Finding Criminals Through DNA of Their Relatives

Frederick R. Bieber, Charles H. Brenner, David Lazer

DNA methods are now widely used for many forensic purposes, including routine investigation of serious crimes and for identification of persons killed in mass disasters or wars (1–4). DNA databases of convicted offenders are maintained by every U.S. state and nearly every industrialized country, allowing comparison of crime scene DNA profiles to one another and to known offenders (5). The policy in the United Kingdom stipulates that almost any collision with law enforcement results in the collection of DNA (6). Following the U.K. lead, the United States has shifted steadily toward inclusion of all felons, and federal and six U.S. state laws now include some provision for those arrested or indicted.

At present, there are over 3 million samples in the U.S. offender/arrестee state and federal DNA databases (7). Statutes governing the use of such samples and protection against misuse vary from state to state (8).

Although direct comparisons of DNA profiles of known individuals and unknown biological evidence are most common, indirect genetic kinship analyses, using the DNA of biological relatives, are often necessary for humanitarian mass disaster and missing person identifications (1, 2, 9). Such methods could potentially be applied to searches of the convicted offender/arréste DNа databases. When crime scene samples do not match anyone in a search of forensic databases, the application of indirect methods could identify individuals in the database who are close relatives of the potential suspects. This raises compelling policy questions about the balance between collective security and individual privacy (10).

To date, searching DNA databases to identify close relatives of potential suspects has been used in only a small number of cases, if some-
times to dramatic effect. For example, the brutal 1988 murder of 16-year-old Lynette White, in Cardiff, Wales, was finally solved in 2003. A search of the U.K. National DNA Database for individuals with a specific single rare allele found in the crime scene evidence that identified a 14-year-old boy with a similar overall DNA profile. This led police to his paternal uncle, Jeffrey Gafoor (11). Investigation of the 1984 murder of Deborah Sykes revealed a close, but not perfect, match to a man in the North Carolina DNA offender database, which led investigators to his brother, Willard Brown (12). Both Gafoor and Brown matched the DNA from the respective crime scenes, confessed, and were convicted.

Although all individuals have some genetic similarity, close relatives have very similar DNA profiles because of shared ancestry. We demonstrate the potential value of kinship analysis for identifying promising leads in forensic investigations on a much wider scale than has been used to date.

Let us assume that a sample from a crime scene has been obtained that is not an exact match to the profile of anyone in current DNA databases. Using Monte Carlo simulations (13, 14), we investigated the chances of successfully identifying a biological relative of someone whose profile is in the DNA database as a possible source of crime scene evidence (15). Each Monte Carlo trial simulates a database of known offenders, a sample found at a crime scene, and a search. The search compares the crime sample with each catalogued offender in turn by computing likelihood ratios (LRs) that assess the likelihood of parent-child or of sibling relationships (1, 16). We used published data on allele frequencies of the 13 short tandem repeat (STR) loci on which U.S. offender databases are based and basic genetic principles (17–19). A high LR is characteristic of related individuals and is an unusual but possible coincidence for unrelated individuals. The analysis of each simulation therefore assumes that investigators would follow these leads in priority order, starting with those in the offender database with the highest LR for being closely related to the owner of the crime scene DNA sample.

Our simulations demonstrate that kinship analysis would be valuable now for detecting potential suspects who are the parents, children, or siblings of those whose profiles are in forensic databases. For example, assume that the unknown sample is from the biological child of one of the 50,000 offenders in a typical-sized state database. Of the 50,000 LRs comparing the “unknown” sample to each registered offender in the database, the child corresponds to the largest LR about half the time, and has a 99% chance of appearing among the 100 largest LRs (see chart). An analysis of potential sibling relationships produced a similar curve (13).

These results could be refined by additional data—for example, large numbers of single-nucleotide polymorphisms (SNPs). Better and immediately practicable, a seven-locus Y-STR haplotype analysis on the crime scene and the list of database leads would eliminate 99% of those not related by male lineage (20). Data-mining (vital records, geographical and geographical data) for the existence of suitable suspects related to the leads can also help to refine the list.

The potential for improving effectiveness of DNA database searches is large. Consider a hypothetical state in which the “cold-hit” rate—the chance of finding a match between a crime scene sample and someone in the offender database—is 10%. Suppose that...
among criminals who are not (yet) in the database themselves, even 5% of them have a close (parent/child or sibling) relative who is. From our projections that up to 80% (counting the 10 best leads) of those 5% could be indirectly identified, it follows that the kinship analyses we describe could increase a 10% cold-hit rate to 14%—that is, by 40%. There have been 30,000 cold hits in the United States up to now (5). Kinship searching has the potential for thousands more.

Success of kinship searching depends most saliently on a close relative of the perpetrator actually being in the offender database. Studies clearly indicate a strong probabilistic depend-

(23, 24). These DNA collections have sparked considerable controversy, especially in light of recent trends to expand collections to arrestees and those convicted of minor crimes and misdemeanors (25). Although use of retained samples for other purposes is prohibited by federal and several state laws, sample retention also has been a controversial practice.

Debates on the expansion of the scope of DNA collections for offender and arrestee databases, as well as collections of volunteer samples, e.g., through DNA dragnets, have concentrated on the balance between society’s interests in security and privacy interests of individuals who might be included in the database and on the fairness and equity of including some in the databases but not others (26, 27). Privacy interests include genetic privacy [as DNA samples can yield medical and other information (28)] and locational privacy (where the contributor has been and left DNA). As with any investigative technique, these DNA matching strategies will lead to investigation of the innocent.

Existing state and federal statutes do not specifically address familial searches, and it is unlikely such search strategies were even considered at the time original statutes were written. Use of familial searching methods described herein could raise new legal challenges, as a new category of people effectively would be placed under lifetime genetic surveillance. Its composition would reflect existing demographic disparities in the criminal justice system, in which arrests and convictions differ widely based on race, ethnicity, geographic location, and social class. Familial searching potentially amplifies these existing disparities. These issues need to be confronted, as widespread use of various familial searching tools, including formal kinship analysis, is foreseeable. The de facto inclusion of kin into DNA data banks may lead some to oppose familial searching. It may lead others to support calls for a universal DNA database (29), which to date have been rejected. Other options include limiting familial searching methods to investigation of the most serious crimes and defining statistical thresholds that minimize intrusions on innocent parties (30).

The rapid proliferation and expansion of DNA collections along with the results of our analyses require careful consideration of the implications of familial searching methods. Every agency or country considering such methods should evaluate attendant policy, ethical, and legal implications, in addition to their valuable investigatory potential.

References and Notes
5. A. F. Griffiths et al., An Introduction to Genetic Analysis (Freeman, New York, ed. 2, 2004).
6. Data from Y-Chromosome Haplotype Reference STR database (YHRD), see (www.yhrd.org).
17. We gratefully acknowledge the American Society of Law, Medicine and Ethics, through NIH grant 1R01 HG018203-01, and the NSF (grant D131923 to D.L.) for partial support. We thank E. Smith for research assistance.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1122655/DC1

Published online 11 May 2006; 10.1126/science.1122655
Include this information when citing this paper.

10.1126/science.1122655