Review

Forensic utilization of familial searches in DNA databases

Cassandra J. Gershaw*, Andrew J. Schweighardt, Linda C. Rourke, Margaret M. Wallace

John Jay College of Criminal Justice, 445 West 59th Street, North Hall, New York, NY 10019, United States

ABSTRACT

DNA evidence is widely recognized as an invaluable tool in the process of investigation and identification, as well as one of the most sought after types of evidence for presentation to a jury. In the United States, the development of state and federal DNA databases has greatly impacted the forensic community by creating an efficient, searchable system that can be used to eliminate or include suspects in an investigation based on their DNA evidence. Recent changes in legislation have begun to allow for the possibility to expand the parameters of DNA database searches, taking into account the possibility of familial searches.

This article discusses prospective positive outcomes of utilizing familial DNA searches and acknowledges potential negative outcomes, thereby presenting both sides of this very complicated, rapidly evolving situation.

© 2010 Elsevier Ireland Ltd. All rights reserved.

Contents

1. Introduction ................................................................. 16
2. General DNA database information ................................ 17
3. Familial searching ...................................................... 17
4. Policies dictating familial searching ................................ 18
5. Arguments for and against familial searching ..................... 19
6. Successfully utilized familial DNA searches ....................... 19
7. Conclusion and future directions ................................... 20
References ......................................................................... 20

1. Introduction

Such major advancements have occurred in the field of forensic science within the past fifteen years that investigative techniques are hardly comparable to what they were in the not-so-distant past. In the 1980s, forensic biologists began to establish DNA as a pillar of the investigative process. The application of DNA to forensic problems underwent prolific growth in the years that followed, and today DNA is one of the most highly regarded tools available to the forensic scientist. In fact, one of the most recent developments in forensic DNA technology is also potentially one of the most significant advancements in forensic science of the past fifteen years – the establishment of the Combined DNA Index System (CODIS), a computer database funded by the Federal Bureau of Investigation (FBI) which stores DNA profiles submitted by local, state, and federal crime laboratories in the United States of America.

Individual United States state codes (which vary from state to state) dictate not only which DNA profiles are uploaded to DNA databases, but also how the profiles in the databases can be searched. Proposed considerations to broaden the boundaries of forensic DNA searches have already become a highly contested issue, but policymakers (and the public) often find the information about this issue to be sparse and incomplete. The primary purpose of this article is to fill the void in the scientific arena by centralizing information surrounding familial DNA searches, thus making the issue more comprehensible. This article will present general information about DNA databases and the process of familial searching, as well as describe the new policies regarding familial searches in DNA databases. The authors seek to provide a well-rounded perspective by acknowledging both positive and negative outcomes of familial DNA searches.
potential outcomes of familial searches. The article will conclude by highlighting several case examples in which familial searching has been utilized successfully.

2. General DNA database information

Each of the fifty states in the United States has developed its own individual DNA database, with varying inclusion criteria. Many states are currently working to expand their inclusion criteria, mostly to include DNA profiles from all felonies [1]. Forty-seven states currently collect DNA samples from all felony convictions (Idaho, Nebraska, and New Hampshire collect from most but not all felonies) [2].

In 1990, the FBI created the CODIS database, a pilot program designed to allow federal, state, and local laboratories in the United States to store and compare DNA profiles in a way previously impossible [3]. CODIS consists of various searchable indices: Convicted Offender, Forensic, Arrestee (if permitted by that particular U.S. state code), Missing Persons, Unidentified Human Remains, and Biological Relatives of Missing Persons [3]. The Convicted Offender Index consists of DNA profiles of individuals who have been convicted of a criminal offense and the Forensic Index comprises DNA profiles obtained from crime scene evidence. When DNA evidence is recovered from a crime scene where investigators lack a suspect, the DNA profile generated from the evidence is searched against the DNA profiles within the Convicted Offender Index. If the search of the database yields a match between the unknown profile and a profile contained in the database, a sample obtained from the identified suspect is analyzed to confirm the match. If no match exists in the database, the DNA profile is then searched against the profiles within the Forensic Index. If there is a match within this index, it is a clear indication that these two (or more) crimes are linked and the appropriate investigators are notified so that they can pool their resources and information [1].

CODIS stores DNA profiles resulting from the analysis of 13 Short Tandem Repeat (STR) loci, designated by the FBI. These particular loci were chosen for their “high discrimination potential,” as well as their overlap with databases maintained by other countries [4]. These 13 loci were also chosen due to their lack of connection to genes whose presence may indicate a predisposition for certain genetic diseases [1].

When utilizing the CODIS software, three levels of search (or match) stringencies may be specified: high, moderate, or low [5]. With a high stringency search, the administrator requires that all of the alleles from the forensic DNA profile match all of the alleles in the CODIS DNA profile; in a moderate stringency search, the administrator relaxes the match criteria such that the software identifies DNA profiles in which each allele is represented, but a perfect match is not required [5]. With a low stringency search, at least one allele must match at each locus between the forensic and CODIS DNA profiles [5]. When the search stringencies are lowered, the search will yield a higher number of matches, resulting in further investigative leads and the potential for familial searching.

3. Familial searching

Familial searching is the process of purposefully searching a DNA database for a match at only a limited subset of the available typed loci (i.e. low stringency search) in an attempt to locate previously unknown relatives in order to open up new investigative leads. As previously described, such searches would yield a larger number of possible suspects by incorporating low stringency matches – DNA profiles that match the unknown profile at fewer loci. These low stringency matches may indicate a close relative to the source of the unknown forensic sample, thereby broadening the inclusion criteria of the searched DNA database to include not only offenders, but also these offenders’ relatives. In order to increase the probability of locating a previously unknown relative in the DNA database (and thereby obtaining a new investigative lead), several suggestions have been made that include modifying the search software to search for rare alleles, a high number of matching alleles, or perhaps utilize a likelihood ratio that indicates relatedness.

Searching for rare alleles might not be productive due to the fact that these alleles are rare and therefore one would not expect to find an abundance of them in any given population. Searching for a high number of matching alleles appears logical. It is based on the notion that a high number of matching alleles across all twenty-six possible alleles (two per each of the thirteen CODIS loci) might indicate relatedness between the perpetrator and the partial match. However, a high number of matching alleles might not be probative if the alleles that match are common. Due to the inherent flaws of a search based on rare or matching alleles, many researchers have turned their attention toward a search based on likelihood ratios.

A likelihood ratio is a way of predicting if the source of the unknown crime scene sample might be related to a particular individual in the DNA database. While certainly not infallible, a larger likelihood ratio is typically indicative of potential relatedness. Likelihood ratios may be used to compare the three alternatives for conducting familial searches because this approach considers rare alleles (e.g., allele frequencies) and matching alleles, while at the same time eliminating the difficulties that would be encountered by working with either alternative by itself.

Bieber et al. used a likelihood ratio formula to determine the probability of a relative being the closest database match to the crime scene sample [6]. A simulated database of offenders was created using published allele frequency data [7]. Using the laws of Mendelian inheritance and the available frequency data, genetic profiles were simulated for children of those already in the database. Each simulated child profile was then compared (using likelihood ratios) to every simulated offender (including the simulated parent) in a database of 50,000 individuals. It was discovered that parent–child kinships were identified 62% of the time as the best match (i.e., highest likelihood ratio). Although 62% is fairly reasonable, this also means that an unrelated individual would be identified as the best match 38% of the time. Therefore, an investigator might have to pursue several false leads before getting the right one, thus placing undue suspicion on individuals in the database (and their families) whose profile yielded a partial match by coincidence. This underscores the fact that unrelated people may also have similarities in their genetic profile. Additionally, it should be noted that the percentage provided above (62%) is proportional to the size of the database. In this particular example, the size of the database was 50,000 (the approximate size of an American state DNA database). However, supporting online material for the research of Bieber et al. demonstrates that the chance of a relative yielding the best match diminishes as the size of the database increases, for example to the size of a national DNA database [8].

In the same analysis it was also found that parent–child kinships were identified 99% of the time in the top 100 matches. Again, these data may seem to support the utility of familial searches, but it should be noted that it might not necessarily be helpful if a parent–child kinship is identified in the top 100 matches. In other words, this might mean that an investigator would have to pursue 99 false leads before finding an actual relative of the perpetrator. According to Bieber et al., the chance of a relative being in the top ten leads is 80% [6]. Some supporters of familial searching like those odds, especially when it might not be uncommon for police to go through dozens of leads if the crime is serious. However, those who are impressed by the statistics are
cautioned to remember that these data are based on a fairly substantial (and perhaps optimistic) assumption that a relative of the perpetrator is actually in the database.

It is still unclear whether future familial searches will be based on rare alleles, common alleles, or likelihood ratios. As promising as it already seems, it has been suggested that the likelihood ratio method could be further enhanced by accounting for geographic location, among other factors [8]. Having considered all three alternative methods of conducting a familial search, it seems that the use of likelihood ratios that account for allele frequencies may offer the best opportunity for correctly identifying a relative of the perpetrator in the database, if one exists.

4. Policies dictating familial searching

When a laboratory searches for a DNA match by comparing the profile of an unknown forensic sample against either the profiles of convicted offenders or other unknown samples, the search will yield a number of results, most of which will not match the profile at all 13 loci (if a perfect match exists in the database at all). While these partial matches may arise in a routine searches, partial matches can also be purposefully sought by conducting a familial search. The ability to deliberately perform these searches as well as the authority to release information to investigators on partial matches that commonly arise is dictated by individual United States state codes and ever-broadening policies. While most U.S. states have remained ambiguous on the concept of familial searching, only New York and Massachusetts have specifically addressed the issue of utilizing low stringency searches in forensic cases [9]. Both require a minimum of four loci to complete a forensic search against their respective state DNA databases but do allow for exceptions under certain circumstances (e.g. the involvement of relatives, sample degradation, limited sample availability, etc.) [for full relevant code information, see Mass. Regs. Code 515 § 2.14 and N.Y. Comp. Codes R. & Regs. 9 § 6192.3] [9].

However, it should be noted that while the New York and Massachusetts state codes may allow low stringency searches, this does not mean that the states’ forensic laboratories are permitted to report information gleaned from partial matches to law enforcement. For example, the State Police Crime Laboratory in Massachusetts bars staff from seeking and/or reporting partial matches, although this ban is currently being reconsidered (note: this “reconsideration” has been on-going for over a year) [10]. As for New York, the New York State Commission on Forensic Science is currently pursuing a regulatory change which would allow laboratory officials to inform law enforcement about inadvertent partial matches that arise during routine searches of the DNA Databank [11]. The Commission is a twelve member panel comprised of representatives from “law enforcement, criminal defense, the judiciary, the state Department of Health, forensic laboratories...empowered to develop minimum standards and a program of accreditation for all public forensic laboratories in New York State” [11]. Prior to June 2009, New York State forensic laboratory officials were only permitted to report to law enforcement if a crime scene DNA sample matched a particular individual in the state's DNA Databank [11]. After thorough research by the Commission’s DNA Subcommittee, the Commission has decided to develop a new regulation to allow partial match information to be utilized discreetly and responsibly [11]. Once the Commission has drafted its partial match policies and procedures, it will then review and debate the recommendations over the upcoming months. However, this new policy is being drafted to address partial match information release only; it will not permit familial searching [11].

As previously mentioned, New York and Massachusetts are the only two of the United States to specifically address low stringency searches in their state codes. Other states are not so clear in their language on appropriate DNA searches, or on how many matching alleles are required in order to identify an individual to investigators. For example, in Virginia, in order for a partial match to be disclosed to investigators, the DNA partial match must be “very, very close” and in California, the partial match must “appear useful” before it may be disclosed [9]. Florida’s policy language is more lucid: 21 of the possible 26 alleles must match in order for investigators to be informed of the partial match [9].

While the language addressing forensic familial searches may be lacking or unclear, a number of states in the U.S. are updating and expanding their policies regarding familial searches in their DNA databases. California is the vanguard, having developed the most aggressive policy to date. Adopted in 2008, California’s familial searching policy allows the state’s crime lab to inform the police of partial matches that arise during routine searches of California’s offender DNA database [12]. The lab follows up by performing a series of tests and calculations to determine the “likelihood of a biological relationship between the person found in the database and the unknown offender believed to have left DNA at the crime scene” [12]. If these partial matches do not arise or fail to produce a lead, a more customized familial search can be performed utilizing computer software which scans the database proactively for possible relatives and measures the probability of relatedness based on the rarity of shared alleles. A series of meetings occurs throughout this process between police, prosecutors, and a Department of Justice committee. The purpose of this collaboration is to ensure all parties concerned that the low stringency search is utilized as a last resort, that both the prosecutor and the law enforcement agency are committed to continuing the investigation, and that the relative’s name is indeed vital to the investigation [12].

Colorado has also aggressively been pursuing the process of familial searching. In fact, Denver’s District Attorney, Mitchell R. Morrissey, has been lobbying for new legislation since 2005 [13]. Colorado recently performed a test run of familial search software on its state-wide DNA database [14]. After running their new familial search software against a local database of unsolved cases, Denver investigators discovered new leads on cold cases, many of which had remained unsolved for years [13]. Denver’s DNA Human ID Research is laying the procedural groundwork for other states to follow. All of the unknown forensic samples (900 in total) were searched against Denver’s Local DNA Index System (LDIS) (which contains 1700 profiles) using Dr. Charles Brenner’s DNA-VIEW software [15]. In six separate cases, investigators found a 90% chance that two individuals were related, and follow-up Y-STR (short tandem repeat on the Y chromosome) DNA testing indicated that the individual had the same Y-DNA type [15]. On a larger scale, Colorado performed the same process utilizing their State DNA Index System (SDIS) – all forensic unknown samples were searched against Colorado’s state DNA database (in this instance, approximately 2000 samples against 80,000 profiles) [16]. Using the same DNA-VIEW software, 13 separate cases were discovered in which there existed a 90% chance that two individuals were related [16]. In another 15 separate cases, a 75% chance existed that two individuals were related [16]. Following Y-STR testing, the names of the offenders are obtained from the state database and this important lead is utilized in a routine investigation [16]. Colorado hopes to create policies which will ensure the forensic unknown samples will be run against the SDIS on a semi-annual basis [15].

In 2005, Florida updated its policy to allow database operators to give investigators the names of convicted offenders if the offender matched the crime scene sample at a minimum of 21 alleles [10]. In a related effort, Florida has also found a new way to solve rape cases. Operators use the DNA of children born to rape victims to search the database for the children’s fathers, thereby leading them to the rape suspect [10].
In contrast to the previously discussed states that are attempting to broaden their utilization of their DNA databases, one state has adamantly denied the expansion. In April 2008, Maryland banned familial searching in a law adopted to enlarge the state database to include the DNA of anyone charged with a violent crime [17].

With the ever-growing body of opinions and science surrounding this new tool in investigations, it is interesting to note that many of the policy changes are occurring on the coasts. States like Missouri and Illinois have yet to discuss the possibility and their laws do not address familial searching. In a random polling of police departments about the subject, only a few of the larger precincts even knew familial searching in DNA databases was possible [18].

As for the FBI and CODIS, the process of familial searching has not yet been accepted. In fact, the FBI is barred from performing familial searches within CODIS by federal privacy laws [10]. Thomas Callaghan, the former director of the CODIS program, has stated that the “FBI would be more comfortable with congressional authorization to conduct familial searches” [17]. An interim policy has been enacted to deal with the partial matches that arise – the policy allows states to make the final decision on whether or not to follow up to see if a relative is involved [17]. An FBI advisory group has suggested a final policy that would subject these partial matches to additional DNA testing and statistical analyses in an effort to aid investigators in locating relatives of profiles in CODIS [17]. The FBI has also hosted a Familial Searching and Genetic Privacy Symposium (January 2008) to address emerging questions with regard to the legality of familial searches.

5. Arguments for and against familial searching

For the United States, the process of familial searching has long been utilized in cases of unidentified human remains, but is only recently being considered as a potential method for identifying suspects. In the United Kingdom, familial searching has been utilized for years, with a 10–14% success rate of identifying offenders [12]. Familial searching holds numerous investigative possibilities, the most important of which is the potential to resolve a higher percentage of cases. With the ability to compare DNA samples against profiles in DNA databases using a fewer number of alleles, investigators would have the opportunity to hone their list of suspects on a truncated timeline, a process that is crucial during time-sensitive inquiries. Bieber et al. suggest that under the right circumstances and including the appropriate kinship analysis, the probability of getting a hit in a DNA database could increase as much as 40% [6]. Researchers agree that the success of DNA databases as a whole is contingent on there being a matching or near-matching DNA profile. Familial searching advocates point to the 1999 U.S. Department of Justice survey in which 46% of jail inmates indicated that they had at least one close relative who had been incarcerated to emphasize the importance of utilizing familial DNA searches [19]. On the other hand, considering this statistic, privacy advocates argue that familial searching may increase discrimination and racial disparities. Statistically, certain groups (e.g. racial, ethnic, social class, etc.) are more likely to be arrested and/or incarcerated, and therefore these same groups are more likely to have submitted a DNA sample for storage in a database. Due to their higher representation in DNA databases, families belonging to a certain demographic may be unfairly scrutinized [20]. Unfortunately, this disparity neither begins nor ends with DNA databases and in order to eradicate the problem, the entire justice system would need to be reworked, not just the process of familial searching [21].

A major apprehension touted by privacy advocates is that police officers will abuse the information gained by familial DNA searches to harass and embarrass innocent people and/or family members of the perpetrator. This contention, supporters of familial searching argue, is invalid. Legislators and laboratory personnel are working together to create procedures and policies that will dictate when and how information obtained through familial DNA searches is released to investigators, and to verify that such information will not be discharged without careful evaluation. In California, various factors are assessed before any names are released to police personnel, thus reducing unnecessary intrusions. Such factors include family structure, criminal history, whether the individual was in or out of custody at the time of the crime, and the number of owed samples [21]. Y-STR profiles are also a condition for disclosure [21]. This background information is verified via computer to avoid confrontations until absolutely necessary.

A related concern is the lack of procedural consistency with respect to familial searches. Due to the novelty of this issue, very few police agencies and laboratories have standard operating procedures (SOPs) in which familial searching has been explicitly addressed. Even so, there is a substantial degree of variation in those organizations that do have well-defined SOPs. Having a clear policy on this issue is important because the hits generated by familial searches can be so voluminous [22]. In many cases, partial matches yielded by a familial search belong to individuals not related to the unknown offender at all [22], and these “red herrings” may waste crucial investigative time. It should be noted that researchers and law-makers have recognized the need both for standardization and the need for scientific research that corroborates the usefulness of familial searching (i.e. California’s intensive multiple step process).

6. Successfully utilized familial DNA searches

For the United States, the process of familial searching might be a new and controversial issue but Great Britain has been successfully utilizing familial searching for years. In tandem with traditional investigative techniques, familial searches have been effective in identifying offenders in a variety of forensic cases, from car break-ins and robberies to sexual assaults and homicides. For instance, in 1993 a female passenger was sexually assaulted by her cab driver. Initial searches within the UK National DNA Database failed to produce a match but scientists were able to create 700 possible familial DNA groupings [14]. Using these groupings, detectives began eliminating suspects by age, ethnicity, and/or geographical location and were able to narrow the list down to thirty suspects whose DNA closely resembled that of the rapist [14]. Included on that short list was the brother of the rapist. Detectives then obtained a DNA sample from the offender and were able to arrest Tahir Mahmood, the cab driver rapist [14].

In another case, a serial rapist was able to evade capture for over twenty years. James Lloyd sexually assaulted women as they walked home alone, taking their shoes with him after the assaults as trophies [14]. Lloyd’s DNA was not included in the UK National Database but by utilizing a familial DNA search, police narrowed down their suspect pool to forty suspects and planned to interview each individual [23]. The third door they knocked on was that of Lloyd’s sister, whose DNA had been entered into the National Database after her arrest for driving under the influence [23]. Police were then able to match Lloyd’s DNA to a sample taken from one of the sexual assaults he committed.

Similar to James Lloyd, another serial rapist able to avoid arrest for over two decades was arrest and convicted with the assistance of familial searching. In the 1980s, Derek Young attacked three women at knife-point [14]. DNA evidence taken from the assaults yielded no matches in the National Database but when the cases were reopened, a familial search was run and the DNA evidence linked the offender to Young’s son [14]. Young was arrested and sentenced to fifteen years [14].
The United States has also seen the utility of familial DNA but in a slightly different manner. In these United States examples, cases closed with the assistance of familial DNA deal with the process of investigating partial matches that commonly arise during routine database searches rather than the purposeful search for low stringency matches, as is more common in the United Kingdom. For example, in 1989, a 29-year-old woman was found murdered in her South Carolina home. DNA evidence and fingerprints were found at the crime scene but neither matched any database samples. The DNA evidence from the 1989 case was routinely run through the database but did not yield any matches – perfect or near – until seventeen years after the crime. The offender’s brother was arrested (for a different crime) and his DNA profile was entered into the system. When the brother’s profile came back as a partial match, investigators followed up and obtained a blood sample from Tony Oliveo Mack, whose DNA was a perfect match [14]. Mack is now serving a thirty-year sentence.

In another more famous case, familial DNA led to the downfall of the notorious BTK serial killer, Dennis Rader. Investigators had zeroed in on this church-going community man but lacked evidence. A court order was obtained for Rader’s daughter’s sample from Tony Oliveo Mack, whose DNA was a perfect match [17]. Her DNA profile was such a close match to that of DNA evidence taken from several BTK crime scenes that investigators were able to secure an arrest warrant for Rader, thus ending his thirty-year spree.

Utilizing familial DNA evidence has also proven useful in exonerating the innocent. In 2004, innocent North Carolina man Darryl Hunt was freed after DNA evidence recovered from the rape/murder scene was run against the state DNA database and a partial match came back. The partial match was the DNA profile of a felon included in the database. By investigating relatives of the felon, officers were able to obtain a DNA sample from the felon’s older brother, Willard Brown, who subsequently confessed to the crime [10].

7. Conclusion and future directions

Familial DNA searching can prove to be an invaluable tool for forensic investigations if the process and resulting information are utilized as just that – an investigative tool. Advocates of familial searching advise that these searches should be neither the first step nor the sole source of information, but should instead be used only as a portion of the investigation when all other leads have been exhausted. On the contrary, critics remain adamant about the potential for privacy violations and police who might abuse the practice. With certain American states (e.g. California and Colorado) utilizing the available manpower and funding money in their budgets to develop new DNA database software, the groundwork is being laid for more jurisdictions to rely on the familial searches.

Advocates assert that the science that corroborates familial DNA searches is strong, yet policymakers seem reluctant to acknowledge this. Supporters of familial searching are advised to press law-makers for structured procedural consistency surrounding this issue. SOPs for investigations and uniform disclosure policies would help to assuage the concerns of privacy advocates. Researchers can also help to expedite the process by providing evidence that bolsters the scientific underpinnings of familial searching. For example, there is a need to develop standards for the statistics required to justify an investigation based on a forensic familial DNA search. Familial searching may be the future of DNA evidence, but it is far from universal acceptance. With states like California forging ahead, other states may follow this lead after allowing time to learn from the mistakes and successes of the pioneers. Alternatively, familial searching could be abandoned because of unresolved privacy concerns, lack of scientific data, and/or a weak legal framework (despite the fact that each of these issues could be resolved with the proper dedication). These problems are reminiscent of the deep-rooted ethical dilemma that is often encountered when innovative technologies become available: practically, it is possible to perform familial searches but ethically, should they be performed?

References

[18] J. Kohler, Crime Scene DNA can Lead to Knock on Relative’s Door, St. Louis Post, June 17, 2008.