Forensic Science International: Genetics 5 (2011) 16–20

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

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

ARTICLE INFO

Article history:
Received 30 January 2010
Received in revised form 8 June 2010
Accepted 25 July 2010

Keywords:
Forensic science
DNA
CODIS
Familial search
Low stringency search
DNA legislation

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 bas 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.

Contents

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 the 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

---

* Corresponding author. Present address: 6 Skytop Drive, Denville, NJ 07834, United States. Tel.: +1 973 615 1282.
E-mail address: cjg252@nyu.edu (C.J. Gershaw).

1872-4973/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2010.07.005

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 the most viable of 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