Finding Criminals Through the DNA of Their Relatives—Familial Searching of the California Offender DNA Database

by Eva Steinberger and Gary Sims

Although the primary intent of DNA data banks is to find exact matches between forensic DNA profiles developed from unknown sources at crime scenes and DNA profiles submitted to the data bank by convicted offenders or arrestees, it is also possible to use the same genetic information to uncover familial relationships connected to the source of the forensic profile. Those familial connections may provide investigators focus and direction in their search for an offender's identity. While broadening the data bank search beyond the offender might help solve more crimes, there is also concern over the constitutional balance between privacy and governmental intrusion.

A familial search policy, *Partial Match Reporting and Modified CODIS Search Policy* (see Information Bulletin 2008-BFS-01*),* was issued by the Department of Justice (DOJ) in April 2008. The drafters of this policy sought to minimize privacy concerns through a carefully crafted set of conditions that must be met before DOJ will release the identity of an offender in the DNA data bank who may be the relative of the actual perpetrator. Furthermore, familial searching will only be allowed in major violent crimes where there is a serious risk to public safety, and all other investigative leads have been exhausted.

The DOJ manages and administers the DNA Data Bank. California's DNA Data Bank, formally established in 1990, is the fourth largest DNA data bank in the world and produces about 200 hits per month through weekly searches of 1.1 million offender DNA profiles against more than 20,000 forensic profiles. The data bank consists of a database of DNA profiles from convicted offenders and arrestees and a database of forensic profiles from crime scene evidence. These two databases form the California

CODIS (Combined DNA Index System). In a standard DNA data bank search, a forensic profile is searched against the offender database, and a hit is declared if the forensic profile is exactly the same as an offender's profile.

Familial searching is the expansion of a traditional DNA data bank search and is designed specifically to detect potential relatives. It can be useful when the standard DNA data bank search has not resulted in a direct match. Familial searching utilizes the fact that genes (alleles) in humans are inherited in a one-for-one relationship from the father and mother. Therefore, other close relatives can be expected to share more alleles, especially more rare alleles, than unrelated individuals. If the DNA profiles are not exactly the same, but close, with a significant number of alleles in common, the source of the forensic profile may be a relative of the offender. A familial search relies on scientifically valid mathematical tools designed to determine whether the similarity of two DNA profiles is more likely to be due to a familial relationship (kinship) or chance alone.

The Familial Search Program and the Development of Familial Search Tools

The Familial Search Program at the DOJ DNA Data Bank in Richmond is responsible for the familial search process. Law enforcement agencies may formally request a familial search for a particular case. The Familial Search Committee reviews every familial search request to ensure that familial searching is only used for major violent crimes where there is a serious risk to public safety and all other investigative leads have been exhausted. Multiple conditions, as outlined in the policy, must be met before

a familial search of the DNA data bank can begin. Serious consideration is given to balancing the need to develop a suspect in a major crime where every lead has been exhausted against the need to maintain constitutional privacy rights by applying only the most solid, rigorously validated scientific methods.

Discussion of Terms

The terms "partial matching" and "familial searching," although often used interchangeably, are related but are not exactly the same. A partial match can be the preliminary result of a DNA data bank search performed under standard conditions. More correctly termed a "candidate match," the offender profile is a very clear and complete DNA profile while the forensic sample may not be that at all. Or the candidate match may be the reflection of a DNA mixture from several individuals, or it may not consist of a complete DNA profile because the biological material had been degraded even before analysis. A candidate match will, upon close inspection, be declared a true data bank match (a hit), a non-match, or a situation where the two profiles, although clearly not the same, show a considerable degree of similarity. When the two DNA profiles show a considerable degree of similarity, familial search tools can be applied to determine whether the similarity might be due to kinship or chance.

California's DNA Data Bank, formally established in 1990, is the fourth largest DNA data bank in the world and produces about 200 hits per month through weekly searches of 1.1 million offender DNA profiles against more than 20,000 forensic profiles.

Locus Designation	Evidence Profile		Offender Profile		Alleles Shared at Locus
D8S1179	13	14	13	14	2
D21S11	29	30	29	29	1
D7S820	8	8	8	9	1
CSF1PO	10	11	10	11	2
D3S1358	14	15	15	15	1
TH01	7	7	6	7	1
D13S317	11	12	12	12	1
D16S539	9	11	9	11	2
D2S1338	19	19	19	20	1
D19S433	13	14	13	13	1
VWA	16	17	16	17	2
TPOX	8	8	8	9	1
D18S51	15	16	15	15	1
D5S818	11	12	11	12	2
FGA	20	21	20	20	1
Total Number of Shared Alleles					20 out of 30

Example of a partial match.

The DNA database is a collection of DNA profiles. Each DNA profile consists of a panel of DNA types at a particular set of loci. The typing technology used is called STR (Short Tandem Repeats) and is the national standard for forensic DNA testing. A locus is a location on the DNA molecule where variation may be observed in the form of different alleles of a gene. Each locus reflects two alleles of the same gene; one allele originated from the father, the other one from the mother. When comparing DNA profiles of individuals with each other in the context of familial searching, some of the same alleles may appear in unrelated people as well as in relatives. In general, however, close relatives, such as parents, children, or full siblings,

are expected to share more alleles with each other than with unrelated people. If the occurrence of an allele in the population is rare, sharing a rare allele is more significant than sharing a common allele. In any case, when two profiles share alleles (a partial match), additional tests must be conducted to sort out a true kinship from a similarity by chance.

Another fundamental technology used in familial searching is Y-STR typing. The DNA of the Y chromosome is inherited from a male's father, and, therefore, a human male and his male siblings will have the same Y-STR profile as their biological male children, their father, and their father's father. Y-STR typing also has limitations due to the way the Y chromosome is inherited, essentially as only one locus, called a haplotype.

Familial Search Tools

The fundamental requirement for a familial search tool is that it has the ability to find the true relative while minimizing the number of "false" relatives. On the surface, it may seem that the CODIS software itself should be a good tool for identifying relatives. One could just perform data bank searches targeted towards a partial match instead of the perfect match and investigate the partial matches further for kinship. To discuss the subject of searching DNA profiles, it is important to have a clear understanding of what the existing software in CODIS can do.

The CODIS software supports three match (or search) stringency levels: high, moderate, and low. These levels of stringency were originally designed to allow searching of more complex forensic profiles against offender samples. By adjusting the stringency of the search, matches that are not exact can still be made, allowing for further investigation.

High stringency matches require all alleles to match, with the same number of alleles in each sample. For example, at high stringency, an evidence sample that is type 11, 12 will only match an offender who is type 11, 12 (see Case A).

Match stringency.

Thus, as the stringency is relaxed from high to moderate to low, the number of allowable matches increases among the possible genetic combinations that could be compared. Research conducted at the DOJ Richmond DNA laboratory has shown that CODIS, in its current configuration, is a very poor tool for finding familial relationships. CODIS looks for allele-sharing patterns based on the level of stringency specified in a search and does not take into account the rarity in the population of a shared allele. And in order to cast the search net wide enough for a true relative, the stringency of a search would have to be relaxed so much that too many adventitious matches would obscure anything real to the point of being unmanageable.

Instead, looking for relatedness has to encompass the analysis of allele sharing as well as the rareness of each shared allele. A better way to approach the task is a statistical evaluation of the offender DNA profiles compared to the forensic DNA profiles, generating a pool of familial candidates based on their likelihood of being relatives of the source of the forensic sample. A likelihood ratio value above one may favor relatedness, while a number below one likely indicates that the DNA profiles are not related. As a practical matter, the DOJ approach is limited to looking for close relatives, i.e., parents and their children or full siblings. The data bank team developed a statistical software application, humorously called "The Ratiometer," that calculates kinship likelihood ratios of a given forensic profile with 1.1 million data bank profiles according to the following algebraic formula:

$$\text{Likelihood Ratio} = \frac{(KI) \times (Y\text{-factor})}{N}$$

KI = Kinship Index, which depends on the rarity of the shared alleles in the population and the type of kinship examined (full sibling or parent/child)
Y-factor = Y-haplotype frequency of the forensic sample, inverted
N = Number of offender samples compared

The Ratiometer filters for possible relatives by setting threshold values for the kinship index and/or the presence of at least 15 shared alleles.

Next, the output from the Ratiometer is ranked based on the kinship values. Once a ranked pool of familial candidates exists, additional DNA testing using Y-STRs can filter out the majority of unrelated persons. Assuming a reasonable probability that the relative, if in the data bank, would be in the upper rankings (for instance in the top 50 but in any case within a reasonable number of samples), subsequent Y-STR testing of this group should eliminate almost all of the unrelated persons.

	Y-STR Locus Designation				
	DYS393	DYS391	DYS439	DYS635	DYS392
Evidence	14	10	12	20	13
Offender 1	12	10	12	21	13
Offender 2	13	11	13	23	13
Offender 3	14	10	12	20	13
Offender 4	13	11	13	23	14
Offender 5	12	10	12	20	14

Comparison of offender Y-STR profiles with evidence Y-STR profile.

If the Y-haplotype of the offender and the forensic sample match, the results are presented to the Familial Search committee. Because there is still the possibility that the offender is not a true relative, DOJ has added another step prior to releasing the name. The DOJ Bureau of Investigation and Intelligence collects additional investigative data bearing on relatedness, and the findings are presented to the committee. The committee reviews all available information and decides whether to release the identity of the offender to the local investigative agency where the case originated.

As described, this combination of procedures constitutes a robust familial search tool that is the first of its kind in the nation. It strikes an effective balance between privacy concerns and the need to provide information that may solve a violent crime or series of crimes. As long as the familial search is limited to close relatives (i.e., parents and their children or full siblings who are in the database), there is a reasonable likelihood that such a relative can be identified using this approach. Familial searching is officially online at the DOJ DNA Laboratory in Richmond, and California's first familial DNA data bank search was conducted in October 2008.

Eva Steinberger received her undergraduate degree in biology in Switzerland and her Ph.D. from Cornell University in Ithaca, New York. She has been with the Department of Justice since 1994 and has been assistant chief in charge of DNA Programs since 2002.

Gary Sims received his undergraduate and graduate degrees in criminalistics/ forensic science from the University of California at Berkeley. He has been a practicing forensic scientist for 32 years and has been with the California Department of Justice DNA Laboratory since 1990. He currently serves as manager and technical leader of the laboratory's casework section.