Issues in the Use of Genetic Technologies in Forensics

Bruce Budowle, Ph.D.
Forensic Science Laboratory, Federal Bureau of Investigation

DR. McCABE: We're going to go ahead and get started. Our next speaker is Bruce Budowle, who will speak on issues in the use of genetic technologies in forensics. Dr. Budowle is from the Forensic Science Laboratory, Federal Bureau of Investigation.

DR. BUDOWLE: Thank you, Ed, and thanks for inviting me to speak today. Just first a word from my sponsor. It's not a Webbie but it is an icon. I was asked to talk about genetic technologies and issues in forensics. I didn't plan to talk about microbial forensics. I thought of focusing on humans. But there is a large field of initiatives in microbial forensics on much of the issues that Claire talked about, and I'd stress the same things about databases. Many of the people with purse strings today think that sequencing one organism is all you need to do and that's sufficient for information. That would be a bad decision to make both for forensics and understanding diversity for the significance and weight of an observation, as well as for therapeutics, drugs, and so forth.

The other main issue to be concerned about is big discussions in security and dissemination of information, where there are some very diametrically opposed viewpoints on information. It gives bad people tools to make bad weapons, versus information helps good people make good tools to fight disease.

I thought I'd start off with the questions that were posed to me so my talk would address the questions that I think Sarah sent me on what are the technologies and the purposes, samples being collected, the kinds of people we're collecting samples from, expanding the breadth of individuals that we might be collecting samples from, are those samples available for other things, privacy issues, legal/social issues, and databases.

So first, we ought to understand what forensics is about and have a case, and I could have put up a violent rape case or whatever, but this makes the point just as well, where we have an individual that's been murdered obviously by a sharp object, and we have to now resolve who was the perpetrator of the crime by being able to identify one individual at the exclusion of all the others in this population here. We're going to use tools, genetic tools and other kinds of tools -- a fingerprint can be a tool, whatever it may be -- to identify who could have done it and who could not have been the perpetrators of the crime. That's our ultimate goal.

So forensic science is basically just science with legal matters. We're going to use science and technologies to be witnesses to help us resolve an event. It may not be the absolute identifier, it may not be the smoking gun. It's just a tool that can be used in addition with other evidence to help, and it plays a special role. But we like it because it is objective. We can make a test and we can interpret the evidence. But I wanted to stress that it's for identity purposes. It does not prove guilt or innocence for human identification. It only identifies an individual or a group of individuals that might be the source, as well as identifies those who may not be the source of the particular material found at the crime scene.

There can be situations where a sample is found and it has no bearing on a crime. For instance, you could have an estranged husband and wife, the wife kills the husband -- which would probably be justifiable in most cases -- and you find a small blood spot in the house. Well, the wife also resides in the house, and having a small blood spot in the house that matches her does not necessarily mean that she killed him or not. I should have said allegedly killed her husband.

There are many kinds of applications for human identity tests, and I've listed them here so I won't go through them. But realize that it's used more than just in criminal cases. It's used in paternity testing, historical investigations, evolutionary studies, for the military, mass disasters, mixed-up samples in hospitals. These are just some numbers to give you an idea, and these may be a little dated because it's probably a little higher now, but there's about 250,000 paternity tests done each year in the United States, about 20,000 criminal cases that are analyzed per year. So it's well ingrained and routinely done. Probably this field has embraced and applied molecular biology tools more so than any other field to date and has had success in resolving cases.

The advantage of using DNA over the prior systems, the protein systems, is that you can look at a large variety of tissues that might be found at a crime scene such as a hair sample. If we look in this room, we'll find hairs from a number of individuals that have been naturally shed, some of us unfortunately more so than others. Right, Ed?

(Laughter.)

DR. BUDOWLE: He wasn't listening, so --

DR. McCABE: I just have less to shed.

DR. BUDOWLE: That's one way of looking at it, yes.

(Laughter.)

DR. BUDOWLE: Makes you a better perpetrator.

DR. McCABE: Yes.

(Laughter.)

DR. BUDOWLE: A single hair could not be typed for the typical isoenzymes or serological proteins, but if there's any nuclear or mitochondrial DNA residing in that hair, it could be exploited and typed.

There are many methods that have been used, and the predominant methods of today are looking at microsatellites, repetitive elements, not a lot of information that can be gained from them on a personal level but only for identity purposes. Mitochondrial DNA for those very challenged samples such as old bones, teeth, a hair shaft; and then there's work also using Y markers for special kinds of cases, not for routine, such as a violent crime where there's very little male DNA in a large background of female DNA, which would be difficult to type by standard systems; then with some of the mass disasters and other issues arising, SNPs have been applied looking at specific sites in the nuclear DNA and mitochondrial DNA.

The kinds of evidence you see, I'm going to run through these quickly so you get an appreciation. It could be anything from stained clothing to bodies being recovered in Kosovo, and we all know the World Trade Center, predators on children. This is a case where DNA was used to identify three missing girls in Virginia, recently resolved. Using the data to solve older cases by database work; we'll talk about that in a minute. Vehicular manslaughter cases being tied up. People being exonerated who have been falsely associated with the evidence or with the case and having now the tools to be able to get better resolution. You can have a man being freed by implicating his brother because he has a similar DNA type.

One of the things that makes it valuable is that it's hard to get statements out of some witnesses. Witnesses don't want to testify. So when you can have scientific evidence, that can help resolve some of the information that's provided. Some witnesses aren't considered credible by some people. A prostitute who is testifying may not have the same credibility to a fact-finder that would be the President of the United States in another kind of DNA case, would not have the same credibility as maybe some other witnesses may have. So again, this has value in that sense.

Victims. This was a woman I talked to who actually was abducted from her home and raped, and then sent back and said if you say anything I'll kill you, and for many years she had this tremendous fear of even going outside, always looking around that somebody was going to attack her, because this person was still out there. Recently they actually identified him with the database, and that brings a different kind of security to the individual subsequently, even though it was years later.

Then there are the parents of victims who are still on vigil looking for the killer of their daughter, for instance. So you can see a wide range of applications and uses that affects society. Paternity testing, as I mentioned. Lineages; this is the lineage study of Thomas Jefferson and Sally Hemmings, an offspring. I might mention that it's the lineage of Thomas Jefferson. It doesn't prove it's Thomas Jefferson because it was a Y marker. It could be any male relative that would be in that lineage.

It can also be used with animals, tying things together in cases. This is a well-known case with Snowball the cat, where the perpetrator was identified, had a jacket, and there were white cat hairs on it, and the victim had a cat. They did some STR typing on the hairs, and the cat matched them up, and that was entered into court to help them assist in determining the course of the case.

It can also be used in poaching, smuggling, population management, all these applications, and also botanicals might be typed to help resolve a case. This was a well-known interesting case in the early '90s where a woman was found buried next to a Paloverde tree out in Arizona, and the suspect had a truck, and in the bay of his truck were Paloverde pods, seed pods. They did a rapid DNA test on it and they matched and went into court. They said it was unique because they typed 18 other trees and they were all different, so it could only have come from that tree. There was a little debate on that and may still be a debate on that. There was a number entered in, but it wasn't uniqueness.

Now that you have an appreciation for some of the ways it can be applied, we use databases to help solve cases by giving us investigative leads, and the idea of a database is to put in your information, to retrieve it to, as I said, get an investigative lead. So a crime is committed, you have no suspect. If you can go to a data bank of convicted felons with the idea of recidivism being that some people commit crimes over and over again, he may be in the database and you can catch him before he does it again.

The database is called CODIS for Combined DNA Index System, and it's a hierarchical database on three levels, from local to state to federal, and it has a number of files in it today, from convicted offenders files, people convicted of certain crimes based on the state's legislation and federal, and each state has slightly different legislation of what samples they can take or who they can take them from. There's a forensic file of case samples from unsolved cases. Now we have a missing persons file and a human remains, which are the same file; relatives of missing persons, which is a special file; a file for using population statistics, why we need databases of anonymous individuals; and I mentioned suspect files, although there is no legislation and the FBI does not have a suspect file of DNA profiles. Some states do have what we'll call shadow databases, and that might raise some concern on privacy, so I bring that up.

There are a number of databases planned around the world. Many countries have implemented them, quite different rules and regulations depending on them. The British, for instance, have probably the most open rule. Just about anybody walking with a tire iron in their hand can be taken a sample from, versus some others where it must be very serious crimes and only for a limited amount of time, to everything in between.

Although there are different policies from all the states and countries, there's sufficient common base on the genetic markers such that we can communicate from state to state. Within the United States, it's all the same. But from country to country if there's any kind of international incident that has to be done.

Now, of course, with success, the first thing that happens is you've got to expand the crimes that you're looking at. Initially it was violent crimes in the United States because we were interested in what we call quality of crime, the worst types of crimes, the ones that affect individuals, rape and sexual assault being the highest level. What's been found since then is that a lot of cases where they've been convicted of rape or murder, a good percentage of them actually were convicted of burglary or other things earlier. So it now expands to all sorts of felonies and maybe other kinds of crimes as well.

Now there are some states, Virginia and New York in particular, that are considering all arrested individuals. I don't know where they're going to get the money to do this, but that's going to be a large database. It can include indicted persons. Terrorism-related offenses are now being part of this system here. There have been studies through the National Institute of Justice, but it does bring up an issue of individuals' rights in society, and there's a real balance here of who should be tested and why. I mean, I personally don't want to be tested. Probably one of the first famous cases in DNA is two teenage girls are murdered, there was sexual assault. It was in England. They go around and collect up 5,000 people, and they do it voluntarily. But, of course, if you don't volunteer, then you're a suspect. So there is this peer pressure and it does create a problem that I hope does not become the norm for the United States. But that's an issue you have to think about when you're applying these.

Now, when you have the data, you're going to have all these samples, and I mentioned these things, but I'll mention the relatives of missing persons. Those are in there voluntarily for one purpose only, and that's to compare for when you find the remains of some putative missing person. Well, it is a missing person but putative to the family members. They cannot be searched against any other data files, and they can be removed upon request.

The interesting one is that you have unidentified human remains that are found because you just found remains, and currently they can be searched against the criminals' samples and the convicted offender files, and that becomes interesting because it could be the brother of somebody in there, as well as you could search a crime where the brother committed it but he's not in the database and look for partial homology because of the relatedness of the individuals, or a parent or whatever, so you have to think about these things.

I mentioned the suspects issue, and the last I'll mention is the DNA samples. Currently, we collect the blood or a buccal swab, extract the DNA or maintain the sample, and that's stored. The reason it's stored is for really good reasons. It's a quality control factor. Let me take a step back. In forensics, when you make a case and you're going to err, you'd like to err on the side of a false exclusion as opposed to a false inclusion, because you'd rather let a guilty man go free than convict anybody who is innocent.

In a database, you want to err the other way. You'd rather have a false inclusion than a false exclusion, because if you falsely exclude, you don't make the match and you don't solve the crime, you don't get that person off the street. If you falsely include them, since you have the original samples remaining in the data set, you can go back and confirm it and resolve those kinds of mismatches. That's a real value in the system.

The second is that as technologies change, although we don't foresee this technology changing for a long time for a lot of other reasons, stability being one of them, and cost being another, you would want to go back and re-do those samples so that as you're typing new cases you have the same individuals typed for the current set of genetic markers. If you throw away the samples, you may never get them again. So there's some really legitimate reason for it, but using it for any other purpose is a serious concern, and I can talk about that as you want.

So questions you might ask are can the suspects in that shadow suspect database who have been exonerated, because if they'd been convicted they're in another database now, can they be searched against other cases on a future date, whether that's a legitimate thing to do? I mentioned some of these other ones.

Because of the tools now, we can look back at older cases, and because of the lag time because of resources to analyze cases, some crimes fall under statute of limitations. So now there are some very clever ways that have been constructed -- I don't want to say contrived because that sounds nefarious -- constructed dealing with an unknown perpetrator, and I think this first case was in Wisconsin of creating a John Doe warrant. In other words, I don't have a name but I do have a profile that represents a person, and that now no longer allows you to be stuck under that statute of limitations. So you have to think about other issues that arise when you have the tools to analyze things.

You could also extend the statute of limitations, you could eliminate the statute of limitations, and various jurisdictions are considering these.

Now, for privacy concerns, there are issues that have been addressed about limited access based on the law and for identification purposes only. So this is not that someone can go out and say I'd like to get a bone marrow repository enhanced by all these convicted felons. That's not the purpose. But I do think one or two states do have this humanitarian clause, but most of them do not.

So what is the federal law? The federal law dates back to 1994. It authorized the National DNA Index and put certain requirements in there for this index, and participation, who can actually create a file and who can access those files. So the records are of persons convicted of crimes, recovered from crime scenes, the unidentified human remains I mentioned and so forth, and the people for the index shall only include information on quality assurance and proficiency testing. There is no other kind of identifiers in this information. It's all coded, so when you look into it you'd have to go back to the state itself or, if it's a federal prisoner, to the government to get that. So there would be limited disclosure of information.

Now, the limited access for these data by law is for law enforcement identification purposes, judicial proceedings, and for criminal defense purposes so a defendant can have access to the samples and analyses performed in connection with his or her case. That one I think is a very ill-defined one, and I don't know how that's going to play out. When you say samples involved in the case, a database search could be 1.2 million individuals. Do you get access to all that? Do you get access to just the fact that you got a hit? I think those are still arguments being played out in the court today.

Now, access to the database is described, and also the cancellation if you don't meet the QC for quality assurance and the privacy requirements. In other words, if someone fails, like we've heard in the newspapers that some crime labs were not up to snuff, they would be taken out and a lot of their data may be expunged if it didn't meet the requirement.

What's imposed on the states is that to participate in CODIS or the national index system, you have to follow these requirements if you participate in CODIS or if you receive federal funding. So that's the lock that gets everybody into it. I'll leave there here because time-wise you can get these all yourselves. But state law, I wanted to bring up a couple of points. This is a little dated data. It may be a little different now, but about a year or two ago about 60 percent of the states did have provisions that penalized people for unauthorized disclosure of DNA information or giving up the samples. Most of it is at a misdemeanor level. Most states can have a felony level for tampering with the evidence, though, and I think that's more because those are traditional kinds of laws, not so much because of the database.

To touch on disease associations, I think I'm with a group of people who understand this, but there is a proposition that we have to be concerned about, that the markers that we use, although they don't encode themselves for proteins, may be associated with disease and provide some information about the person beyond just identity. There are some repeat sequences such as certain expansion repeats that are associated with diseases that are well known. But for the markers we have, they don't seem to have that kind of an association or that kind of functional effect on that. So the relative risk is rather low to provide information to date with the information we have on an association.

Not to confuse cause and effect with association. There's a big difference here between those. So I'll give one example that's known, and there's another one too, but this is just fine. If you look through one of the repeat sequence markers called THO1 that resides in an intron in the tyrosine hydroxylase gene, people have been interested in this gene because it's in the catecholamine pathway, so it might be associated with certain neuropsychiatric diseases, bipolar disease, schizophrenia. So there have been some studies done looking at populations of healthy and, let's say, people with schizophrenia, two studies actually, French and Tunisian, and they found that a particular allele which is relatively rare to have a higher risk in the individuals who have schizophrenia than in those who do not.

When you went back, there was no family history associated with it. Other studies done show no association with the 10 but maybe another allele showing up in that, and that's probably one of the problems I was talking with Francis during the break about the half million people, that I think you need to up the number, because a lot of these kinds of confusing studies, we have very heterogeneous populations, and you need to define them better, make them more homogeneous through a lot of admixture studies and things, so that you can get some of this haziness out of the way so you can find out what's real or not. But in the end, the relative risk would increase from one-half to 1 percent to 1 to 2 percent, hardly enough to be worried about. If I had a 1 to 2 percent risk, I certainly wouldn't be worried about taking any kind of drug therapy at this point in time.

So in conclusion, given how the technologies can be used, given the idea of some of the samples, the idea of expansion, there is one thing that a couple of us voted against, but we lost. Eric will tell you. They can use it for research purposes for forensic tools, developing other tests. I would argue it doesn't make a lot of sense, but that's in there. Privacy issues still abound, but there's a lot of laws in place in many states to give punishment if people use it beyond what it's supposed to be or tamper with it. The databases are not a problem. We all use the same genetic systems throughout the United States. We all use the same software and communication, and there's the same quality assurance standards are applied across the board.

Then there are some legal/social issues. The social ones for privacy as far as knowing something about the individual, probably very low for these markers. Other markers as we go down the line may present a problem. There are a lot of people who want to switch to SNPs because it's the designer marker of the day. There are technologies advancing to it. It's not trivial. To gain the amount of information that we get from the current battery of core loci, we would probably need around 70 or 80 SNPs, and that doesn't get into the issues of mixtures and other kinds of family relationship kinds of tests that would confuse that. That's not trivial to make a robust technology for 70 to 80 SNPs. While we have lots of chips and things, they're not robust enough today to be able to make it into the forensic arena, where we really demand a higher quality, reliable system.

So it's not going to happen quickly, but in some instances, like the World Trade Center where there are these victims, we've gotten down to the dregs of materials, highly degraded, SNPs could make sense. If you take 60 or 70 SNPs, you run out of real estate for unlinked loci, biologically speaking. So you have to keep that in mind, that eventually you're going to hit some marker somewhere that's close to some other gene, but maybe in that particular situation the humanitarian effort of identifying an individual might outweigh the risk, and that's something you have to think about.

So with that, I'm done. Thank you.

(Applause.)

DR. McCABE: Thank you, Bruce.