Issues in the Use of Genetic Technologies in Bioterrorism

Claire M. Fraser, Ph.D.
President and Director, The Institute for Genomic Research

DR. McCABE: Our next speaker is Dr. Claire Fraser. Her topic is issues in the use of genetic technologies in bioterrorism. Dr. Fraser is President and Director of TIGR, the Institute for Genomic Research.

DR. FRASER: Thank you very much. It's really a pleasure to be here. I was delighted to receive the invitation to speak with you today.

This was not my topic. This was one that was assigned to me, and I think it would perhaps be equally well-framed if the title of my presentation was issues in the use of genetic technologies and infectious disease, because I think that the challenges before us, the opportunities that come from having the availability of these new genetic technologies, are very much the same; and I think, as we've seen from the recent outbreak of SARS, that on balance the challenges we face from natural outbreaks of disease are hopefully going to be far greater -- not hopefully, but will likely be far greater than any that we will face in terms of any deliberate acts of bioterrorism.

But nonetheless, I've put together this talk with the bioterrorism aspect as a framework, and let me begin by just quickly reviewing what I think our current vulnerabilities are. This is, again, across the board in dealing with infectious agents in general.

With only a few exceptions, we really don't have adequate systems for rapidly and accurately detecting and recognizing a specific infectious agent, again whether it be deliberately released or a natural outbreak. We clearly have a fundamental lack of basic knowledge regarding the pathogenesis of most infectious disease agents, perhaps more so with biowarfare agents because the number of investigators that have studied these is much less. As a corollary, I would make the point that we also have a fundamental lack of basic knowledge of host response following exposure to infectious agents, and the two of these really go hand in hand. We have the tools now at our disposal to begin to think about tackling these issues.

We don't have adequate forensic methods for the purposes of attribution, and you'll hear certainly much more about this from Bruce Budowle in the next talk, whether we're talking about forensics having to do with human DNA or microbial DNA, as we're now starting to think about more that the issues are very much the same. I think there's good agreement that our arsenal of available vaccines, antimicrobials and antivirals could certainly stand to be beefed up.

Given that as a background, it's not surprising that when the microbial genomics efforts began in the mid-1990s, some of the most important first organisms to be tackled with these approaches were those that cause human disease, and I don't think it's any exaggeration to say now that essentially all of the major human pathogens have genome sequences available, and in many cases for some of these more important organisms we now have sequences available for multiple isolates. As I'll talk about a bit later, that's turned out to be extremely valuable.

The point of all of this was to hopefully use this information to accelerate the development of new and better diagnostics, drugs and vaccines for treating infectious disease. But specifically in terms of bioterror agents, several years ago there was a concerted effort among a number of federal funding

agencies to put these organisms on the list for genome sequence analysis. The smallpox genome sequence was completed some time ago, in the early 1990s, but since then we've now tackled essentially all of these agents under Category A and Category B. These are the lists put out from the CDC, and these tend to be the more esoteric infectious agents, ones that don't often come into typical conversation about infectious disease.

I think that Francis made the point that the more sequence we have, the more we seem to want, because the real power from having this information comes from the ability to do comparative genomics. In the microbial arena, some of the most interesting and most informative comparisons so far have come from looking at differences between species. Sometimes these are distantly related, other times more closely related. I think one of the big surprises in the microbial area to come from this analysis was the observation that the process of lateral gene transfer, exchange of DNA in the environment, seems to be playing a much bigger role in generating genetic diversity among microbial species and pathogens in particular -- this is an important area of this -- than we had previously appreciated.

So it says that these genomes are not static and it suggests mechanisms whereby new isolates with different properties, increased virulence, for example, antibiotic resistance may arise. Just to give you one very quick example of the kind of insights that you can get, at TIGR we've been working on a number of organisms in the Bacillus anthracis, Bacillus cereus, Bacillus thuringiensis group. This is a very closely related group of microorganisms. Anthracis you all know about. I don't need to tell you why that's so well known. Bacillus cereus is an organism found in the soil. It's an opportunistic pathogen in a small number of immunocompromised patients, but in most cases it just goes about its business in the soil. Bacillus thuringiensis is an insect pathogen. Those of you who put BT on your lawn to kill Japanese beetle grubs, BT comes from Bacillus thuringiensis. It's a toxin that interferes with insect feeding.

So we've been working on genome projects on this group of organisms, and I bring this up because what this information can provide when you take a comparative look at it is insights as to how some of these organisms have evolved. In particular, there was an issue of Nature just a couple of weeks ago with two reports on the complete genome sequence from Bacillus anthracis. This is work from TIGR by Tim Read; Bacillus cereus, work from a group that integrated genomics.

The take-home message here is that, as had been thought prior to having complete genome sequence information, these two organisms are extremely similar, especially if we're thinking about these as two different bacterial species. Their genome size is very much the same, just over 5 million base pairs. Many of the predicted proteins in one have best hits in the other, and vice versa, and we see a great deal of gene syntony. That is conservation of gene order in very large clusters when we do whole-genome comparisons.

This has confirmed that, in fact, these two organisms are very closely related and most likely shared a common ancestor. One of the questions is are there additional virulence factors in Bacillus anthracis that we don't know about? Can we get insights from taking a comparative approach? One approach that was taken -- and this is work that was done previously. There was a major transcriptional regulator. This is a master switch, if you will, in this group of organisms that, when activated, turns on a whole cascade of events, with the end result being increase in expression of a number of virulence factors. It's thought that this is a key trigger in activating the virulence pathways in thuringiensis, in cereus, and Bacillus anthracis.

One of the things that you can do when you have whole-genome sequence information is scan the genome for the known binding sites for this transcriptional regulator that occur upstream of genes. In doing this, we were able to identify a much larger number of these genes in both the cereus and the anthracis genome. Given what we know about this transcriptional regulator being important in virulence, we can

then assume as a new starting point for follow-on investigation that the genes that have been identified, the additional genes that are preceded by this binding motif here may also be involved in virulence.

When we put together a list of what those are, we see a much larger number of potential virulence factors than had previously been appreciated, and interestingly those on the chromosome are essentially shared entirely between Bacillus anthracis and its much less virulent relative, Bacillus cereus. So this gives us a way now to begin to focus in on a subset of genes and ask what role do these play in virulence and begin to help put the relationship between these organisms in better perspective.

Continuing with our anthracis work, we've also, I think, done a great deal in the past few years looking at differences between various isolates of Bacillus anthracis, hoping to better understand the known differences that have previously been described in phenotype. There are some isolates of Bacillus anthracis that are almost avirulent, and others, like the Ames strain that was sent through the U.S. mail, that are highly virulent. We'd like to understand what those differences are and ultimately use that information in the design of novel vaccines, novel drugs, et cetera.

It just so happens that in 2001, at the time of the anthrax letter attacks, TIGR was working on finishing the genome sequence of Bacillus anthracis, and we decided that we would ask the question could we use comparative genome analysis and leverage the genome data that already existed to try to get additional insight about the attack strain that had been sent through the U.S. mail. At the time of the attack, the VNTR analysis done in Paul Keim's lab at Northern Arizona University had shown that the bioterror isolate was the Ames strain, but there was not much that could be said beyond that.
This is a summary, if you will, of the history or what we think the history of the Ames isolate is. It originally came from a dead cow in Texas in 1981, was sent to Fort Dietrich, and from there it was distributed to labs around the world. We received our isolate for genome analysis from Horton Down in the U.K. To make a very long story short, what we did in terms of comparative genomics was get access to the isolate that came from the CSF of the first patient to die of inhalation anthrax in Florida, and by doing a high-draft sequence coverage of this Florida isolate, we compared it with the reference Ames strain that we were working on and, in fact, found a number of additional polymorphic loci between the two strains. These were then used by Paul Keim, who was a key collaborator on this project, to further extend his genotyping assays.

The outcome of this was that we were able, using this new information, to further distinguish a small number of Ames isolates, all shown here in blue, that had previously been indistinguishable based on existing genotyping information back in October of 2001. We were very encouraged by these results, and we have been funded now by the National Institute of Allergy and Infectious Diseases to take this a bit further and to look more broadly within the Bacillus anthracis project to develop new tools to analyze these closely related genomes to look at these clonal isolates, and at the same time to develop new methods for automated SNP discovery and to further explore B. anthracis as a model system for comparative genomics.

Shown here on this slide is a family tree, if you will, of Bacillus anthracis. This is also work from Paul Keim's lab, genotyping assays looking at many hundreds of isolates that had been collected around the world. It's suggested that there are two major groups of Bacillus anthracis, A and B, and within each of these groups there are a number of subgroups, and we have looked, at least we've started to look very broadly across this phylogenetic tree looking at isolates with different geographic distributions, with differences in phenotype.

This is still work in progress, but with the available sequencing capabilities that we have, we're able to generate data very, very rapidly. I think one of the most exciting and completely unexpected results from

this work is the discovery of an entirely new taxonomic group of Bacillus anthracis that Paul Keim has designated as Group C. This was originally isolated from Louisiana several years ago. There are two strains that exist of this new Group C. They contain one of the key plasmids that encode the important virulence genes but not the other.

You might say, well, so what? What's the importance of one more group of Bacillus anthracis isolates? I think it becomes important when you put this information in phylogenetic context. Here is the new Group C that has been identified. These are the existing Groups A and B that we are working on. It turns out that this Group C Bacillus anthracis looks to be much more closely related to Bacillus cereus, that we think may have been an ancestor of Bacillus anthracis. So it looks as if, as part of doing this work, we're beginning to perhaps get better insight as to where these more highly virulent isolates of Bacillus anthracis might have come from. So this is work in progress and we're going to be continuing with this over the next 18 months or so.

The ultimate goal is to develop a comprehensive database of Bacillus anthracis isolates, and I'm not going to say too much about databases. You'll hear much more about this from Bruce Budowle. We're taking a number of approaches, and I think it's important not to just limit this activity to Bacillus anthracis. I think that this is the kind of information that could be critically important for all major human pathogens, because if this kind of information existed for the family of corona viruses, for example, we would have been able to ask when the SARS outbreak first occurred where did this come from, where was this last seen geographically, what does this most closely resemble in terms of other corona virus isolates, and I think we would have been much further ahead than we were, although I have to say the speed at which the developments in tackling the SARS analysis have taken place have really been very laudable.

So the bottom line is that I think if we have a better understanding of DNA variation among various bacterial isolates and strains, it will be important for a number of reasons, both in terms of epidemiological studies and microbial forensics. It's not out of the question to think that if we have this information and we can do additional follow-up work, that we may be able to predict clinical outcomes based on having genotype information.

In the case of bioterrorism events, this kind of information would certainly increase our ability to rapidly detect genetically modified strains, and for a lot of work that's currently going on in terms of vaccine development, development of novel therapeutics, this is critically important information. You don't want to be developing a new vaccine against a protein that's not widely distributed among natural isolates of an infectious agent. So we really want to know what the variability is, and as we've seen already from some key examples, the variability can be quite profound. So it would be nice to have that information up front.

Bottom line is I think that over the next several years we'll begin to see this information applied in new ways. This is perhaps a bit of an exaggeration of where we are today in terms of our ability to make identifications, but I think with genome sequence information, microarray technology, proteomics technologies, et cetera, it's not out of the question to think that at some point in the future we will be able to much more rapidly collect this kind of information about a particular isolate, and that will give us a tremendous advantage over where we are currently.

Again, I remind you that this is not about bioterrorism. I think more importantly it's about anticipating, understanding natural outbreaks of disease. I think this Newsday headline is very, very topical, that the worst bioterrorist may be nature itself. We've seen too many examples of that recently. What could be added now to this slide of a number of emerging and reemerging diseases over the past 20 years or so is

West Nile, SARS, and perhaps even the most recent reports of the outbreak of Monkey pox in the Midwest.

So let me conclude, then, by just briefly talking about how microbial genomics approaches have had an impact already on our understanding and treatment of infectious disease. One of the real hopes in undertaking a lot of these studies initially was that this kind of information would greatly accelerate the discovery of new antimicrobials and would provide new targets for antimicrobial development over the current three pathways that are targeted with all existing antibiotics, those being DNA synthesis, protein synthesis, and cell wall biosynthesis.

Many academic investors, many biotech companies, many large pharma companies have certainly used genomics information to begin to identify new sets of targets, taking very different approaches, some of which are summarized on this slide. I heard Gail Cassell speak about a year ago. Gail used to be head of infectious disease research for Eli Lilly, and she said microbial genomics has more than delivered on its promise to identify new targets. In fact, the pharmaceutical company now has probably close to 200 new targets for follow-up. Not all of these make good targets, and the next step is, in a large number of new targets, identifying those which make the best targets.

But I find it ironic that in this era where there are potential new avenues of investigation for development of new antimicrobials, that a large number of pharmaceutical companies are shutting down their infectious disease research programs. This is extremely troubling, because I think there is very much a consensus that our current arsenal of antibiotics is not sufficient. With one exception, there hasn't been a new class of antibiotics developed in 30 years, and many of those that are out there, we see many, many examples, ever increasing, of antibiotic resistance. There have been too many reports over the past few years of Vancomycin-resistant Staph. aureus, Vancomycin being considered the last line of defense.

So I think that this is a critically important issue. I think that we collectively are poised to begin to try to exploit some of this information, but unfortunately the economics of the bottom line are just not making that a reality, and I find that very disturbing.

The other area where we've seen some notable successes using genomics data is in the area of vaccine research. Whether it be for infectious disease or bioterrorism, the goal is to develop new vaccines to protect all groups of civilians, and there are two needs here, to develop better vaccines against microbes for which vaccines currently exist -- I think the example with the smallpox vaccine is a great one -- and also the opportunity exists to develop new vaccines against pathogens for which none currently exist, and the list there is quite long.

This is just one example of how this kind of information can accelerate vaccine development. This is work that I know well because it came from TIGR, in collaboration with a group at Chiron Corporation, but there have been many reports published since this first appeared in March 2000 in Science using a similar approach to vaccine development based on having genome sequence information available.

In a very short period of time now, we can generate the complete sequence of a pathogen genome of interest. Back when we were doing this, three to twelve months. We can now say this is, depending upon the center you're looking at, three to twelve days or three to twelve hours, depending upon who is doing the work. In a very short period of time, this information can be mined using a number of available algorithms to look for potential new vaccine targets.

In the case of Neisseria meningitides here, 570 putative secreted or surface proteins were identified. These were all expressed recombinantly in E. coli, used to immunize mice. The sera were collected and

screened in a number of assays. In this very large-scale triage process, seven proteins emerged with very high titers. In all of the assays, and these were finally assessed for their sequence variability, something that pathogens seem to be able to do well is generate sequence variability in their cell surface proteins.

To make a long story short, two of these proteins out of this screening are now in Phase I clinical trials, and our Chiron colleagues have told us that having this information and taking this new approach accelerated this part of the vaccine development process by about three to four years, so that's very encouraging. But again, we're also finding ourselves in the situation with vaccine development of perhaps not having sufficient incentive for large pharma to develop vaccines.

Finally, with the availability of the human genome sequence, we can think about tackling infectious disease as well from the point of view of the host. I think some of the most exciting work that will come in this arena in the next few years will be further exploring both innate and adaptive immunity, and we should probably add to this slide host susceptibility to disease, because there certainly have been a number of reports already that have suggested that, in fact, there is individual variation. So when we're talking about SNPs and variation and susceptibility to disease, it's important to include susceptibility to infectious disease in that group as well.

This is work from David Relman's lab. I think this represents some of the most exciting potential application of human genome sequence information to infectious disease. What this is, is information from a microarray analysis of human lymphocytes following exposure to a number of pathogens or microbial components known to induce an immune response. It's too early to know whether we'll ever be able to generate signature expression patterns that will be able to say that an organism has been exposed to a specific agent, but it certainly looks from this information that we can say that an organism has been exposed to an infectious agent in general.

One of the hopes with this kind of approach and technology is that with the development of these kinds of assays, it may in fact be possible to get a readout like this before symptoms ever appear. So when we're thinking about the bioterrorism scenario, it would be wonderful to think about being able to screen large numbers of patients who were potentially exposed and get an answer before symptoms develop, because sometimes that can be too late.

Final slide here is that as we continue to make progress in the area of microbial genomics, and there has been a focus on biowarfare pathogens more recently, there have been some renewed discussions about whether some of this information should be kept out of the public domain, the fear being that somehow perhaps we may be providing information on new virulence genes, mechanisms of pathogenesis that may facilitate engineering of a superpathogen, if you will. There are a number of discussions ongoing. This is not a dead issue by any means.

My own opinion is that the more open these databases are, the better, because without this kind of information, we're not going to see the follow-up work taking place that we've already had a glimpse of already. But these are terribly complex issues. It's not so simple to say we need to have this information out there to accelerate research. I think the real difficulty comes in not being able to adequately assess the risk or the potential risk of this information for those who might want to use it for malicious intent. So this is a somewhat recent development but one that we're going to be following very closely.

Let me just conclude by acknowledging my colleagues at TIGR and a number of important outside collaborators who have played key roles in our microbial efforts over the past several years. Thanks.

(Applause.)