Prospects for applying virulence factor–activity relationships (VFAR) to emerging waterborne pathogens

Gerard A. Cangelosi

ABSTRACT

This paper addresses the potential for using virulence factor–activity relationships (VFAR) to discover and detect emerging waterborne pathogens. Emerging pathogens present two challenges to drinking water safety. The first is to improve the predictive value of tests for novel organisms that have recently been identified as emerging waterborne health threats, but are not yet well characterized. There are potential applications for VFAR in meeting this challenge, most notably to distinguish pathogens from closely related nonpathogenic organisms that may also be present in water samples. The second challenge is to prospectively identify new organisms that are not yet known to be health risks to humans, but have the potential of becoming so. This goal is more ambitious. Many microbial virulence factors have multiple functions, some of which are innocuous. Moreover, dedicated virulence factors can be found in pathogens that infect non-human hosts, but pose no threat to humans. These issues complicate the near-term use of VFAR as a first line indicator of pathogenicity. However, our understanding of microbial pathogenesis continues to expand exponentially. As known virulence factors increase in number, so too does the feasibility of a prognostic VFAR strategy.

Key words | biomarkers, host range, *Legionella*, *Mycobacterium*, *Vibrio*, zoonoses

Gerard A. Cangelosi
Seattle Biomedical Research Institute,
Seattle WA,
USA
Departments of Global Health and Epidemiology,
University of Washington,
Seattle, WA 98109-5219,
USA
E-mail: *jerry.cangelosi@sbri.org*

INTRODUCTION

The Institute of Medicine and the Centers for Disease Control classify infectious diseases as "emerging" if their incidence in humans has increased in the past two decades, or threatens to increase in the near future (http://www.cdc.gov/ncidod/EID/about/background.htm). The term is sometimes applied to well-known diseases, such as malaria, which have resurged in parts of the world. However, it is more often applied to new pathogens that evolve from animal pathogens (e.g. HIV), to known human infections that have spread to new geographic areas or populations (e.g. hepatitis E, West Nile virus) or to infections that have increased with the increased population of humans who are elderly, immunocompromised or otherwise susceptible to rare diseases (e.g. *Mycobacterium avium* complex, Legionnaire's disease).

When an infectious disease emerges, there is a need to rapidly characterize its biology, ecology and epidemiology. There may also be a need for new methods to detect the pathogen in water and other sources of environmental exposure. The VFAR concept was developed in part to help meet these needs (Board on Environmental Studies & Toxicology 2001; NRC 2002; Jenkins *et al.* 2004; Tourlousse *et al.* 2007). New genetic, genomic and bioinformatic approaches make it possible to identify virulence factors–the genetic "teeth and claws" that pathogens use to infect and survive in human hosts and, in some cases, cause disease–more quickly than was previously possible. Such factors can help categorize hypothetical etiological agents as likely or unlikely pathogens. They can also be applied to environmental monitoring and surveillance, as markers

doi: 10.2166/wh.2009.045

to help distinguish new pathogens from closely related non-pathogenic organisms that might be found in the same environmental samples.

The bipartite definition of emerging pathogens (recently increased incidence or threat of future increase) presents two challenges relevant to drinking water safety. The first is to rapidly develop and maximize the predictive value of methods for detecting recently emerged human pathogens in environmental samples. The second (and much more ambitious) goal is to prospectively identify new organisms that are not yet known to be health risks to local populations, but have the potential of becoming so. This paper explores the feasibility of applying VFAR to both goals.

DETECTION OF KNOWN EMERGING PATHOGENS IN DRINKING WATER

Virulence is the result of an evolved strategy for replicating on or within a host, leading to pathology. The pathology may result from infection of normally sterile sites in the host, because the mechanisms for subverting or avoiding host defenses cause host damage or because the mechanisms for transmission to new hosts cause host damage. Such strategies are hypothesized to be relatively rare among organisms that have not evolved in close contact with animal hosts. However, a number of known human pathogens exist primarily in the external environment. There are reasons to believe that there are also currently unrecognized pathogens that reside primarily in the external environment, including drinking water.

This section focuses on applications for VFAR in improving the predictive value of methods used to detect known emerging pathogens in drinking water. In the context of drinking water safety, "predictive value" refers to the probability that a positive test for the presence of a known human pathogen corresponds to a true threat to human health (i.e. it predicts or explains an elevated incidence of disease among susceptible people who are exposed).

Virulence as a variable trait

Tests for microbial contaminants in water are mechanistically diverse, ranging from traditional microbial cultivation methods to sophisticated tests for molecular markers in environmental samples. A shared feature is that they correlate taxonomic identification with the risk of disease. For example, a water supply that tests positive for *Cryptosporidium parvum* is understood to present a risk of Cryptosporidosis. The assumption is that all cells and strains of this pathogen have approximately equal potential to cause disease. This assumption breaks down when applied to pathogen species that are internally diverse with regard to virulence potential. Detection of non-virulent strains constitutes background "noise" that reduces the predictive value of pathogen detection.

In the case of many environmental pathogens such as *Pseudomonas aeruginosa*, *Legionella pneumophila* or *Mycobacterium avium* complex (MAC), most cell divisions are thought to occur outside of human hosts. These organisms occupy diverse environments and do not need hosts for replication. Therefore, there is evolutionary opportunity for individual strains and isolates to lose the ability to infect humans. For example, MAC occurs naturally in over 25 different serotypes, of which only a handful are commonly isolated from human samples (Falkinham 1996*a*,*b*; Cangelosi *et al.* 2003; Turenne *et al.* 2008). Genetic heterogeneity is a hallmark of MAC (Turenne *et al.* 2007). Recently, a combination of microarray and polymerase chain reaction (PCR) was used to compare the genomes of 44 *M. avium* strains to that of the genome sequence strain, *M. avium* subsp. *hominissuis* 104 (Semret *et al.* 2004). These methods detect large-sequence polymorphisms (LSPs), typically defined as insertions or deletions (indels) that encompass one or more complete genes. Many *M. avium* strains exhibited LSPs relative to strain 104 that encompassed >10% of the strain 104 genome. These analyses generally do not include the second MAC species, *M. intracellulare* (MI), which may exhibit even greater divergence from *M. avium* strain 104. In comparison, similar approaches applied to *M. tuberculosis*, a close cousin of MAC that is a near-obligate human pathogen, found <4% variation between isolates (Kato-Maeda *et al.* 2001; Tsolaki *et al.* 2004).

Microarrays and other methods have also revealed extensive LSPs among strains of *E. coli* (Fukiya *et al.* 2004), *V. cholera* (Thompson *et al.* 2003) and *C. parvum* (Morgan *et al.* 1999; Xiao *et al.* 1999, 2000, 2002; Caccio *et al.* 2001; Widmer 2004). Some indels result from loss (deletion) of

genetic information, while others result from acquisition (insertion) of information. Genetic information can be acquired from a variety of external bacterial, eukaryotic and viral sources. This may occur frequently in biofilms and other complex natural environments. Waterborne pathogens believed to carry externally acquired "pathogenicity islands" and other genes include the bacterial genera *Legionella, Helicobacter, Escherichia, Vibrio, Mycobacterium* and *Pseudomonas* (Hacker & Carniel 2001; Cohen *et al.* 2007; Danelishvili *et al.* 2007; Izutsu *et al.* 2008) and the protozoan *Cryptosporidium* (Striepen *et al.* 2004). Some genomic clades of a species can have full complements of virulence-related genes while others do not. For example, *V. cholera* exhibits variable occurrence of its major pathogenicity island and other virulence genes in waterborne isolates, even within the O1 serogroup that is normally associated with disease (Faruque *et al.* 1998; Vital Brazil *et al.* 2002).

Most genetic and genomic analyses are conducted on clinical isolates of environmental pathogens, which can be expected to possess full complements of virulence genes. Environmental isolates are examined far less frequently, and are likely to exhibit greater diversity than clinical isolates. This can be tested by using genomic approaches such as microarray-based whole-genome comparisons. However, it is important to note that quantitative genomic diversity does not always correspond to diversity in the ability to cause disease. For many waterborne pathogens, the extent of phenotypic diversity with regard to virulence remains unknown. In some cases this can be investigated through the use of non-human disease models, which can quantify virulence-associated phenotypes such as the ability to invade and grow within host cells (Widmer *et al.* 1998; Ohkusu *et al.* 2004; Widmer 2004; Danelishvili *et al.* 2007).

For those pathogens which are confirmed by phenotypic and genotypic investigations to be diverse with regard to the ability to cause disease, tests for known virulence markers can help to improve the positive predictive value of their detection in drinking water.

The challenge of host range

VFARs have been compared to structure–activity relationships (SARs) in that both link innate properties of a contaminant (biological properties of microorganisms and structural properties of chemical compounds) with a real or potential ability to harm humans (Jenkins *et al.* 2004). However, microbial contaminants may differ from chemical contaminants when it comes to host range. It can be difficult to differentiate virulent strains of a pathogen that infect humans from closely related strains or species that infect other animals but not humans. Virulence factors present in strict animal pathogens pose a challenge to VFAR-based pathogen detection.

Many emerging human pathogens have close relatives that infect animals. For example, the Simian Immunodeficiency Virus (SIV), from which the Human Immunodeficiency Virus (HIV) emerged in recent human history, is nearly identical to HIV in terms of genetic organization. There are nucleotide and amino acid sequence differences between orthologous genes in the two viruses: however, in quantitative terms this diversity is no greater than that seen between two strains of HIV, a famously hypermutable pathogen.

A similar situation is seen within the *M. tuberculosis* complex, which encompasses six closely related species that include human pathogens (*M. tuberculosis*, *M. africanum* and *M. canettii*), pathogens with broad host range (*M. bovis* and *M. pinnipedii*) and a strict rodent pathogen (*M. microti*). Species within the complex share 99.9% sequence similarity and they share most known virulence-associated genes (Sreevatsan *et al.* 1997; Brosch *et al.* 2001, 2002). In both HIV and *M. tuberculosis*, most known virulence factors do not differ in animal and human pathogens.

AIDS and TB have been subjected to intensive research leading to the identification of at least a few reliable genotypic markers to differentiate human species from animal species. For example, the ESAT6 region of the *M. tuberculosis* genome is missing in *M. microti* (Brodin *et al.* 2002). The same is rarely true of environmental and emerging pathogens, most of which remain inadequately characterized relative to HIV and *M. tuberculosis*. Very few host range factors have been identified in environmental pathogens.

Impact of genetic variability and host range: the *Mycobacterium* example

The genus *Mycobacterium* offers a good test case for the VFAR concept. Extensive characterization of *M. tuberculosis* has identified a set of well-accepted mycobacterial virulence

factors. The genus also includes a large number of environmental and zoonotic species that are emerging as pathogens of susceptible human populations (non-tuberculous mycobacteria, or NTM). Many of these species can be isolated from water supplies. Based on phenotypic as well as genotypic criteria, the genus is divided into two distinct phylogenetic clades, the "slow growing" and "fast growing" species (Harmsen *et al.* 2003). The slow growers include *M. tuberculosis* and most, but not all, of the other human pathogenic species of *Mycobacterium*. Knowledge derived from the characterization of *M. tuberculosis* can be applied to the less well-characterized NTMs, especially slow growing species.

Protein products of *M. tuberculosis* that function directly as virulence factors include the early secreted antigen target EsxA (ESAT6), the iron-regulated heparin binding hemagglutinin HbhA, the exported repetitive protein Erp (PirG) and the ESX5 export complex, which includes the required virulence gene Rv1798 (Brodin *et al.* 2002; Cosma *et al.* 2003; Abdallah *et al.* 2006). Table 1 compares the predicted occurrence of these proteins in environmental and zoonotic mycobacteria for which relevant genomic sequence data are available. These include the zoonotic pathogen *M. bovis,* which is nearly identical to *M. tuberculosis*; the fish pathogens *M. marinum* and *M. shottsii,* which are closely related to *M. tuberculosis* and *M. bovis*; the water-associated opportunistic pathogens *M. avium* subsp. *homissuis* and *M. abscessus*; and the non-pathogenic species *M. smegmatis.* Not surprisingly, the genome sequences of *M. bovis* and *M. marinum* include strongly homologous orthologs of all four factors. *M. avium* subsp. *homissuis* has two of them, *hbhA* and ESX5. *M. abscessus,* a fast growing species that is an opportunistic pathogen of humans and is acquired from water (Petrini 2006), lacks strong homologs to any of the four virulence factors, consistent with its greater phylogenetic distance from *M. tuberculosis*. The same is true of the fast growing non-pathogen *M. smegmatis,* with the exception of ESAT6, which is strongly homologous to the *M. tuberculosis* virulence factor.

Several issues are evident from Table 1. First, the extent to which the NTMs in the table "possess" or "lack" the four *M. tuberculosis* virulence markers is somewhat subjective. In theory, an ortholog that exhibits modest homology to a

Table 1 | Virulence factors in environmental and zoonotic mycobacteria

Species	Classification	Disease in humans	Animal hosts	Percent amino acid identity to *M. tuberculosis* gene			
				ESAT6 (*esxA*)	Heparin-binding hemagglutinin (*hbhA*)	Exported repetitive protein (*erp*)	Rv1798 (ESX5)
M. bovis	Slow growing	Yes	Cattle, badgers, others	100%	100%	100%	100%
M. marinum	Slow growing	Yes (skin)	Fish, amphibians	91%	88%	77%	93%
M. shottsii	Slow growing	No	Striped bass	Unknown	Unknown	80%	Unknown
M. avium ssp. *hominissuis*	Slow growing	Yes	Swine, others	27%	87%	60%	89%
M. abscessus	Fast growing	Yes	Fish, others	34%	58%	42%	32%
M. smegmatis	Fast growing	No	None	91%	69%	55%	32%

known virulence factor in *M. tuberculosis* protein could perform the exact same function. Conversely, an ortholog with 99% amino acid identity to a known virulence factor might perform an entirely different (and innocuous) function, if its active site differs. Thus, even strong sequence homology can be risky to interpret. The intermediate levels of homology seen in most of Table 1 are even more so.

If we arbitrarily choose a threshold level of homology to indicate a functional match with a virulence factor, say 80% amino acid identity, then the issue of host range becomes evident from Table 1. For example, *M. marinum* is a virulent pathogen of fish and amphibians. Outbreaks of human infections have been associated with swimming, bathing, fish processing and other water-related activities. Such infections are almost always confined to skin lesions, most likely because the organism cannot replicate at human body temperature. *M. marinum* is incapable of disseminated infection and cannot be acquired by inhalation or ingestion. *M. marinum* is closely related to *M. shottsii* and *M. pseudoshottsii*, recently described fish pathogens that have never been associated with human disease (Rhodes *et al.* 2005; Kaattari *et al.* 2006). Together with *M. tuberculosis, M. bovis, M. marinum, M. ulcerans* and a few other species, these organisms form a phylogenetic cluster called the "*M. tuberculosis* clade" that share >98% sequence homology in their small subunit rRNA genes (Kaattari *et al.* 2006). The *erp* gene of *M. shottsii* has been sequenced and its homology to the *M. tuberculosis erp* gene exceeds that most of the human pathogens in Table 1. The other three virulence genes have not yet been sequenced in *M. shottsii*: however, given its pathogenic lifestyle and its close phylogenetic relationship with *M. tuberculosis*, it is reasonable to expect strong homology, at least of the order of *M. marinum*. Nevertheless, this water-associated bacterium poses no threat to humans. It may lack other, yet-to-be identified, genetic factors that are required for infection of humans or it may simply be unable to replicate at human body temperature.

Table 1 illustrates additional pitfalls for VFAR strategies. ESAT6 is present in the non-pathogen *M. smegmatis* but absent in the waterborne pathogen *M. avium* subsp. *hominissuis*. This illustrates the apparently multi-functional nature of many virulence factors, a problem that has been cited elsewhere (Tourlousse *et al.* 2007). Conversely, *M. abscessus*, a clinically significant NTM, lacks all known mycobacterial virulence factors. *M. avium* subsp. *hominissuis* is closely related to other *M. avium* serotypes (e.g. 3–6 and 8–11) that are almost never associated with human disease, but are likely to carry most, if not all, of the same virulence factors.

To address these challenges, realistic VFAR strategies must incorporate very large numbers of virulence factors in combination with statistical methods to discern consequential from non-consequential matches. In addition, the characterization of virulence factors will not by itself be sufficient for the application of VFAR to human health; "host range" factors must also be characterized. In some cases, host range may simply be a function of growth temperature range, a parameter that may be difficult to discern by genotypic analysis. An alternative approach, described in the next subsection, is to combine VFAR methods with traditional phylogenetic approaches. The combined approaches may offer predictive values that exceed the sums of their parts.

Combined approaches incorporating virulence factors and phylogeny

In some cases phylogenetic markers have been identified that segregate with virulence and host range. For example, the mobile genetic element IS901 is common in bird isolates of *M. avium,* but is almost never seen in isolates from humans and other large mammals (Kunze *et al.* 1992; Bono *et al.* 1995; Pavlik *et al.* 2000). IS901 carries no genetic information other than that needed for its own transposition, so the relationship between this element and preference for bird hosts is likely to be a matter of phylogeny (i.e. the evolutionary lineage that co-evolved with birds happens to have acquired or retained the element, while that which co-evolved with humans did not). Nonetheless, phylogenetic markers such as IS901 may be useful adjuncts to virulence markers when the goal is to detect strains of a pathogen that are virulent to humans. Interestingly, it was demonstrated that insertion elements can affect the expression of neighboring genes in mycobacterial genomes, sometimes with visible phenotypic consequences (Laurent *et al.* 2002; Safi *et al.* 2004). Thus, phylogenetic and functional markers may in some cases be the same.

Identification of new virulence and host range markers

When a new pathogen is discovered, several approaches can be brought to bear relatively quickly to identify its virulence factors and, in some cases, host range factors. When known virulent and non-virulent (or human and animal) isolates are available, comparative genomic methods such as microarrays, subtractive hybridization or high-throughput genomic sequencing can identify DNA sequences that are unique to virulent human isolates. These approaches are rapid but limited by the size and diversity of the samples tested. Moreover, many of the markers they identify are "guilty by association", i.e. present in virulent genotypic clades but not necessarily directly involved in virulence.

True VFARs can be identified by using high-throughput mutational approaches such as signature tagged mutagenesis (Chiang *et al.* 1999) and transposon-site hybridization (Sassetti & Rubin 2003). These methods use sequence tagging systems or microarrays to identify large numbers of random transposon insertion mutants that are unable to grow under specific conditions such as growth in animal models, but not under other conditions such as free-living laboratory cultures. Genes that are disrupted in such mutants are assumed to be specifically required for life in animal hosts (some such genes are true virulence factors, while others code for more general metabolic and catabolic functions that are used in host environments). These approaches have mainly been applied to bacterial pathogens, including *Legionella pneumophila* (Edelstein *et al.* 1999) and *Vibrio cholera* (Merrell *et al.* 2002).

In some cases, it is possible to directly screen transposon mutant libraries for mutants that fail to exhibit virulence-associated properties that are visible during free-living growth in the laboratory. For example, MAC grown on agar plates forms multiple colony types that vary with regard to virulence. Mutagenesis of "white" colony type variants, which are virulent and multi-drug-resistant, has yielded a large number of mutants that form "red" colonies, which normally are non-virulent. Some of these factors were subsequently shown to be required for virulence in disease models (Philalay *et al.* 2004; Cangelosi *et al.* 2006).

An alternative strategy is to identify genes that are differentially expressed, if not necessarily required, during growth in or on host cells. There are many ways to identify differentially expressed genes; examples that have been applied to waterborne and environmental pathogens include promoter probing (Rankin *et al.* 2002), differential display (Schroeder *et al.* 1999), selective capture of transcribed sequences (SCOTS) (Hou *et al.* 2002) and *in vivo* induced antigen technologies (IVIAT) (Rollins *et al.* 2008). Others have identified genes that are upregulated during growth in macrophages (Li *et al.* 2005). Subsequent mutational and comparative analyses often result in the positive identification of virulence factors.

The research methods used to identify virulence factors in emerging pathogens may eventually be applied to the question of host range. For example, it has been observed that IS901-positive strains of *M. avium* isolated from birds express a protein, p40, which is not expressed in other phylogenetic clades of MAC, including human isolates of *M. avium* (Inglis *et al.* 2001). Interestingly, the genetic information for p40 is present in almost all mycobacterial pathogens, but its high-level expression is seen only in bird isolates of MAC. In a separate microarray-based study, Bermudez and colleagues identified several putative virulence genes, including a putative pathogenicity island, that occur variably in *Mycobacterium* species with different host ranges (Harriff *et al.* 2008). Expansion of such analyses may help to identify factors that are potentially useful as "host range factor–activity relationship" markers.

Conclusions: use of VFAR to improve the predictive value of detecting known emerging pathogens in drinking water

Virulence is a variable trait within many pathogen species. Detection of non-virulent strains compromises the predictive value of pathogen detection. In such cases, VFAR can be a useful adjunct to purely phylogenetic approaches to pathogen detection. When VFARs are used to detect human pathogens, virulence factors carried by closely related microorganisms that infect animals but not humans are potential sources of background. Few "host range factor–activity relationships" have been identified. However, the methods used to identify virulence factors can be brought to bear toward this end. In some cases, phylogenetic markers of pathogen clades that infect humans vs. animals may be

used in conjunction with VFAR to improve specificity for human pathogens. Traditional taxonomic identifiers of pathogenic species will remain important and useful as adjuncts to VFAR.

USE OF VFAR TO DISCOVER NEW PATHOGENS

An ambitious goal for the use of VFARs is the predictive identification of emerging pathogens *before* they are conclusively correlated with a disease outbreak (Board on Environmental Studies & Toxicology 2001; NRC 2002; Jenkins *et al.* 2004). For example, *Legionella pneumophila* has been found in clinical and environmental samples dating back to 1947, but was not known to cause human disease until after the famous Legionnaire convention outbreak in Philadelphia in 1976 (Ching & Meyer 1987). Had our *current* knowledge of virulence factors existed at the time, and had modern methods for detecting such factors been available, might it have been possible to identify *L. pneumophila* as a dangerous waterborne pathogen prior to 1976? Probably not, given the complexities discussed above and the fact that our understanding of microbial pathogenesis remains incomplete. For example, analysis of the factors in Table 1 would fail to identify *M. abscessus* as a human pathogen and might instead falsely identify *M. shottsii* as a potential threat. That being said, our understanding of pathogenesis continues to expand exponentially and insights into host range are starting to emerge. Thus, a reliable VFAR strategy remains a feasible goal for the future, and a strong rationale for continued research into the mechanisms of microbial pathogenesis.

Sources and occurrence of virulence factors in the environment

The evolution of microbial pathogenesis is relevant to the development of VFAR strategies. Many selective pressures that microbes encounter in human hosts are also encountered in the environment. Mechanisms that microbes evolve to cope with these pressures can have multiple uses. For example, the formation of biofilms is an ecological strategy that improves microbial persistence in many habitats (Costerton *et al.* 1999; Donlan 2001). This adaptation may have enhanced the abilities of some microbes to resist clearance by immune defenses in host tissues. Many microbial factors that enable persistence in hosts, especially adhesion factors, fimbria, scavenging mechanisms and quorum-sensing systems, may also play important roles in persistence in the environment. Such factors can be found in numerous non-pathogenic microorganisms that adhere to surfaces in the environment and would not be ideal choices for VFAR analysis.

Even pathogen-specific factors, such as the serovar-specific glycopeptidolipids (ssGPLs) of MAC, can play dual roles. Genes encoding ssGPLs can occur variably in MAC strains and have been implicated in virulence of MAC (Krzywinska & Schorey 2003; Krzywinska *et al.* 2005). However, they are also required for biofilm formation and planktonic dispersal of the pathogen in water habitats (Freeman *et al.* 2006). Thus, the presence of ssGPL genes is not by itself evidence of a pathogenic lifestyle.

Many microbial factors important to human disease are adaptations to living within protozoal or animal hosts. *Legionella pneumophila,* for example, lives within phagocytic environmental amoebae, an adaptation that may have enabled it to adapt to life within the professional phagocytic cells of humans (Swanson & Hammer 2000; Gal-Mor & Segal 2003). The same may be true of MAC (Cirillo *et al.* 1997; Steinert *et al.* 1998; Miltner & Bermudez 2000). *Vibrio, Aeromonas* and *Pseudomonas* species colonize the surfaces of zooplankton using bacterial factors that also contribute to colonization of the human tissues (Dumontet *et al.* 1996; Colwell *et al.* 2003; Pruzzo *et al.* 2003). As with biofilm formation genes, these factors are widespread in the environment and far from specific to human pathogens.

An additional source of background already discussed is strains and species that infect animals but not humans. Thus, a great many factors cited in discussions of VFAR (Board on Environmental Studies and Toxicology 2001; Jenkins *et al.* 2004) are far from specific to human pathogens. There is a need for improved understanding of virulence, taking these issues into account.

Additional challenges

In applying VFAR to the predictive identification of new human pathogens, a "positive result" would be the accurate

prediction of a new waterborne disease. Such events are very rare in drinking water public health. Therefore, the validation of this approach must be viewed as a long-term endeavor.

From the technical standpoint, the use of microarrays and other high-throughput methods to directly interrogate complex environmental samples for very large numbers of virulence genes (Board on Environmental Studies and Toxicology 2001; Jenkins *et al.* 2004) might be problematic, because the analytical sensitivity of such methods is limited. To yield enough DNA for detection, a given pathogenic organism would have to be present in large numbers. This can be addressed to some extent by incorporating DNA amplification into the labeling steps or by using maxi-arrays that sacrifice throughput in exchange for sensitivity. The throughput afforded by microarrays, which can interrogate samples for tens of thousands of genes, might not be needed for identifying virulence factors that might number in the hundreds.

Steps toward useful VFAR strategies

In order to realize the VFAR paradigm, the most critical unmet need is a greatly expanded understanding of how microbial pathogenesis works. Current research in this area is focused on a few pathogens, most of which are not water-derived. Because mechanisms of pathogenesis vary between groups of organisms, there is a need to expand research on water-associated pathogens.

As our understanding of microbial pathogenesis improves, databases of known and potential VFARs (Jenkins *et al.* 2004; Tourlousse *et al.* 2007) will become increasingly useful. A potentially useful addition would be a database of VFARs detected in complex water samples.

A database that correlates specific factors with specific microbial species might also be useful. Such a database could be developed by cultivating microorganisms from water and applying high-throughput methods to interrogate each isolate individually for the presence of virulence factors. Additional taxonomic data, such as small subunit rRNA sequences, could be collected and cross-referenced with virulence factor complements. When future outbreaks of unknown etiology are associated with water supplies, the database can be consulted in conjunction with traditional epidemiological methods. Microorganisms that possess virulence factors would be prioritized over other microorganisms in the search for novel etiological agents. Unfortunately, such a database would not include virulence factors present in non-culturable microbial species, an important goal of the VFAR approach (Board on Environmental Studies and Toxicology 2001).

An additional measure recommended recently (Cangelosi & Freitag 2005) is the integration and consolidation of existing databases. Currently, most databases of information on microorganisms specialize in organisms from either the environment or from clinical sources, but not both. These databases curate diverse information, ranging from molecular information to host and habitat figures. Clinical and environmental databases could be integrated and better coordinated. A VFAR database currently under development (Jenkins *et al.* 2004) may help to fulfill this need. Smaller, organism-specific models for this type of enhanced database exist; the NSF and NIH have coordinated efforts to compile a database on the West Nile Virus, for example (Cangelosi & Freitag 2005).

Conclusions: prospects for using VFAR to discover new pathogens

The goal of using VFAR as a first-line tool to identify new pathogens is undeniably ambitious. There are many practical challenges. Many microbial virulence factors perform multiple functions, some of which have nothing to do with disease in humans. Such factors may be widespread in environmental organisms. Therefore, a significant background signal can be expected when virulence factors are sought in environmental samples, and even in pure cultures of microorganisms.

Fortunately, our understanding of microbial pathogenesis continues to expand. As known virulence factors increase in number, so too does the feasibility of a VFAR strategy. This remains a desirable goal and a strong rationale for continued research into the mechanisms of microbial pathogenesis. As our understanding of virulence mechanisms expands, databases that link virulence factors with specific environmental sites and culturable organisms will become increasingly useful.

ACKNOWLEDGEMENTS

The author is grateful to Makeda Semret and Anthony Newsome for their critical reviews of the manuscript. The writing of this article was supported by grant no. 833030010 from the US Environmental Protection Agency.

REFERENCES

Abdallah, A. M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R. J. P., Vandenbroucke-Grauls, C. M. J. E., Appelmelk, B. J., Luirink, J. & Bitter, W. 2006 A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol. Microbiol.* **62**, 667–679.

Board on Environmental Studies and Toxicology 2001 *Classifying Drinking Water Contaminants for Regulatory Consideration*. National Academies Press, Washington, DC.

Bono, M., Jemmi, T., Bernasconi, C., Burki, D., Telenti, A. & Bodmer, T. 1995 Genotypic characterization of *Mycobacterium avium* strains recovered from animals and their comparison to human strains. *Appl. Environ. Microbiol.* **61**(1), 371–373.

Brodin, P., Eiglmeier, K., Marmiesse, M., Billault, A., Garnier, T., Niemann, S., Cole, S. T. & Brosch, R. 2002 Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect. Immun.* **70**(10), 5568–5578.

Brosch, R., Pym, A. S., Gordon, S. V. & Cole, S. T. 2001 The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol.* **9**(9), 452–458.

Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van Soolingen, D. & Cole, S. T. 2002 A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl Acad. Sci. USA* **99**(6), 3684–3689.

Caccio, S., Spano, F. & Pozio, E. 2001 Large sequence variation at two microsatellite loci among zoonotic (genotype C) isolates of *Cryptosporidium parvum*. *Int. J. Parasitol.* **31**(10), 1082–1086.

Cangelosi, G. A. & Freitag, N. E. 2005 *Pathogens in the Environment. American Academy of Microbiology Critical Issues Colloquia*. ASM Press, Washington, DC.

Cangelosi, G. A., Clark-Curtiss, J. E., Behr, M., Bull, T. & Stinear, T. 2003 Biology of pathogenic mycobacteria in water. In *Pathogenic Mycobacteria in Water* (ed. J. Bartram & G. Rees), pp. 23–44. World Health Organization–US Environmental Protection Agency, Geneva.

Cangelosi, G. A., Do, J. S., Freeman, R., Bennett, J. G., Semret, M. & Behr, M. A. 2006 The two-component regulatory system mtrAB is required for morphotypic multidrug resistance in *Mycobacterium avium*. *Antimicrob. Agents Chemother.* **50**(2), 461–468.

Ching, W. T. & Meyer, R. D. 1987 *Legionella* infections. *Infect. Dis. Clin. North Am.* **1**(3), 595–614.

Chiang, S. L., Mekalanos, J. J. & Holden, D. W. 1999 *In vivo* genetic analysis of bacterial virulence. *Ann. Rev. Microbiol.* **53**(1), 129–154.

Cirillo, J. D., Falkow, S., Tompkins, L. S. & Bermudez, L. E. 1997 Interaction of *Mycobacterium avium* with environmental amoebae enhances virulence. *Infect. Immun.* **65**, 3759–3767.

Cohen, A. L., Oliver, J. D., DePaola, A., Feil, E. J. & Fidelma Boyd, E. 2007 Emergence of a virulent clade of *Vibrio vulnificus* and correlation with the presence of a 33-kilobase genomic island. *Appl. Environ. Microbiol.* **73**(17), 5553–5565.

Colwell, R. R., Huq, A., Islam, M. S., Aziz, K. M. A., Yunus, M., Khan, N. H., Mahmud, A., Sack, R. B., Nair, G. B., Chakraborty, J., Sack, D. A. & Russek-Cohen, E. 2003 Reduction of cholera in Bangladeshi villages by simple filtration. *Proc. Natl Acad. Sci. USA* **100**(3), 1051–1055.

Cosma, C. L., Sherman, D. R. & Ramakrishnan, L. 2003 The secret lives of pathogenic mycobacteria. *Ann. Rev. Microbiol.* **57**, 641–676.

Costerton, J. W., Stewart, P. S. & Greenberg, E. P. 1999 Bacterial biofilms: a common cause of persistent infections. *Science* **284**(5418), 1318–1322.

Danelishvili, L., Wu, M., Stang, B., Harriff, M., Cirillo, S., Cirillo, J., Bildfell, R., Arbogast, B. & Bermudez, L. E. 2007 Identification of *Mycobacterium avium* pathogenicity island important for macrophage and amoeba infection. *Proc. Natl Acad. Sci. USA* **104**(26), 11038–11043.

Donlan, R. M. 2001 Biofilm formation: a clinically relevant microbiological process. *Clin. Infect. Dis.* **33**, 1387–1392.

Dumontet, S., Krovacek, K., Baloda, S. B., Grottoli, R., Pasquale, V. & Vanucci, S. 1996 Ecological relationship between *Aeromonas* and *Vibrio* spp. and planktonic copepods in the coastal marine environment in Southern Italy. *Comp. Immunol. Microbiol. Infect. Dis.* **19**(3), 245–254.

Edelstein, P. H., Edelstein, M. A. C., Higa, F. & Falkow, S. 1999 Discovery of virulence genes of *Legionella pneumophila* by using signature tagged mutagenesis in a guinea pig pneumonia model. *Proc. Natl Acad. Sci. USA* **96**(14), 8190–8195.

Falkinham, J. O., III 1996 Molecular epidemiology techniques for the study of *Mycobacterium avium*-complex infection. *Mycobacterium avium-Complex Infection: Progress in Research and Treatment* (ed. J. A. Korvik & C. A. Benson). Marcel Dekker, New York.

Falkinham, J. O., III 1996 Epidemiology of infection by nontuberculous mycobacteria. *Clin. Microbiol. Rev.* **9**(2), 177–215.

Faruque, S. M., Albert, M. J. & Mekalanos, J. J. 1998 Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev. MMBR* **62**(4), 1301–1314.

Freeman, R., Geier, H., Weigel, K. M., Do, J., Ford, T. E. & Cangelosi, G. A. 2006 Roles for cell wall glycopeptidolipid in surface adherence and planktonic dispersal of *Mycobacterium avium*. *Appl. Environ. Microbiol.* **72**(12), 7554–7558.

Fukiya, S., Mizoguchi, H., Tobe, T. & Mori, H. 2004 Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J. Bacteriol.* **186**(12), 3911–3921.

Gal-Mor, O. & Segal, G. 2003 The *Legionella pneumophila* GacA homolog (LetA) is involved in the regulation of *icm* virulence genes and is required for intracellular multiplication in *Acanthamoeba castellanii*. *Microb. Pathog.* **34**(4), 187–194.

Hacker, J. & Carniel, E. 2001 Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. *EMBO Rep.* **2**(5), 376–381.

Harmsen, D., Dostal, S., Roth, A., Niemann, S., Rothganger, J., Sammeth, M., Albert, J., Frosch, M. & Richter, E. 2003 RIDOM: comprehensive and public sequence database for identification of *Mycobacterium* species. *BMC Infect. Dis.* **3**, 26.

Harriff, M. J., Wu, M., Kent, M. L. & Bermudez, L. E. 2008 Species of environmental mycobacteria differ in their abilities to grow in human, mouse, and carp macrophages and with regard to the presence of mycobacterial virulence genes, as observed by DNA microarray hybridization. *Appl. Environ. Microbiol.* **74**(1), 275–285.

Hou, J. Y., Graham, J. E. & Clark-Curtiss, J. E. 2002 *Mycobacterium avium* genes expressed during growth in human macrophages detected by selective capture of transcribed sequences (SCOTS). *Infect. Immun.* **70**(7), 3714–3726.

Inglis, N. F., Stevenson, K., Davies, R. C., Heaslip, D. G. & Sharp, J. M. 2001 Unique expression of a highly conserved mycobacterial gene in IS901 + *Mycobacterium avium*. *Microbiology* **147**(6), 1557–1564.

Izutsu, K., Kurokawa, K., Tashiro, K., Kuhara, S., Hayashi, T., Honda, T. & Iida, T. 2008 Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in Kanagawa phenomenon-positive *Vibrio parahaemolyticus* strains. *Infect. Immun.* **76**(3), 1016–1023.

Jenkins, T. M., Scott, T. M., Cole, J. R., Hashsham, S. A. & Rose, J. B. 2004 Assessment of virulence-factor activity relationships (VFARs) for waterborne diseases. *Water Sci. Technol.* **50**(1), 309–314.

Kaattari, I. M., Rhodes, M. W., Kaatari, S. L. & Shotts, E. B. 2006 The evolving story of *Mycobacterium tuberculosis* clade members detected in fish. *J. Fish Dis.* **29**, 509–520.

Kato-Maeda, M., Rhee, J. T., Gingeras, T. R., Salamon, H., Drenkow, J., Smittipat, N. & Small, P. M. 2001 Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* **11**(4), 547–554.

Krzywinska, E. & Schorey, J. S. 2003 Characterization of genetic differences between *Mycobacterium avium* subsp. *avium* strains of diverse virulence with a focus on the glycopeptidolipid biosynthesis cluster. *Vet. Microbiol.* **91**(2–3), 249–264.

Krzywinska, E., Bhatnagar, S., Sweet, L., Chatterjee, D. & Schorey, J. S. 2005 *Mycobacterium avium* 104 deleted of the methyltransferase D gene by allelic replacement lacks serotype-specific glycopeptidolipids and shows attenuated virulence in mice. *Mol. Microbiol.* **56**(5), 1262–1273.

Kunze, Z. M., Poraels, F. & McFadden, G. 1992 Biologically distinct subtypes of *Mycobacterium avium* differ in possession of insertion sequence IS901. *J. Clin. Microbiol.* **30**, 2366–2372.

Laurent, J. P., Faske, S. & Cangelosi, G. A. 2002 Characterization of IS999, an unstable genetic element in *Mycobacterium avium*. *Gene* **294**(1–2), 249–257.

Li, Y., Miltner, E., Wu, M., Petrofsky, M. & Bermudez, L. E. 2005 A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. *Cell. Microbiol.* **7**(4), 539–548.

Merrell, D. S., Hava, D. L. & Camilli, A. 2002 Identification of novel factors involved in colonization and acid tolerance of *Vibrio cholerae*. *Mol. Microbiol.* **43**(6), 1471–1491.

Miltner, E. C. & Bermudez, L. E. 2000 *Mycobacterium avium* grown in *Acanthamoeba castellanii* is protected from the effects of antimicrobials. *Antimicrob. Agents Chemother.* **44**(7), 1990–1994.

Morgan, U. M., Xiao, L., Fayer, R., Lal, A. A. & Thompson, R. C. A. 1999 Variation in *Cryptosporidium*: towards a taxonomic revision of the genus. *Int. J. Parasitol.* **29**(11), 1733–1751.

NRC 2002 *Classifying Drinking Water Contaminants for Regulatory Consideration*. National Academy Press, Washington, DC.

Ohkusu, K., Bermudez, L. E., Nash, K. A., MacGregor, R. R. & Inderlied, C. B. 2004 Differential virulence of *Mycobacterium avium* strains isolated from HIV-infected patients with disseminated *M. avium* complex disease. *J. Infect. Dis.* **190**, 1347–1354.

Pavlik, I., Svastova, P., Bartl, J., Dvorska, L. & Rychlik, I. 2000 Relationship between IS901 in the *Mycobacterium avium* complex strains isolated from birds, animals, humans, and the environment and virulence for poultry. *Clin. Diagn. Lab. Immunol.* **7**(2), 212–217.

Petrini, B. 2006 *Mycobacterium abscessus*: an emerging rapid-growing potential pathogen. *APMIS* **114**, 319–328.

Philalay, J. S., Palermo, C. O., Hauge, K. A., Rustad, T. R. & Cangelosi, G. A. 2004 Genes required for intrinsic multidrug resistance in *Mycobacterium avium*. *Antimicrob. Agents Chemother.* **48**(9), 3412–3418.

Pruzzo, C., Tarsi, R., Mar Lleo, M., Signoretto, C., Zampini, M., Pane, L., Colwell, R. R. & Canepari, P. 2003 Persistence of adhesive properties in *Vibrio cholerae* after long-term exposure to sea water. *Environ. Microbiol.* **5**(10), 850–858.

Rankin, S., Li, Z. & Isberg, R. R. 2002 Macrophage-induced genes of *Legionella pneumophila*: protection from reactive intermediates and solute imbalance during intracellular growth. *Infect. Immun.* **70**(7), 3637–3648.

Rhodes, M. W., Kator, H., McNabb, A., Deshayes, C., Reyrat, J. M., Brown-Elliott, B. A., Wallace, R., Jr., Trott, K. A., Parker, J. M., Lifland, B., Osterhout, G., Kaattari, I., Reece, K., Vogelbein, W. & Ottinger, C. A. 2005 *Mycobacterium pseudoshottsii* sp. nov., a slowly growing chromogenic species isolated from Chesapeake Bay striped bass (*Morone saxatilis*). *Int. J. Systemat. Evol. Microbiol.* **55**, 1139–1147.

Rollins, S. M., Peppercorn, A., Young, J. S., Drysdale, M., Baresch, A., Bikowski, M. V., Ashford, D. A., Quinn, C. P., Handfield, M., Hillman, J. D., Lyons, C. R., Koehler, T. M., Calderwood, S. B. & Ryan, E. T. 2008 Application of *in vivo* induced antigen technology (IVIAT) to *Bacillus anthracis*. *PLoS ONE* **3**, e1824.

Safi, H., Barnes, P. F., Lakey, D. L., Shams, H., Samten, B., Vankayalapati, R. & Howard, S. T. 2004 IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **52**(4), 999–1012.

Sassetti, C. M. & Rubin, E. J. 2003 Genetic requirements for mycobacterial survival during infection. *Proc. Natl Acad. Sci. USA* **100**(22), 12989–12994.

Schroeder, A. A., Lawrence, C. E. & Abrahamsen, M. S. 1999 Differential mRNA display cloning and characterization of a *Cryptosporidium parvum* gene expressed during intracellular development. *J. Parasitol.* **85**(2), 213–220.

Semret, M., Zhai, G., Mostowy, S., Cleto, C., Alexander, D., Cangelosi, G., Cousins, D., Collins, D. M., van Soolingen, D. & Behr, M. A. 2004 Extensive genomic polymorphism within *Mycobacterium avium*. *J. Bacteriol.* **186**(18), 6332–6334.

Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. & Musser, J. M. 1997 Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl Acad. Sci. USA* **94**(18), 9869–9874.

Steinert, M., Birkness, K., White, E., Fields, B. & Quinn, F. 1998 *Mycobacterium avium* bacilli grow saprozoically in coculture with *Acanthamoeba polyphaga* and survive within cyst walls. *Appl. Environ. Microbiol.* **64**(6), 2256–2261.

Striepen, B., Pruijssers, A. J. P., Huang, J., Li, C., Gubbels, M. J., Umejiego, N. N., Hedstrom, L. & Kissinger, J. C. 2004 Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl Acad. Sci. USA* **101**(9), 3154–3159.

Swanson, M. S. & Hammer, B. K. 2000 *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. *Ann. Rev. Microbiol.* **54**(1), 567–613.

Thompson, F. L., Thompson, C. C., Vicente, A. C. P., Theophilo, G. N. D., Hofer, E. & Swings, J. 2003 Genomic diversity of clinical and environmental *Vibrio cholerae* strains isolated in Brazil between 1991 and 2001 as revealed by fluorescent amplified fragment length polymorphism analysis. *J. Clin. Microbiol.* **41**(5), 1946–1950.

Tourlousse, D. M., Stedtfeld, R. D., Baushke, S. W., Wick, L. M. & Hashsham, S. A. 2007 Virulence factor activity relationships: challenges and development approaches. *Water Res.* **79**(3), 246–259.

Tsolaki, A. G., Hirsh, A. E., DeRiemer, K., Enciso, J. A., Wong, M. Z., Hannan, M., de la Salmoniere, Y. -O., Aman, K., Kato-Maeda, M. & Small, P. M. 2004 From the cover: functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl Acad. Sci. USA* **101**(14), 4865–4870.

Turenne, C. Y., Wallace, R., Jr. & Behr, M. A. 2007 *Mycobacterium avium* in the postgenomic era. *Clin. Microbiol. Rev.* **20**(2), 205–229.

Turenne, C. Y., Collins, D. M., Alexander, D. C. & Behr, M. A. 2008 *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* are independently evolved pathogenic clones of a much broader group of *M. avium* organisms. *J. Bacteriol.* **190**(7), 2479–2487.

Vital Brazil, J. M., Alves, R. M., Rivera, I. N. G., Rodrigues, D. P., Karaolis, D. K. R. & Campos, L. C. 2002 Prevalence of virulence-associated genes in clinical and environmental *Vibrio cholerae* strains isolated in Brazil between 1991 and 1999. *FEMS Microbiol. Lett.* **215**(1), 15–21.

Widmer, G. 2004 Population genetics of *Cryptosporidium parvum*. *Trends Parasitol.* **20**(1), 3–6.

Widmer, G., Tzipori, S., Fichtenbaum, C. J. & Griffiths, J. K. 1998 Genotypic and phenotypic characterization of *Cryptosporidium parvum* isolates from people with AIDS. *J. Infect. Dis.* **178**, 834–840.

Xiao, L., Morgan, U. M., Limor, J., Escalante, A., Arrowood, M., Shulaw, W., Thompson, R. C. A., Fayer, R. & Lal, A. A. 1999 Genetic diversity within *Cryptosporidium parvum* and related *Cryptosporidium* species. *Appl. Environ. Microbiol.* **65**(8), 3386–3391.

Xiao, L., Morgan, U. M., Fayer, R., Thompson, R. C. A. & Lal, A. A. 2000 *Cryptosporidium* systematics and implications for public health. *Parasitol. Today* **16**(7), 287–292.

Xiao, L., Sulaiman, I. M., Ryan, U. M., Zhou, L., Atwill, E. R., Tischler, M. L., Zhang, X., Fayer, R. & Lal, A. A. 2002 Host adaptation and host–parasite co-evolution in *Cryptosporidium*: implications for taxonomy and public health. *Int. J. Parasitol.* **32**(14), 1773–1785.