Forensics, DNA Fingerprinting, and CODIS

By: Karen Norrgard, Ph.D. (*Write Science Right*) © 2008 Nature Education
Citation: Norrgard, K. (2008) Forensics, DNA fingerprinting, and
CODIS. *Nature Education* **1(1):35**

How ethical is it to keep a database of convicted felons' DNA profiles? Can we rely on DNA fingerprints for conviction? Many ethical issues surround the use of DNA in forensic technology.

DNA is present in nearly every cell of our bodies, and we leave cells behind everywhere we go without even realizing it. Flakes of skin, drops of blood, hair, and saliva all contain DNA that can be used to identify us. In fact, the study of forensics, commonly used by police departments and prosecutors around the world, frequently relies upon these small bits of shed DNA to link criminals to the crimes they commit. This fascinating science is often portrayed on popular television shows as a simple, exact, and infallible method of finding a perpetrator and bringing him or her to justice. In truth, however, teasing out a DNA fingerprint and determining the likelihood of a match between a suspect and a crime scene is a complicated process that relies upon probability to a greater extent than most people realize. Government-administered DNA databases, such as the Combined DNA Index System (CODIS), do help speed the process, but they also bring to light complex ethical issues involving the rights of victims and suspects alike. Thus, understanding the ways in which DNA evidence is obtained and analyzed, what this evidence can tell investigators, and how this evidence is used within the legal system is critical to appreciating the true ethical and legal impact of forensic genetics.

How Does DNA Identification Work?

Although the overwhelming majority of the human genome is identical across all individuals, there are regions of variation. This variation can occur anywhere in the genome, including areas that are not known to code for proteins. Investigation into these noncoding regions reveals repeated units of DNA that vary in length among individuals. Scientists have found that one particular type of repeat, known as a short tandem repeat (STR), is relatively easily measured and compared between different individuals. In fact, the Federal Bureau of Investigation (FBI) has identified 13 core STR loci that are now routinely used in the identification of individuals in the United States, and Interpol has identified 10 standard loci for the United Kingdom and Europe. Nine STR loci have also been identified for Indian populations.

As its name implies, an STR contains repeating units of a short (typically three- to four-nucleotide) DNA sequence. The number of repeats within an STR is referred to as an allele. For instance, the STR known as D7S820, found on chromosome 7, contains between 5 and 16 repeats of GATA. Therefore, there are 12 different alleles possible for the D7S820 STR. An individual with D7S820 alleles 10 and 15, for example, would have inherited a copy of D7S820 with 10 GATA repeats from one parent, and a copy of D7S820 with 15 GATA repeats from his or her other parent. Because there 12 different alleles for this STR, there are therefore 78 different possible genotypes, or pairs of alleles. Specifically, there are 12 homozygotes, in

which the same allele is received from each parent, as well as 66 heterozygotes, in which the two alleles are different.

The Statistical Strength of a 13-STR Profile

Within the U.S., the 13-STR profile is a widely used means of identification, and this technology is now routinely employed to identify human remains, to establish or exclude paternity, or to match a suspect to a crime scene sample.

In order to utilize STR information as a means of human identification, the FBI established the frequency with which each allele of each of the 13 core STRs naturally occurs in people of different ethnic backgrounds. To this end, the FBI analyzed DNA samples from hundreds of unrelated Caucasian, African American, Hispanic, and Asian individuals. Assuming that all 13 STRs follow the principle of independent assortment (and they should, as they are scattered widely across the genome) and that the population randomly mates, a statistical calculation based upon the FBI-determined STR allele frequencies reveals that the probability of two unrelated Caucasians having identical STR profiles, or so-called "DNA fingerprints," is approximately 1 in 575 trillion (Reilly, 2001).

This very small number needs to be put into perspective. Note that this figure refers to pairs of people, and there are many pairs of people in the world. Indeed, for the 100 million Caucasians in the world, there are 5,000 trillion pairs of people, so roughly eight or nine pairs would be expected to match at the 13 STR loci. This predicted matching does not specify which profile is shared by two people, and the chance that anyone matches the particular profile associated with a crime is still very small. The distinction between two people sharing a profile and one person having a particular profile is an example of the so-called "birthday problem." Here, the probability that a person has a particular birthday is 1 in 365, ignoring February 29, but there is a 50% chance that two people in a random group of 23 people have the same unspecified birthday (Weir, 2007).

DNA Extraction and Analysis

To perform a forensic DNA analysis, DNA is first extracted from a sample. Just one nanogram of DNA is usually a sufficient quantity to provide good data. The region containing each STR is then PCR amplified and resolved according to size, giving an overall profile of STR sizes (alleles). The 13 core STRs vary in length from 100 to 300 bases, allowing even partially degraded DNA samples to be successfully analyzed. The costs of analysis, both in time and reagents, are significantly reduced by the amplification of all 13 STRs in just two multiplex PCR reactions.

Depending on the complexity of the repeat unit, the different alleles of an STR can vary by as little as a single nucleotide. For instance, the aforementioned D7S820 STR is relatively simple and contains between 5 and 16 repeats of GATA. Another STR, D21S11, has a more complex repeat pattern consisting of a mixture of tetra- and trinucleotides, and it therefore has alleles that vary in size by a single base pair. Because of the need to differentiate single-base differences, PCR products are typically resolved using automated DNA sequencing technologies with software that recognizes allele patterns by comparison to a known "ladder."

Making an STR Match

In order to match, for example, crime scene evidence to a suspect, a lab would determine the allele profile of the 13 core STRs for both the evidence sample and the suspect's sample. If the STR alleles do not match between the two samples, the individual would be excluded as the source of the crime scene evidence. However, if the two samples have matching alleles at all 13 STRs, a statistical calculation would be made to determine the frequency with which this genotype is observed in the population. Such a probability calculation takes into account the frequency with which each STR allele occurs in the individual's ethnic group. Given the population frequency of each STR allele, a simple Hardy–Weinberg calculation gives the frequency of the observed genotype for each STR. Multiplying together the frequencies of the individual STR genotypes then gives the overall profile frequency.

Table 1: Example DNA Profiles Showing the STR Alleles for Each Sample and the Genotype Frequency of Suspect B for Each STR Locus

STR Locus	Evidence Sample	Suspect A	Suspect B	Suspect B's Genotype Frequency for Each STR
D3S1358	15, 17	17, 17	15, 17	0.13
vWA	15, 16	18, 19	15, 16	0.22
FGA	23, 27	21, 23	23, 27	0.31
D8S1179	12, 13	14, 15	12, 13	0.34
D21S11	28, 30	27, 30.2	28, 30	0.06
D18S51	12, 18	14, 18	12, 18	0.11
D5S818	13, 13	9, 12	13, 13	0.29
D13S317	12, 12	12, 12	12, 12	0.21
D7S820	10, 11	9, 10	10, 11	0.26
CSF1PO	8, 11	11, 12	8, 11	0.18
TPOX	7, 8	8, 8	7, 8	0.30
THO1	9.3, 9.3	6, 9.3	9.3, 9.3	0.38
D16S539	9, 13	11, 12	9, 13	0.10

In the fictional case shown in Table 1, Suspect A is excluded as the source of the crime scene sample. Suspect B, on the other hand, matches the crime scene sample at all 13 STRs. A calculation of the frequency of Suspect B's genotype, based upon the STR allele frequencies within Suspect B's ethnic group, reveals that the likelihood that a random member of this ethnic group has this profile is about 1 in 1.5 billion. It is important to understand that this number is the probability of seeing this DNA profile if the crime scene evidence did not come from the suspect but from some other person. To regard the number as the probability that the suspect is the source of the crime scene evidence is to commit the "prosecutor's fallacy";

this is the logical fallacy of equating "the probability that an animal has four legs if it is an elephant" (high) with "the probability that an animal is an elephant if it has four legs" (low). To go from one probability to another requires the use of Bayes' theorem and some prior (before the DNA profile) probabilities of the suspect being the source of the evidence. In addition, it is important to note that the probability of 1 in 1.5 billion is substantially increased if the actual source is a person related to the suspect, especially a sibling.

Confounding Circumstances of DNA Profiling

Sometimes, the DNA from crime scene evidence is in a very small quantity, poorly preserved, or highly degraded, so only a partial DNA profile can be obtained. When fewer than 13 STR loci are examined, the overall genotype frequency is higher, therefore making the probability of a random match higher as well. For instance, in the fictional case in Table 1, if data were only obtained for the first four STRs listed in Table 1, the likelihood of encountering this genotype would be roughly 1 in 331. In this instance, prosecutors would need additional types of evidence against Suspect B to convince a jury that he or she was the source of the evidence sample. In addition, if an individual happens to have STR alleles that are very common in his or her ethnic group, the genotype frequency can also be quite high, even when all of the core 13 STR loci are examined. It is also important to note that crime scene samples sometimes contain DNA from several different sources. This can make teasing out the sources of the DNA extremely difficult.

Databases of DNA Profiles

DNA evidence is used in court almost routinely to connect suspects to crime scenes, to exonerate people who were wrongly convicted, and to establish or exclude paternity. DNA data is considered to be more reliable than many other kinds of crime scene evidence. For this reason, tissue samples are frequently collected by law enforcement officials from those individuals who are implicated (even loosely) in a crime. The unique profile of each DNA sample is analyzed for comparison to crime scene evidence. The DNA profile can also be stored in a database to compare with crime scene evidence from past and future crimes.

But under what circumstances should an individual be compelled to provide a sample for a DNA database? Originally, statutes mandating collection of tissue for DNA typing applied only to those people convicted of sex crimes or murder. This was due to the fact that there was usually an abundance of DNA at the scene of a rape or murder to compare to a suspect's DNA. More recent DNA collection laws have applied to all convicted felons, reflecting advances in DNA technologies that allow sufficient DNA samples to be obtained from scenes of more common crimes, such as burglary (Figure 1).

Country (Year established)	Reference profile size	Crime-scene sample size	Suspect to scene hits	Scene to scene hits	Entry criteria for suspects	Entry criteria for convicted offenders	Removal criteria
UK (1995)	2.5 million	200,000	550,000	30,000	Any recordable offence*	Entered as suspect	Never removed, including suspects
USA (1994)	1.52 million	67,000	Figure unavailable	Figure unavailable	No suspects entered, but under revision	Depends on state law	Depends on state law
Germany (1998)	286,840	54,570	13,700	5,500	Offence leading to >1 yr in prison	After court decision	After acquittal or 5–10 years after conviction, if prognosis is good
Austria (1997)	64,740	11,460	3,200	1,350	Any recordable offence*	Entered as suspect	Only after acquittal
New Zealand (1996)	44,000	8,000	4,000	2,500	No suspects entered	A relevant offence (including ≥7 yr in prison)	Never removed, unless conviction quashed
Switzerland (2000)	42,530	7,240	4,840	5,540	Any recordable offence*	Entered as suspect	After acquittal or 5–30 years after conviction
France (2001)	14,490	1,080	50	70	No suspects entered	Sexual assault and serious crime	40 years after conviction
Finland (1999)	8,170	5,450	2,080	780	Offence leading to >1 yr in prison	Entered as suspect	Only after acquittal
Slovenia (1998)	4,820	2,360	370	80	Any recordable offence*	Entered as suspect	Depends on severity of crime
Netherlands (1997)	4,260	13,700	2,520	4,260	No suspects entered#	Offence leading to >4 yr in prison	20–30 years after conviction
Sweden (2000)	3,980	9,860	2,500	4,750	No suspects entered	Offence leading to >2 yr in prison	10 years after release from prison

Figure 1: Characteristics of some national DNA databases.

© 2004 **Nature Publishing Group** Jobling, M. A. & Gill, P. Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 5, 745. All rights reserved. ⓘ

Figure Detail

As of July 2008, all 50 states in the U.S. have laws requiring convicted sex offenders to submit DNA samples, and 46 states have laws that require other types of convicted felons to do so. Additionally, 11 states require DNA samples from those convicted of certain misdemeanors, and 12 states have laws authorizing DNA sampling from arrestees of certain types of crime, usually felonies (National Conference of State Legislatures, n.d.). In comparison, in the United Kingdom, all arrestees, regardless of the degree of the charge, can be compelled to provide a sample. In fact, even those who are merely suspects in a crime can be forced to offer samples (Oak Ridge National Laboratory, n.d.).

Forced DNA Profiling

Those people opposed to DNA banking for law enforcement purposes note that arrestees are often found innocent of crimes. Retention of an innocent person's DNA can be seen as an intrusion of personal privacy and a violation of civil liberties. It is interesting to note that in the United States, under any other circumstance, the provision of a DNA sample would require informed consent and other protections for the donor. In contrast, an arrestee's DNA profile, once entered into a database, can be accessed by police, forensic scientists, or researchers without the consent of the donor. Another problem with the DNA database system is an exacerbation of the ethnic bias already present in the criminal justice system. If people from one ethnic group are more often arrested, tried, and convicted of felonies, they will be overrepresented in the database, potentially leading to even more arrests within that ethnic group.

Proponents of DNA databanks argue that major crimes often involve people who have also committed other offenses. Having DNA banked could potentially make it easier to identify suspects. It could also significantly cut down the cost of an investigation if an automated computer search could eliminate suspects or link a suspect to a crime scene

Does the DNA Databank System Help Solve Crimes?

The current DNA database maintained by the FBI, known as the Combined DNA Index System (CODIS), contains case samples (DNA samples from crime scenes or "rape kits") and individuals' samples (collected from convicted felons or arrestees) that are compared automatically by the system's software as new samples are entered. As of February 2007, CODIS had produced over 45,400 "hits," which assisted in more than 46,300 investigations (Federal Bureau of Investigation, n.d.). However, contrary to how DNA analysis is portrayed on popular television shows, DNA samples are not analyzed within the course of an hour. Rather, the U.S. currently has an enormous backlog of samples waiting to be typed and entered into the database. Some of these samples are from cases that have outlasted their statutes of limitation, so even if these samples could help solve a crime, the crime can no longer be tried.

This delay brings up the dilemma of the validity of statutes of limitation. These statutes were established at a time when large quantities of physical evidence were required to match a suspect to a sample and when extended time periods significantly decreased law enforcement's ability to find a match, as well as the likelihood of successful prosecution. With the advent of DNA databanks and the possibility of storing samples indefinitely, the very notion of a statute of limitation now seems extremely outdated.

Of course, there are many other debatable issues concerning DNA banking. For instance, should the original tissue sample be stored indefinitely after the DNA profile has been entered into the database? Detractors note threats to genetic privacy, but proponents argue that future DNA typing methods will undoubtedly be developed and that old samples might have to be reanalyzed using new techniques. Also at issue is the reopening of old cases on the basis of new (DNA-based) evidence. Which cases should be eligible for reanalysis in light of this new evidence? Can equitable rules be established to allow reexamination of cases that were analyzed with less powerful lab techniques? Further public awareness of the power of DNA forensic technology will help lawmakers decide these issues in a way that seeks to strike a balance between protecting individuals' genetic privacy and protecting innocent citizens from crime.

Conclusion

DNA evidence is easy to obtain because genetic material is found in all human cells, save red blood cells. As a result, when we leave behind small biological bits of ourselves, these bits can be used to identify us and link us to the places we've been. With modern technology, the amount of DNA required for analysis can be obtained from even a miniscule biological sample, which allows police to match crime scene evidence with suspects. However, because forensics is a science largely rooted in probabilities, even a confirmed "match" does not supply concrete proof of guilt. In addition, DNA databases designed to simplify the process of connecting past offenders to recent crimes are fraught with concerns involving individual genetic rights, as well as problems related to delayed sample entry, both of which hinder the ultimate usefulness of

these databases. As a result, even though forensics is undeniably important to the modern justice system, its personal ramifications and ethical questions are topics of continuing discussion within the scientific, law enforcement, and legal communities.

References and Recommended Reading

Federal Bureau of Investigation. "CODIS–NDIS Statistics." (accessed August 1, 2008)

Jobling, M., *et al.* Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 5, 739–751 (2004) doi:10.1038/nrg1455 (link to article)

Oak Ridge National Laboratory. "DNA Forensics." (accessed August 1, 2008)

National Conference of State Legislatures. "DNA Databanks." (accessed August 1, 2008)

Reilly, P. Legal and public policy issues in DNA forensics. *Nature Reviews Genetics* 2, 313–317 (2001) doi:10.1038/35066091 (link to article)

Weir, B. The rarity of DNA profiles. *Annals of Applied Statistics* 1, 358–370 (2007) doi:10.1214/07–AOAS128