In this inaugural interview for Forensic Genomics, David Mittelman, Editor-in-Chief, talks to Dr. Bruce Budowle, a well-known member of the forensic science community. After more than 25 years of working for the Federal Bureau of Investigation, Dr. Budowle is currently the director of the Center for Human Identification and a professor in the Department of Microbiology, Immunology and Genetics, at The University of North Texas Health Science Center in Fort Worth, Texas.

Dr. David Mittelman: Thank you, Dr. Budowle, for agreeing to do this interview for our new journal, Forensic Genomics. Tell us about the organization you run and what it does?

Dr. Bruce Budowle: The Center for Human Identification is a center of excellence that combines different aspects. One aspect is service in the sense of providing genomics, forensic genomics, and forensic anthropology analyses to help identify missing persons, unidentified persons, and to help solve a variety of investigations of that nature. In addition, we do traditional case work such as sexual assaults and property crimes. Our laboratory utilizes Combined DNA Index System (CODIS) to enable uploads of DNA data and subsequent database searches. This association with the Federal Bureau of Investigation (FBI)’s managed CODIS makes our center rather unique—we are the only nontraditional government laboratory to have such access.

We also manage NamUs, which is supported by the Department of Justice through its National Institute of Justice (NIJ). Although the DNA testing is performed in our operational Missing Persons Unit, NamUs is an additional part of the intricate system to identify missing persons. In part, NamUs is a database, like a clearing house, but also provides a number of services to help identify people. With NamUs, the public and the government work together as force multipliers to help identify missing people and unidentified human remains.

The Center for Human Identification also has a research laboratory in the forensics genomics area—developing new technology and other applications to enhance capabilities to analyze challenging forensic evidence. Since we are a part of an academic institution, we take on and mentor graduate students, teach courses, and train professionals. This allows us to provide a wide spectrum of functions and opportunities.

Dr. Mittelman: Are these DNA testing services you provide offered to everyone across the United States, not just in Texas?

Dr. Budowle: The missing persons work is provided to agencies across the country, with funding from the Department of Justice and its NIJ. We serve Texas and run the missing person database for Texas (with support from the state). The missing persons services provided to agencies across the nation are at no cost to those agencies—these services will continue as long as that funding holds. In fact, more than half of the unidentified remains that are in CODIS (that is National DNA Index System, NDIS) today were typed by our laboratory.

Dr. Mittelman: Wow, half of the unidentified remains in CODIS came from University of North Texas Center for Human Identification?

Dr. Budowle: The work came from all over the country.

Dr. Mittelman: The testing was performed at your center?
Dr. Budowle: Correct. For the other forensic genetics work we do, such as sexual assault casework, it is only for the state of Texas. Most states have their own crime laboratory systems that handle the more traditional types of casework.

Dr. Mittelman: For unidentified remains, you have markers that you can upload to CODIS. But is there anything else done for the missing persons? Do family members contribute their DNA to CODIS?

Dr. Budowle: The sheer volume of missing and unidentified person cases poses major challenges to local and state agencies and is a tragedy to the countless families and communities who have suffered lost loved ones. In addition to the trauma and workload demands, missing and unidentified person investigations are difficult—if not impossible—to resolve without a central repository of relevant case information.

NamUs fills that gap. NamUs doesn’t do any DNA typing itself or house any DNA data. It is a database that brings investigative information, such as the reported missing loved one and maybe their personal items they had and maybe some other kind of metadata. The DNA testing is done separately in our missing persons unit that is a part of the center and those results generated from the remains will first be uploaded into our local database, then up to the state level database, and lastly up to the national database (NDIS).

We upload the DNA profiles of the remains. Most of these remains cannot be identified unless there are reference DNA profiles to compare them with. These reference samples are family members that are necessary to generate viable associations to assist in the identification process. It is a relationship testing, a kinship or familiar relationship testing, whatever terms you want to use. It is necessary to get participation from family members around the country through a variety of mechanisms.

The DNA profiles are used to generate a pedigree or a family tree, in our cases mostly first-degree relatives or close relatives. Then the DNA profiles of the human remains are compared with these pedigrees to determine whether the evidence profile would be supported genetically to be that missing person of a particular family or whether there is more support to be unrelated to that pedigree.

Dr. Mittelman: Are the people who decide that they want to submit a pedigree or genetic material to establish that first-degree relationship depositing information into an internal data storage that you use to search or is it also going to CODIS?

Dr. Budowle: We use the CODIS system for the human remains identification work in the United States. We type family reference samples and we type more than half of all the family reference samples as well.

Dr. Mittelman: So, all of those profiles end up in CODIS. Now if family members have contributed their profile to CODIS and one of their missing loved ones turns up at your center as an unidentified person, then the missing persons unit does an analysis of the DNA markers and through a CODIS search should be able to hit the family member?

Dr. Budowle: Yes, if there is a hit to be obtained. There are limitations that must be considered. First, not all missing persons or human remains are in the database, and not all families with missing loved ones are in the database. We have this open data set that limits our ability to make associations. Ideally, every family that has a missing loved one should be in the database so we would have the best searching capabilities. Thus, we can search and not obtain a hit. Indeed, we have more than 7000 remains in the CODIS database that have yet to be identified. Right now, those reference samples in the database are insufficient or not representative of those people who are the human remains.

Mitelman: How many total unidentified remains do you think there are in the United States, not just documented in NamUs but in general?

Dr. Budowle: On average, between 80,000 and 90,000 missing person cases are active in the National Crime Information Center (NCIC), which is likely an underestimate. The number of unidentified decedents is underrepresented in the NCIC. At the close of 2019, there were only 7987 active unidentified person entries in the NCIC. NamUs contains 13,379 open unidentified person cases. This disparity of cases is due largely to the fact that most medical examiner/coroner offices—the agencies primarily tasked with identifying remains—do not have access to the NCIC. In addition to existing cases, 4400 new unidentified decedents are recovered across the country each year, with ~1000 of these cases remaining unidentified and becoming cold cases.

It is a huge number of unidentified decedents—we use the term “a mass disaster over time.” I was involved in work on the World Trade Center (WTC). It was a tragic event that none of us will forget—people dying, explosive, terror, chaos, and destruction. Everyone focused on it and a little less than 3000 people died tragically. I do not want to diminish the significance of this terrorist attack as it was tragic, but comparatively there are more unidentified human remains than the number of people who died in the WTC attack. But because these missing persons are not associated with a dynamic event, they are sort of lost in the background and thus do not get the same attention that they and their families deserve.
**Dr. Mittelman:** As your center has had such a big role in bringing in unidentified remains and relatives of missing persons into CODIS, what do you see happening in the next 5 years to make more of a difference for this mass disaster that you are talking about?

**Dr. Budowle:** There already are technologies here or emerging depending on different sides of viewpoints, such as the genetic genealogy systems that actually have been around for a while, although there are different degrees of genetic genealogy to consider.

First, there is familial searching, in which a DNA profile from a crime scene sample is searched in CODIS, and no hit is obtained. However, there are individuals with similar DNA profiles in CODIS but not exact. Those that are similar could be relatives of the donor of the crime scene evidence. The candidates can be ranked from the highest likelihood to the lowest based on the genetic data and then take the top hundred or so are subsequently subjected to Y-short tandem repeat (STR) typing.

This lineage marker filters out adventitious hits of first-degree relatives. Y-STR typing is used because the vast majority of violent crimes are by male perpetrators. So, it makes good sense to go that route and it is easier to do Y-STR typing than mitochondrial DNA (mtDNA) sequencing, which could cover other kinship relationships. Only those very few (typically one or two per candidate set) individuals sharing the same Y lineage would then be pursued further, as there is a high probability, they are male relatives of the true source of the evidence. Now we are seeing in some places in Europe with drag-nets and also in China there is a large database using Y-STRs and the DNA Doe project relying on Y-STRs as lineage markers to identify family lineages, and again, the paternal lineage. Familial searching is good for identifying potential first-degree relatives such as sib—sib and parent—offspring. A Y-STR—based search can extend the potential to identify relatives two to four generations distant. With this approach as the degree of distance increases, the mutation rate of the STRs increases the uncertainty where eventually a distant relative may appear no different than an unrelated individual.

The third genetic genealogy approach is using high-density single nucleotide polymorphism (SNP) data either generated with SNP arrays in the range of 600 thousand to 1 million to 4 million SNPs or whole genome sequencing that in theory could yield all the SNPs within an individual that could be upward to 10 million. Obviously with so much more genetic data, one can infer a relationship with a higher degree of accuracy or reliability and extend to more distant relatives—six degrees and beyond.

Depending on the forensic laboratory right now with regular CODIS searches, the hit rate is ~30%–50%, that of the United Kingdom is ~60% because they have a high per capita entry of profiles in their database. Consider a 30%–50% hit rate actually means there is a 70%–50% failure to hit rate—those who deposited the crime scene samples that are not being identified through CODIS. There are upward of a million crime scene profiles residing in CODIS that have not been identified. That number would suggest an upward bound of a million cases not supported by a CODIS search. These are unsolved cases in CODIS that may be assisted with one of the genetic genealogy tools that we just discussed.

**Dr. Mittelman:** Will your center in the future take a lead role in trying to explore these more advanced forms of relationship testing?

**Dr. Budowle:** Absolutely, not just for the things we have talked about. We have cases right now that elude many people because they do not think about the limitations of current kinship testing. As an example, a skull was found in a county (the jurisdiction is not relevant); we were able to get mtDNA out of the sample, but there were no STR data. We searched our database and there is a woman who has a missing brother and was the only family reference sample that matched the mtDNA sequence of the skull. But that association was not strong.

About a year or two later, remains were found in the same county but they were the lower half of the body, so possibly the lower half of the body and the skull could be from the same individual. Those remains yielded an STR profile and an mtDNA sequence. The mtDNA sequence was the same as the skull and matching of course the sister. But the autosomal STRs suggested that the woman and the remains were not related. The combined data were uninformative. So, what does that conclusion mean?

Well, in relationships of siblings on average 50% of alleles are shared by descent, although the average appears a bit higher due to alleles shared by state. Although 50% of alleles are shared on average, the distribution of shared alleles between a population of sib pairs could range from no alleles in common (a low probability) to all the alleles in common (another low probability), when a limited number of markers are used as is done in most forensic work—between 13 and more than 20 STRs markers. So, stochastically siblings could have less than 50% sharing and even much less than 50%. In those biological first-degree relatives (sibs) may be uninformative or the data actually would support an exculpatory conclusion. Yet they are truly related. So, we are missing (no pun intended) identifying those individuals. We need to address this shortcoming.

In addition, some people do not realize what their true family relationships are and they are not full siblings but instead half siblings. We are going to miss those because the genetic data are not strong enough to make those inferences once you get past that first-degree relative. Our recent article in PLoS Genetics addresses this uncertainty.
and the false positive/false negative rates with the number of paternity/kinship tests that are done—millions over the past decade.

Given the migration issues going on at our border, where people are being genetically tested to confirm the biological relationship that, for example, this man says this is my child, depending on the likelihood ratio threshold that some laboratories actually invoke to make interpretations, there is a reasonable false-negative rate. The false-positive rate is small but still occurs when we consider millions of cases done. These numbers become notable. We need more genetic power to reduce that false-positive and especially the false-negative rates. If you think about the whole migrant situation, not just the United States but also around the world, someone is claiming that is my child and they happen to fall into that percentage of false negatives, the consequences to that person can be substantial. Reducing uncertainty when needed would be helpful.

Dr. Mittelman: By more powerful tests, you are stating using inference from more markers than what you might get from a traditional forensic test?

Dr. Budowle: Exactly.

Dr. Mittelman: We talked about how you probably deposited 7000 unidentified profiles to CODIS and if you look at NamUs, it shows ~13,000 unidentified person records. There are at least 100,000 unidentified persons in the United States and obviously some of them are found and solved but many of them are not. What your thoughts are on the reason for this difference?

Dr. Budowle: First, there is a resource issue. There needs to be a center of excellence, such as we are, to do this work, and the center needs consistent funding. Right now, we operate hand to mouth where we obtain funding year to year from the Department of Justice. The funding fluctuates and is not stable. We cannot maintain a good operation/service under this model.

Second, to handle the country’s needs we need a lot more funding to grow and maintain sustainability to support the work demands. It will be more cost-effective to have the work done in a central location for the country because replicating this laboratory in every state would be an exorbitant cost that does not make any sense. I am a big fan, not because my laboratory does such work per se but because it makes more economic sense for the taxpayers and concentrates expertise.

Another issue for the gap in unidentified remains not being analyzed is that many investigators are actually unaware that this service or even tools exist. We need to provide more training to the investigators even in the more rural entities and there needs to be some resources to get that information out to the people who are most likely to be able to identify remains.

Dr. Mittelman: It sounds like you are saying it is a combination of having the funding to be able to do enough testing and also getting the word out, educating people so they know what the options are for testing?

Dr. Budowle: Right. Of course, everyone says they need funding. We need funding to do this work. So that is a common thread, but it is more about sustainability, which is absolutely necessary. If you are going to do something right, you have to be sustainable, otherwise employees are terminated. There is a cascade effect because it is not just the cost of a person. That expertise is lost. Hiring new people if and when funding is made available requires substantial training (of a year or more) to qualify the individuals as experts. We cannot move employees out as one might do cashiers at Walmart. We cannot do the same thing with forensics genetics work.

Dr. Mittelman: Have you ever been in a situation wherein you must turn away work? Can you take any number of cases or do you plan for how many cases in a year you can accept, based on funding?

Dr. Budowle: Basically, it is the latter. We try to do as much as we can and that of course creates a backlog. Even with the backlog, we actually exceed the expected cases. We do not have enough resources to actually meet the demand, but we are going to do as much as we can.

Dr. Budowle has testified for more than 300 criminal cases and has established the framework for current forensic DNA and statistical analysis methods for human identification. In addition to his work on cold cases, Dr. Budowle has also recently started working on the use of forensic genetics to counter terrorism especially microbial DNA analysis and bioterrorism.