encode enzymes that help it survive the acidic conditions of the stomach while other genes encode proteins that enable this bacterium to adhere tenaciously to the intestinal walls of its host. The analysis also suggests many of these virulence traits were assembled into the O157:H7 genome relatively recently, indicating the strain's extraordinary virulence may have arisen through natural exchanges of genes among microorganisms.

Such genomics-based analyses are providing researchers with valuable clues into identifying potentially important targets for drug and vaccine research. With knowledge of different genes' functions, researchers can determine how those genes contribute to a microbe's life processes and learn how to shut those processes down.

Thus, genomics experts anticipate working closely with their colleagues engaged in applied research to explore new leads and accelerate efforts to develop safe and effective new products to treat or prevent deadly outbreaks. For example, some of the sequence-specific signatures identified through the genomic analysis of *E. coli* O157:H7 may prove useful for identifying and monitoring this pathogenic microorganism in various settings, such as on farms where it may associate with cattle without causing outward signs of disease, to prevent its entry into the food supply.

Genomics Meets Bioinformatics

As genome sequences continue to become available, scientists must develop ways to efficiently deal with the accumulating data. Comprehensive data sets are currently released by investigators and widely shared among scientists, who now can access these vast databases by way of the Internet. The free availability of data about microbial genomes encourages important analytical comparisons with those of other pathogens, as well as with humans.

1970

Werner Arber, Hamilton Smith, Daniel Nathans
Discovered restriction endonucleases, key enzymes that cut DNA at specific sites and are critical tools for DNA analysis

Because the genomes of each one of the microbes may contain upwards of several million base pairs of DNA and thousands of genes, analyzing information on this immense scale looms as a daunting task. Genomics provides far more data than can be interpreted using uncomplicated calculations jotted on the back of an envelope. Rather, such analysis requires the help of sophisticated computers and the use of innovative software.

To undertake such analysis, biologists now depend on the cooperation of a new class of scientists who are knowledgeable not only about the biological properties of the organisms they study but also in newer computer and "bioinformatics" skills essential for making the best use of the huge volume of data.

Some of these approaches will be aimed at developing new diagnostic procedures for rapidly identifying pathogens, including those carrying unique medically important traits such as antibiotic resistance. Someday, such information may enable physicians to more quickly diagnose an infection, identify the microbial culprit, and more closely tailor specific treatments to the disease. In other cases, rapid diagnostic procedures may indicate that a virus is responsible for causing a patient's problem, thereby avoiding antibiotic treatment and thus helping to curb the development of resistance through unwarranted uses of such drugs.

Proteomics

Knowing the complete set of genes within a microbe is just one step in understanding the biology of that organism. Proteins, which are encoded by the genes, do most of the actual work. Researchers therefore are beginning to use genomics in combination with other technologies to determine all of the proteins within an organism, a process dubbed proteomics. Scientists hope to identify each protein in a microbe, learn its functions, and decipher how all the proteins interact to help the organism survive. In the past, investigators have relied on specific biological properties of a microbe to hint at the possible presence of different proteins. In other cases, researchers have broken a virus or microbe into its chemical components to get a rough idea of the number and physical properties of its proteins. Scientists

1973

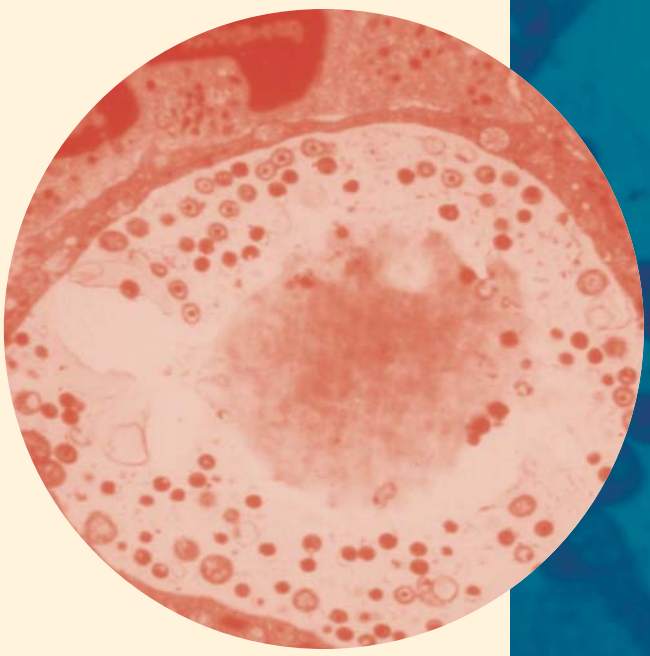
Stanley Cohen, Herbert Boyer
Performed the first successful recombinant DNA experiment, inserting a frog gene into bacteria

A CLOSER LOOK: Microbes and Chronic Diseases

Chronic *Chlamydia pneumoniae* Infections May Contribute to Heart Disease

The bacterium *Chlamydia pneumoniae* causes respiratory tract infections and is a common cause of pneumonia, accounting for about 10 percent of cases in the United States. Infections can also occur with only mild or no symptoms, however, and many adults throughout the world carry antibodies to the bacterium, indicating prior exposure. Researchers are now focusing considerable attention on *C. pneumoniae* because of its possible link to another serious illness—heart disease. Although the story is far from complete, medical researchers have compiled persuasive circumstantial evidence linking *C. pneumoniae* with the blood-vessel plaques that cause coronary heart disease. Studies suggest the persistent presence of these bacteria in blood vessels near the heart could provoke an exaggerated and damaging response of the host immune system.

If *C. pneumoniae* indeed is responsible for a substantial portion of coronary heart disease, it may be possible to reduce the incidence of that disease and save lives with a vaccine that can prevent chronic *C. pneumoniae* infections. Toward that end, researchers are now using the genomic sequence of *C. pneumoniae* and those of other closely related pathogens to identify specific genes whose products might be used in a vaccine to trigger a protective immune response. In addition, researchers are counting on genomic information to uncover other genes that enable the microbe to cause an array of damage in its human hosts.



Chlamydia Bacteria

have then relied on laborious biochemical analyses to isolate and identify individual proteins and determine their function. Genomics greatly accelerates that process because it provides a starting list of genes, and by extension the proteins they encode.

Combining Pathogen and Human Genomics

Researchers are also exploring the details of host-pathogen interactions. For instance, some pathogens make their way into host cells on specific receptors. Once identified, those entry ports could provide researchers with new means for interfering with such pathogens. In addition, detailed knowledge about the bevy of molecules human cells produce in response to certain pathogens may enable scientists to better orchestrate host defenses. At an even more detailed level, such analysis eventually may explain why some individuals are susceptible to a particular infectious disease while others are resistant.

On the pathogen side, researchers are focusing on the genes needed to invade, colonize, and disrupt normal functions in a human host. Cells and organ systems of the infected person also play important roles during such infections. Some cells serve as the target of the pathogen, while other cells and organs play an active role in a patient's immune response. Using powerful microarray analysis, researchers hope to identify the key features of human-pathogen interactions as another very promising way to tilt the contest in favor of humans.

For example, some researchers are studying how cells that are part of the host defense system respond during first encounters with pathogens. Initial experiments indicate that part of the first response appears to be the same regardless of the pathogen involved; several distinct groups of human genes switch on shortly after the cells are accosted by a variety of different types of microbial intruders. Other components of the immune system, however, respond to only one type of pathogen but not others.

1977

Frederick Sanger, Steven Niklen, Alan Coulson, Walter Gilbert, Allan Maxam
Developed techniques for determining DNA sequence

1984

Kary Mullis
Developed the polymerase chain reaction (PCR),which enabled amplification of minute quantities of DNA and revolutionized DNA studies

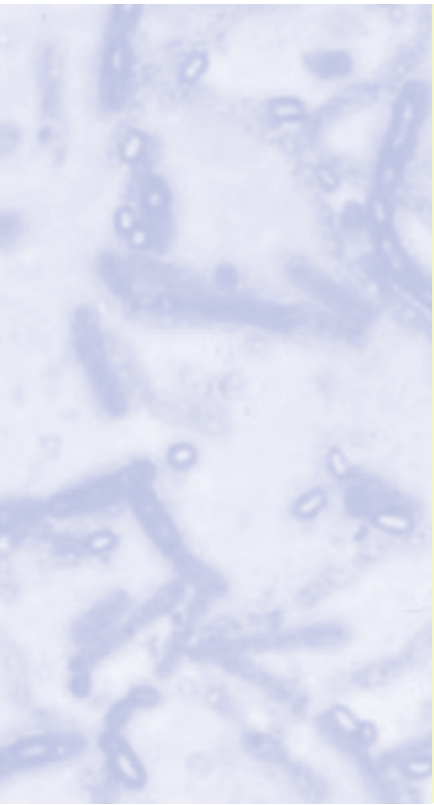
Microarrays

Because of genomics, new technologies are being designed to assist basic and applied research on microbial pathogens. One of these technologies, called a DNA microarray, combines DNA and computer technology to rapidly scan the genetic activity of a given organism. In a microarray, scientists coat a small chip with hundreds or thousands of fragments of DNA; each one corresponds to a different gene within the cell being studied. Those fragments can be made to light up when they encounter their partner gene, an event that means the gene is actively being read by the cell. Microarrays therefore enable researchers to see which genes are being switched on and off under different conditions. Scientists can use microarrays to determine the underlying genetics behind how a microbe responds to drugs, infects different hosts, survives outside of its hosts, and responds to different environmental conditions.

Genomics and Biodefense

The possibility of a bioterrorist attack on U.S. soil has shed new light on the importance of infectious diseases research. In response to that threat, NIAID and other federal health agencies are accelerating basic microbial research and developing new ways to diagnose, prevent, or treat infections that may be caused by the intentional release of a pathogen.

Microbial genomics is a vital part of a comprehensive approach to biodefense. The genomes of several smallpox virus strains are already known, as are those for different hemorrhagic fever viruses. NIAID currently funds the genome sequencing of many other potential bioterrorism microbes and collaborates with the Defense Advanced Research Products Agency (DARPA) to sequence such potential agents of bioterror as the bacteria that cause brucellosis, Q fever, gangrene, and epidemic typhus. In 2002, the genome sequences of the anthrax and plague bacteria also were completed.



Anthrax

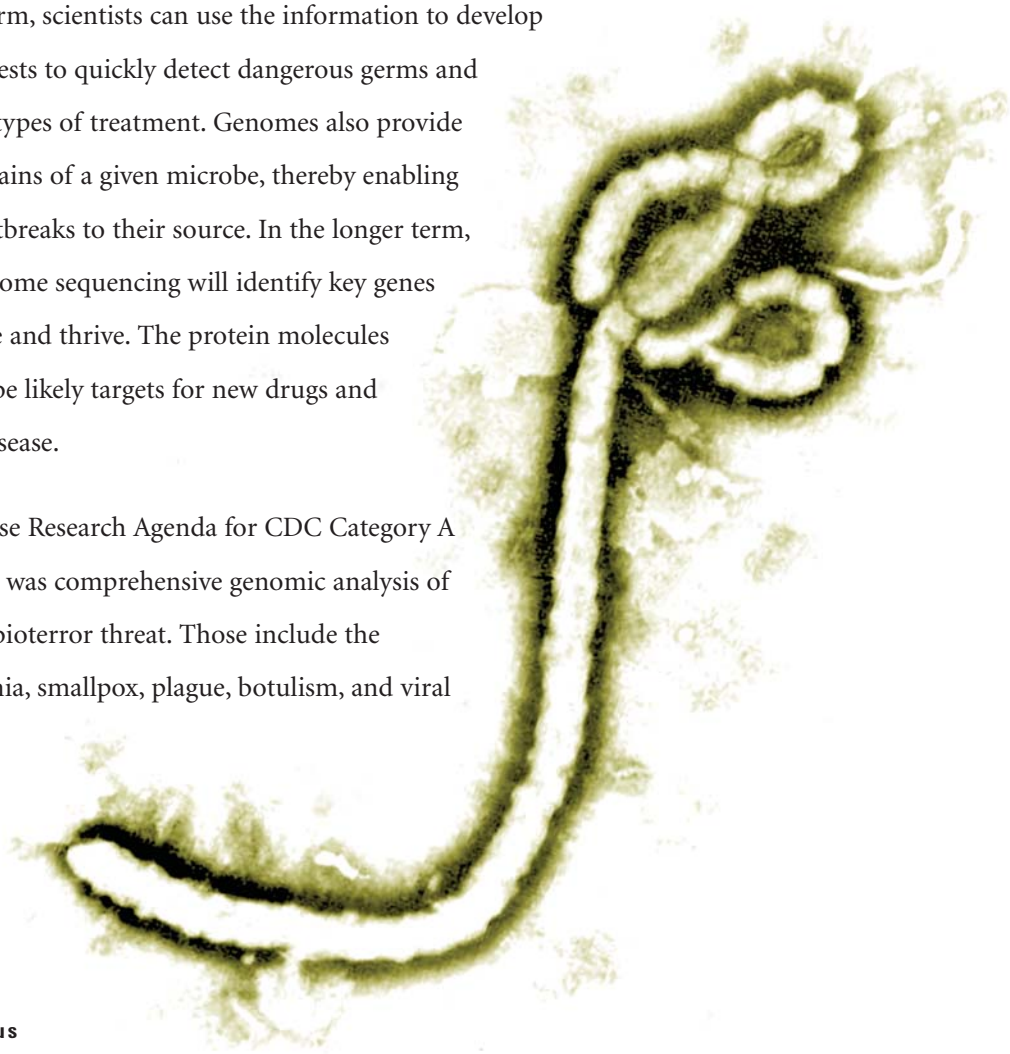
In 2001, a string of letters containing the anthrax bacterium were mailed to several destinations in the United States. The deadly bacterium, *Bacillus anthracis*, infected at least 18 people, killing eight. At the time the letters were mailed, researchers were well on their way to determining the genome sequence of the bacterium. After the mail attack, scientists used genome sequencing technology to determine the genetic blueprints of the strain isolated from one of the letters and compared it to the blueprints of other *B. anthracis* strains. As a result of that work, investigators gathered information on the potential source of the bioterror bacteria. In addition, the gene sequences provided researchers with the vital data required to quickly analyze the bacterium for potential vulnerabilities to new drugs and vaccines. Work on the *B. anthracis* genome continues as NIAID-funded scientists sequence additional strains.

1995

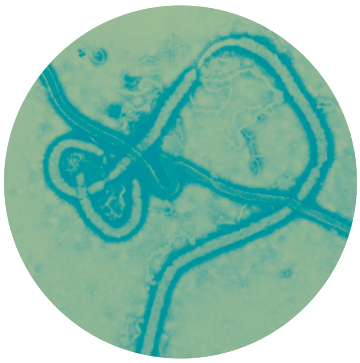
Multiple investigators
Determined first DNA sequence of an infectious bacterium, *Haemophilus influenzae*

Rapid sequencing of microbial DNA quickly unveils many secrets of bioterrorism's most frightening pathogens. In the short term, scientists can use the information to develop gene-based diagnostic and sampling tests to quickly detect dangerous germs and assess their susceptibility to different types of treatment. Genomes also provide molecular fingerprints of different strains of a given microbe, thereby enabling investigators to better track future outbreaks to their source. In the longer term, the genetic blueprints revealed by genome sequencing will identify key genes required for microbes to infect people and thrive. The protein molecules encoded by some of those genes will be likely targets for new drugs and vaccines to protect the public from disease.

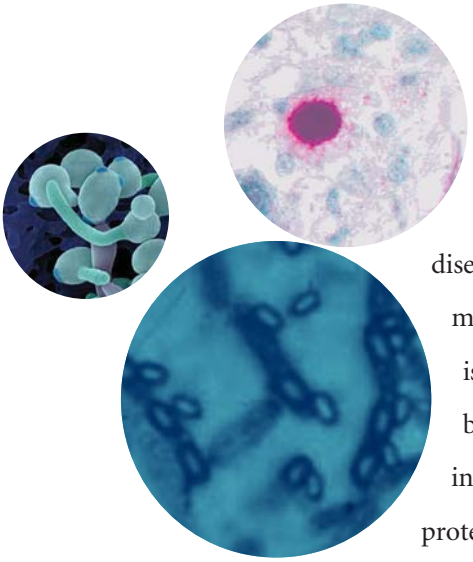
In 2002, NIAID released its "Biodefense Research Agenda for CDC Category A Agents." A key element to that agenda was comprehensive genomic analysis of the microbes considered the greatest bioterror threat. Those include the pathogens that cause anthrax, tularemia, smallpox, plague, botulism, and viral hemorrhagic fevers such as Ebola.



Ebola virus
Diagnostic specimen (above) from the first passage in Vero cells of a specimen from a human patient. This image is from the first isolation and visualization of Ebola virus, 1976.



Ebola virions are extremely varied in appearance. They are flexible filaments with a consistent diameter of 80 nanometers (nm), but they vary greatly in length and degree of twisting.



In Conclusion

As infectious diseases emerge, reemerge, and persist throughout the world, genomics research promises to be a valuable weapon in the fight to keep those diseases in check. By deciphering a pathogen's genetic blueprint, researchers can more quickly learn about the different genes and proteins required for that organism to survive, infect people, and cause disease. Some genes will point to unique biochemical processes that scientists can block with new drugs, thereby destroying a microbe without harming the people it infects. Other genes will identify key proteins on a pathogen that stimulate a person's immune response. Those proteins can then be used as the basis for new vaccines. Genome sequencing also can identify how microbes alter their makeup to avoid the effects of drugs or immune detection, enabling researchers to stay one step ahead by refining vaccines or medicines. Still other genes will provide potential targets for tests to rapidly diagnose an infection or identify the presence of harmful microbes in the field.

Pathogen genome sequencing therefore remains a high research priority at NIAID. Scientists continue to decipher the genetic secrets of infectious microbes and the organisms that transmit them to people. By understanding the underlying genes that make those microbes tick, researchers will continue to make advances in the diagnosis, prevention, and treatment of infectious diseases.

2002

Multiple investigators
Complete genomes published for malaria parasite and its mosquito vector

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
NATIONAL INSTITUTES OF HEALTH

National Institute of Allergy and Infectious Diseases

Division of Microbiology and Infectious Diseases

NIH Publication No. 02-4987
September 2002
www.niaid.nih.gov