The Search for Unrecognized Pathogens

David A. Relman
Stanford University School of Medicine, relman@stanford.edu

Relman, David A., "The Search for Unrecognized Pathogens" (1999). *U.S. Department of Veterans Affairs Staff Publications*. 21.
https://digitalcommons.unl.edu/veterans/21

The Search for Unrecognized Pathogens

David A. Relman

The distribution and diversity of microorganisms in the world are far greater than have been previously appreciated. Molecular, cultivation-independent methods have played a key role in this insight. To what extent do humans remain ignorant of microbial diversity within the human body and the settings in which microorganisms cause human disease? In addition to implicating microbial agents in nontraditional infectious diseases, the use of methods such as broad-range polymerase chain reaction, representational difference analysis, expression library screening, and host gene expression profiling may force a reassessment of the concepts of microbial disease causation.

Within the world of cultivated bacteria, viruses, and unicellular eukaryotes that inhabit the human body, the number and variety of relations between microbe and host are breathtaking. Yet, surveys of many terrestrial and aquatic ecosystems indicate that >99% of microorganisms resist cultivation in the laboratory and can only be identified or characterized with the use of molecular approaches (1). We presume to know much about the ecosystem most intimately related to our very existence—the microbial world within the human body. Nearly all available information about the composition of the endogenous microflora and the role of microorganisms in human disease has been based on laboratory cultivation. Are we as ill-informed about these human-associated segments of the microbial world? We would expect this not to be the case, because such a large effort is dedicated to microbial identification once humans develop clinical signs of infection. But, a number of observations suggest inadequacies in current diagnostic methods. The ensuing discussion describes alternative methods that have the potential to enhance the recognition and characterization of microbial disease, as well as to change the way we view disease causation.

The traditional diagnostic approach of isolating the etiologic agent or demonstrating a substantial change in antibody titer to a specific pathogen requires some understanding of growth conditions or specific antigenic determinants. Yet, these requirements have limited our ability to recognize and identify microbial pathogens. Beginning with the work of Robert Koch in the late 19th century, generations of epidemiologists, clinicians, and pathologists have confronted the same obstacles in their quest to explain disease patterns that are reminiscent of infection. The

Department of Medicine, Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305, USA; VA Palo Alto Health Care System 154T, 3801 Miranda Avenue, Palo Alto, CA 94304, USA. E-mail: relman@cmgm.stanford.edu

causes of Crohn's disease, ulcerative colitis, Wegener's granulomatosis, rheumatoid arthritis, tropical sprue, systemic lupus erythematosus, Kawasaki's disease, and other chronic diseases remain unknown despite disease features that are suggestive of infectious etiology.

Search Methods

Relying on an infectious agent to "perform" in the laboratory for the purposes of detection and identification poses a number of potential pitfalls. Identification of infectious agents by phenotypic traits is insensitive, because we are largely ignorant of the environmental conditions required by some microorganisms for growth and, hence, are unable to replicate them in the laboratory. Phenotypic traits are also difficult to quantitate, and they can be misleading. In contrast, genotypic traits generally provide reliable and quantifiable information for the identification and characterization of infectious agents. Moreover, genotypic traits such as genomic sequence can be analyzed directly in clinical specimens, without purification or isolation of the agent. Among the proven genotypic approaches for detecting and identifying previously unrecognized pathogens are broad-range polymerase chain reaction (PCR), representational difference analysis (RDA), and expression library screening (2, 3). With broad-range PCR, one amplifies genetic loci whose sequences provide reliable phylogenetic information about an infectious agent, using priming sites that are conserved among broad groups of agents. One of the most useful loci is the gene that encodes the small subunit ribosomal RNA. An extensive microbial ribosomal DNA (rDNA) sequence database enhances the usefulness of this marker for diagnostic discrimination and resolution. For example, previously unrecognized members of the bacterial domain can be identified with primers that recognize all bacterial small subunit rDNAs and amplify specific sequences (4–6); families of viruses can be explored with the broad-range PCR approach, using primers

that are complimentary to genomic sequences conserved among all family members (7). The approach is also amenable to high-throughput methods. However, the advantages of broad-range PCR are partially offset by the problem of microbial DNA contamination. Bacterial rDNA can be detected in clinical specimens from sterile sites of healthy individuals (for example, blood samples). Possible explanations include PCR reagent contamination, contamination of the specimen during collection, and the presence of bacteria or their DNA within sterile sites despite the absence of apparent disease. Few investigations have addressed this third, but most intriguing, possibility.

Other genotypic pathogen discovery approaches complement the advantages and disadvantages of broad-range PCR. RDA relies on subtractive hybridization to isolate rare genomic fragments (from an exogenous infectious agent) found in one (infected) specimen but not in an otherwise matched (uninfected) specimen. With a subsequent PCR step, these rare fragments are enriched. With RDA, one addresses the problem of specimen contamination and endogenous microbial sequences by using a matched specimen to reduce the presence of sequences that are common to the two specimens; however, unique genomic fragments from the infected specimen are enriched in a nonselective manner. These fragments may not provide specific information about the putative pathogen; that is, not all genome sequences provide a useful view of phylogenetic history. With expression library screening, pathogen-derived genomic fragments are identified from infected sites using host antisera, but only if the fragments encode antigenic products. Disease accompanied by autoreactive antibodies might pose additional problems. Some of the pathogens revealed with these approaches are listed in Table 1.

Any comprehensive pathogen detection effort faces additional challenges. In some instances, microorganisms cause disease from a distance. For example, toxic shock syndrome results from superantigen toxins released by *Staphylococcus aureus* or *Streptococcus pyogenes* strains that can be sequestered at unsuspected anatomic sites, and botulism usually occurs after ingestion of only the toxin. In other instances, infection initiates an immunologic response that results in disease after the causative agent has been cleared from the host. For example, *Campylobacter jejuni* is proposed to cause Guillain-Barré syndrome, an acute inflammatory neu-

ropathy, by eliciting cross-reactive antibodies to ganglioside that damage peripheral nerves weeks to months after intestinal infection (8). In all of these instances, nucleic acid from the pathogen would not be expected at sites of pathology or in the bloodstream. Of course, detection of nucleic acid does not indicate viability of the putative agent.

To meet some of these challenges, one might turn instead to the host as a source of infection-specific genotypic information. In collaboration with Patrick Brown at Stanford University and Lou Staudt at the National Cancer Institute, we are collecting and analyzing human gene expression responses to a wide variety of infectious agents and other noxious stimuli using microarrays containing complementary DNA (cDNA) for >15,000 human genes. cDNA microarrays measure steady-state messenger RNA (mRNA) abundance in a miniaturized, massively parallel hybridization format. Comparisons of mRNA levels in multiple samples are accomplished with differential fluorescence labeling (9). Our hypothesis is that microbial pathogens induce stereotypic mRNA response profiles that can be used to distinguish among different infectious agents or different pathogenetic mechanisms (2, 10). The complex repertoire of potential human gene transcriptional responses provides a far greater degree of specificity than do the traditional markers of acute inflammation, such as cytokines and acute-phase proteins like C-reactive protein (11). The goal with this approach is to recognize diagnostic signatures in clinical specimens from unexplained illness that would reveal the etiologic agent and the time of initial host exposure. With this approach, infectious diseases might be classified in a radically different manner, on the basis of how a particular host "sees" an integrated collection of virulence mechanisms.

Where to Look?

A wide range of acute and chronic diseases and syndromes warrants broad cultivation-independent searches for microbial pathogens. In short, there is a substantial amount of unexplained illness that resembles infectious disease. For example, most clinicians are familiar with a scenario in which a previously healthy individual develops an acute life-threatening illness with signs of infection and negative diagnostic test results. One of the first efforts to identify and analyze such cases is the Unexplained Deaths project coordinated by the Centers for Disease Control and Prevention (12). Our results from this ongoing project indicate that (i) unexplained death or critical illness occurs in 0.5 to 2.0 persons per 100,000 population in the United States per year, (ii) broad-range bacterial rDNA PCR reveals known disease-causing agents as a probable or definite pathogen in some cas-

Table 1. Examples of human disease and associated microbial pathogens revealed during the past 15 years. All of these infectious agents, except *Helicobacter pylori*, were first identified directly from clinical specimens using genotypic approaches.

Disease	Infectious agents	Reference
Peptic ulcer disease	*Helicobacter pylori*	(27)
Non-A, non-B hepatitis	Hepatitis C virus	(28)
Bacillary angiomatosis	*Bartonella henselae*	(5)
Whipple's disease	*Tropheryma whippelii*	(29)
Hantavirus pulmonary syndrome	Sin Nombre virus	(7)
Kaposi's sarcoma	Kaposi's sarcoma–associated herpesvirus	(30)

es, and (iii) little thought or attention has been directed at the selection and processing of clinical specimens that are appropriate for molecular pathogen discovery approaches.

A number of chronic inflammatory syndromes also bear the hallmarks of microbial disease (some listed above) and are only now, belatedly, the subjects of broad-range molecular pathogen identification methods (13). Crohn's disease, for example, involves the entire intestinal wall in a relapsing-remitting, acute and chronic granulomatous inflammatory process (14). However, any effort to find a microbial etiologic agent for Crohn's disease must address the issues of when and where it might be found in the course of this disease and how to deal with the complex, dense host-specific background of microorganisms within the intestines. Laser-capture microdissection (15) could complement all of the molecular approaches described above by focusing the search on suspect areas or cells

within a tissue section and assisting with the selection of a more closely matched control sample (Fig. 1).

The commensal microflora has been viewed as a source of specimen contamination and occasional opportunistic pathogens, but it may play a more important role in health and disease than was once thought. The intestinal commensal flora is associated with the pathogenesis of inflammatory bowel disease such as Crohn's disease, with nontropical sprue, and other disorders; it is also thought to play an essential role in intestinal epithelial differentiation. The endogenous flora of the mouth is associated with gingivitis, periodontitis, and systemic infections such as endocarditis. Yet, we are largely ignorant of the composition of and the temporal and spatial dynamics within these microbial communities (16, 17). This ignorance results in part from an excessive reliance on microbial cultivation. A substantial proportion of hu-
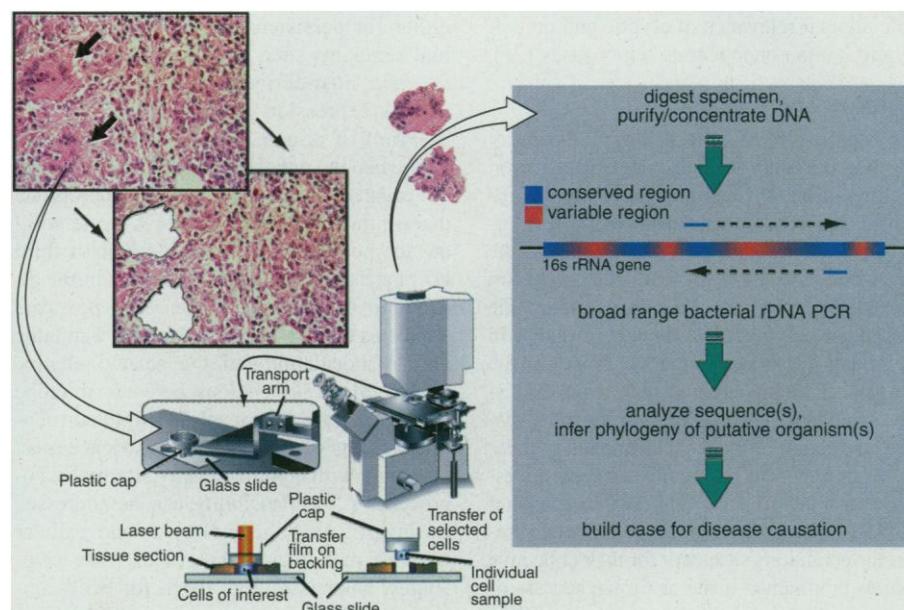
Fig. 1. A targeted broad-range rDNA PCR approach for pathogen discovery in Crohn's disease. Laser-capture microdissection allows one to isolate microscopic regions of tissue that have a greater likelihood of containing the pathogen (for example, granulomata or infected single cells). Granulomata are chronic inflammatory reactions that are frequently associated with persistent microorganisms or microbial antigen; they are found in tissues affected by Crohn's and other idiopathic diseases. Anatomic targeting is particularly important when there is gross microbial contamination of tissues, as in the intestinal wall from Crohn's disease. [Modified from Fig. 1 by Bonner et al. (15); histopathology courtesy of Donald Regula (Stanford University).]

man-associated microorganisms spends time as members of biofilms (for example, on the surface of teeth). As Costerton *et al.* point out, the sessile members of biofilms assume physiological states that render them resistant to cultivation (*18*). With sensitive genotypic methods, one task will be to identify the previously unrecognized members of our endogenous mucosal and cutaneous flora and then to track their possible translocation to other sites. For example, do chlamydia routinely traffic to vascular endothelia? Despite the recognized presence of archaea (methanogens) within the alimentary tract, why have no archaea been incriminated as pathogens? Are we not looking with appropriate methods in the right places? Another task will be to characterize quantitative and qualitative differences in community composition between different anatomic niches and hosts and changes over time.

Ecological concepts such as population heterogeneity and density may be critical in the transition from health to disease (for example, from plaque-associated gingivitis to adult-onset periodontitis). We will need new types of clinical measurements that reflect ecosystem health and stability (*19*). Our understanding of endogenous microflora may also require reassessment. For example, unsuspected endosymbiotic rickettsia-like *Wolbachia* species have been discovered in nematodes and have been found to play a role in host development (*20*). As we explore the details of the human genome and apply sensitive pathogen detection methods to human specimens, we should keep in mind the possible clinical relevance of cryptic and not-so-cryptic human endogenous retroviruses (*21*).

Microbes and Causation

The application of new pathogen discovery methods is certain to alter our perspective on microbial disease causation. For example, improvements in microbial detection and discrimination may generate data that challenge the traditional association of one microorganism with one diseased host. We currently accept the concept of polymicrobial disease causation in the setting of abscess formation; however, the same concept may apply to other types of disease, including rheumatic or "autoimmune" disease, and malignancy (*17*). Interaction and dependency between community members are essential features of microbial communities and biofilms (*18*). When a group of microbes creates the conditions necessary for their collective success in a human niche or for the success of a newcomer, unanticipated consequences might include disruption of local epithelial cells, induction of a persistent virus, or perturbation of the local innate immune system (*22*) and, ultimately, disease. Gingivitis, inflammatory bowel disease, and intestinal sprue may constitute disorders of the local endogenous microflora that result in impaired mucosal integrity and

function. In these situations, the linkage between microbe and disease is not linear and unidirectional. Genotype-based approaches that are quantitative, sensitive, and high throughput are needed.

Other nontraditional concepts of microbial disease causation pose challenges for pathogen discovery methods. Molecular mimicry is a phenomenon involving the induction of self-directed immunity by microbial epitopes (which resemble host epitopes); this phenomenon may explain the role of streptococci in rheumatic heart disease, chlamydia in coronary atherosclerosis (*23*), and *C. jejuni* with Guillain-Barré syndrome. The difficulty in detecting these pathogens at the site of pathology has been mentioned. As part of alternative strategies, one would like to use the host's immune responses to identify the antigenic stimulus. In addition to inducing self-immunity or immunity against the commensal flora, microbes are capable of inducing persistent changes in the local cytokine environment and in helper T cell expression profiles (*24*). These changes may protect the pathogen but may also result in damage to the host. Persistence of the microbial cell wall alone may be sufficient for these effects. Some of these mechanisms may play a role in the etiology of Crohn's disease, in which most or all of the causative agent has largely disappeared by the time the diagnostic features of the disease are recognized. In the absence of an ability to predict where the agent might be found at earlier time points, one must hope either for persistence of diagnostic microbial remnants such as sequence (Fig. 1) or specific host-derived signatures such as mRNA expression profiles.

Proof of disease causation is a daunting goal when the putative etiologic factor cannot be isolated or purified and an appropriate disease model is not available. A world without the possibility of satisfying Koch's third postulate demands equivalent rigor in the attribution of causation. (The third postulate stipulates that disease be induced in a suitable host by inoculation with the isolated putative agent.) Most associations between microbe and disease are noncausal. Multiple corroborating forms of circumstantial evidence must be sought with complementary alternative approaches (*25*); plausibility can be addressed by investigating the molecular and cellular basis of pathology (*26*). All of these challenges emphasize the need for well-controlled epidemiology, careful clinical and histologic observations, and increased attention to specimen collection and processing. One obvious payoff from the development and application of pathogen discovery approaches over the past 10 years has been the discovery of previously unrecognized disease-associated agents. An equally important and less

obvious payoff will be the reassessment of our relations with infectious agents and the way we view microbial disease causation.

References and Notes

1. R. I. Amann, W. Ludwig, K. H. Schliefer, *Microbiol. Rev.* **59**, 143 (1995); N. R. Pace, *Science* **276**, 734 (1997); D. M. Ward *et al.*, *Microbiol. Mol. Biol. Rev.* **62**, 1353 (1998); P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **180**, 4765 (1998).
2. D. A. Relman, *Emerging Infect. Dis.* **4**, 382 (1998).
3. N. Lisitsyn, N. Lisitsyn, M. Wigler, *Science* **259**, 946 (1993); S. J. Gao and P. S. Moore, *Emerging Infect. Dis.* **2**, 159 (1996).
4. C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990).
5. D. A. Relman, J. S. Loutit, T. M. Schmidt, S. Falkow, L. S. Tompkins, *N. Engl. J. Med.* **323**, 1573 (1990).
6. C. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6854 (1998). In working with fixed clinical specimens, one is usually limited to amplification of ≤500–base pair DNA fragments. Although the rDNA locus is universally conserved among cellular life, the breadth of a pathogen search can be "tuned" by selecting different consensus primers. Given these constraints and features, the small subunit rDNA is as useful as any locus and is more useful than most loci in providing reliable phylogenetic information.
7. S. T. Nichol *et al.*, *Science* **262**, 914 (1993).
8. A. F. Hahn, *Lancet* **352**, 635 (1998).
9. M. Schena *et al.*, *Trends Biotechnol.* **16**, 301 (1998); V. R. Iyer *et al.*, *Science* **283**, 83 (1999).
10. P. O. Brown and L. Hartwell, *Nature Genet.* **18**, 91 (1998); M. B. Eisen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
11. C. Gabay and I. Kushner, *N. Engl. J. Med.* **340**, 448 (1999).
12. B. A. Perkins *et al.*, *Emerging Infect. Dis.* **2**, 47 (1996).
13. D. N. Fredricks and D. A. Relman, *Curr. Clin. Top. Infect. Dis.* **18**, 180 (1998).
14. C. Fiocchi, *Gastroenterology* **115**, 182 (1998).
15. R. F. Bonner *et al.*, *Science* **278**, 1481 (1997).
16. P. E. Kolenbrander and J. London, *J. Bacteriol.* **175**, 3247 (1993).
17. P. G. Falk, L. V. Hooper, T. Midtvedt, J. I. Gordon, *Microbiol. Mol. Biol. Rev.* **62**, 1157 (1998).
18. J. W. Costerton, P. S. Stewart, E. P. Greenberg, *Science* **284**, 1318 (1999).
19. S. Naeem and S. Li, *Nature* **390**, 507 (1997).
20. A. Hoerauf *et al.*, *J. Clin. Invest.* **103**, 11 (1999).
21. R. Lower, J. Lower, R. Kurth, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5177 (1996).
22. J. A. Hoffmann, F. C. Kafatos, C. A. Janeway Jr., R. A. B. Ezekowitz, *Science* **284**, 1313 (1999).
23. K. Bachmaier *et al.*, *ibid.* **283**, 1335 (1999).
24. M. Wilson, R. Seymour, B. Henderson, *Infect. Immun.* **66**, 2401 (1998).
25. D. N. Fredricks and D. A. Relman, *Clin. Microbiol. Rev.* **9**, 18 (1996); P. S. Moore and Y. Chang, *Am. J. Epidemiol.* **147**, 217 (1998).
26. P. S. Moore, C. Boshoff, R. A. Weiss, Y. Chang, *Science* **274**, 1739 (1996); C. Boshoff *et al.*, *ibid.* **278**, 290 (1997); C. Bais *et al.*, *Nature* **391**, 86 (1998). Kaposi's sarcoma–associated herpesvirus was discovered with RDA. Examination of several open reading frames suggests viral mechanisms for stimulating cellular proliferation and angiogenesis.
27. B. J. Marshall and J. R. Warren, *Lancet* **i**, 1311 (1984); B. J. Marshall, *JAMA* **274**, 1064 (1995); A. Covacci, J. L. Telford, G. Del Guidice, J. Parsonnet, R. Rappuoli, *Science* **284**, 1328 (1999).
28. Q.-L. Choo *et al.*, *Science* **244**, 359 (1989).
29. K. H. Wilson, *et al.*, *Lancet* **338**, 474 (1991); D. A. Relman, T. M. Schmidt, R. P. MacDermott, S. Falkow, *N. Engl. J. Med.* **327**, 293 (1992).
30. Y. Chang *et al.*, *Science* **266**, 1865 (1994).
31. I thank J. Theriot (Stanford University) for a critical reading of this manuscript and K. Morrow (VA Palo Alto) for assistance with Fig. 1. D.A.R. is supported by grants from the National Institute of Allergy and Infectious Diseases at NIH, the U.S. Department of Veterans Affairs, the U.S. Environmental Protection Agency, and the Defense Advanced Research Projects Agency.