

The Evolution of DNA Databases: Expansion, Familial Search, and the Need for Reform

Michael Seringhaus, PhD - Yale Law School

INTRODUCTION	2
I. SCIENTIFIC FOUNDATIONS OF FORENSIC DNA	3
<i>DNA, the Genome, and STRs</i>	3
<i>Inheritance</i>	8
<i>Summary: The Genome, from a Forensic Standpoint</i>	9
II. ORIGIN, CONTENT AND FUNCTION OF DNA DATABASES	10
<i>American forensic DNA databases: Origin and current organization</i>	10
<i>An Odds Game: The 13 CODIS markers</i>	12
III. DNA DATABASES: LEGAL FOUNDATIONS	16
<i>Statutory Authorization</i>	16
<i>DNA Collection is a Fourth Amendment Search</i>	18
<i>Similarity to Fingerprinting</i>	21
<i>Summary of Legal Standards</i>	22
IV. NEW CHALLENGES: WILL THE LEGAL FOUNDATION HOLD?	22
<i>Including Arrestee DNA</i>	22
<i>Familial Search</i>	26
<i>Other Issues: Phenotypic Profiling, Gems in the Junk, and Sample Retention</i>	36
V. TECHNICAL AND POLICY REFORM	38
<i>Technical Reform: Expanding the CODIS Loci</i>	39
<i>Policy Reform: In Support of a Universal DNA Database</i>	41
VI. CONCLUSION	45

INTRODUCTION

Every human on earth – with the exception of identical twins – carries unique genetic material. Although the vast majority of this material, called the genome, is shared by all people, about 15 million nucleotides are not. It is this tiny fraction of divergent genetic material that makes forensic DNA analysis possible.

The Federal Bureau of Investigation (FBI) Combined DNA Index System (CODIS) has been fully operational for just over a decade, built upon a set of just 13 genetic markers from a certain class of convicted felons. Very recently, this system of DNA databases has grown to include arrestees not convicted of any crime, and also to allow familial DNA searching: A type of partial-stringency matching designed to return close relatives of individuals profiled in the database. Through this ambitious expansion, the FBI may already have overstepped its legal authorization for the database: A once-solid Fourth Amendment justification that authorizes compulsory DNA collection based upon offenders' diminished expectation of privacy.

This paper summarizes the scientific and legal foundations for CODIS, examines the aforementioned developments, details other challenges to CODIS including racial bias, advances in phenotypic profiling and the related issue of sample retention, and argues both for an immediate expansion in CODIS markers, and more generally for a sample-free, records-only universal DNA database.

Section I presents scientific background to acquaint those new to the field with the biology necessary to fully appreciate what the CODIS markers are, and what they are not. Section II outlines the history and organization of DNA databases, and explores in detail some features of the 13 CODIS markers. Section III presents the legal underpinnings of forensic DNA databases, including statutory authorization and legal treatment by the courts, primarily on Fourth Amendment grounds. Section IV outlines recent and coming challenges facing CODIS, including rapid expansion, familial DNA

searching, phenotypic profiling, and the issue of racial bias. Section V sets forth the two suggested reforms, a universal DNA database and an expansion of CODIS loci. Section VI concludes.

I. SCIENTIFIC FOUNDATIONS OF FORENSIC DNA

DNA, the Genome, and STRs

A human being consists of roughly 50-100 trillion cells, virtually all of which carry a full diploid copy of that individual's complete genetic information, or genome. This genome in turn consists of some 3 billion nucleotides of deoxyribonucleic acid (DNA) – the four component nucleotides of which are chemical bases commonly represented by the letters A, C, G and T – arranged into 23 long strands, called chromosomes.¹ There are 22 autosomal chromosomes (numbered 1 through 22) and one sex chromosome; the latter can take two forms (X or Y). A normal human cell carries two copies of each autosomal chromosome, and two copies of the sex chromosome, making 46 chromosomes in total; the normal sex chromosome arrangement is XX in females, and XY in males.² The human genome, then, is a duplicate set of long chemical strands that for simplicity can be envisioned as a single concatenated text string of about 3 billion A, C, G and T characters. To the unaided observer, the order of these characters appears largely random.

¹ Each chromosomal strand is itself a long DNA double helix, coiled around a protein scaffold. Each double helix consists of two complementary strands of DNA base-paired together: that is, A pairs with T, and G pairs with C. Because the two strands are complementary, the sequence of one dictates the sequence of the other, and as such, scientists need only report the sequence of one of the strands.

² Some cells differ in their DNA complement. For instance, mature red blood cells do not have a nucleus, and thus do not carry genomic DNA. Sex cells (sperm and egg cells) each carry only a half complement, namely a single copy of each autosomal chromosome and a single X or Y.

Every living thing on Earth carries its genetic material in strands of DNA, although their chromosomal number and arrangements differ.³ These long strings composed of just four chemicals are capable of coding for a stunning diversity of life.⁴

An individual's precise genetic DNA sequence, or genotype, helps determine its appearance, or phenotype.⁵ Two individuals of a given species – say, two humans, or two hamsters – will share largely identical genomes, with small differences in sequence contributing to differences between individuals. In two humans selected at random, the sequence of nucleotides will be 99.5% identical.⁶

This near-identity can cause some confusion when discussing the human genome. The draft human genome, sequenced and widely publicized in 2001,⁷ was in actual fact a mixture of several individuals' genomes called a reference genome. Researchers at that time were interested in elucidating the sequence of the vast majority of DNA shared by all humans. In this context, sequencing the 'human genome' was a monumental achievement that continues to shed light on the genetic and biological underpinnings of our species.

³ Some types of virii carry their genetic information as ribonucleic acid (RNA), a chemical compound very similar to DNA. However, a virus is not considered free-living, since it must infect a host cell and hijack its machinery to reproduce.

⁴ The genome of a rosebush, for instance, is very different from that of a whale, both in sequence length, chromosomal number and arrangement, and sequence content. However, both genomes consist of long strings of these same four nucleotides. It is possible to compute similarity scores between any two species based on their DNA, and thereby to deduce their evolutionary distance and relatedness. This is called phylogenetics.

⁵ Genotype is important, but not all-determining. The standard rubric taught to biology undergraduates is that genotype + environment = phenotype. Environmental conditions (which includes, among many other things, availability of nutrients at given points in the developmental process) can influence biological development as well. This is readily observable in identical twins, or more recently, in cloned animals: these individuals, despite being genetically identical with one another, nonetheless exhibit some slight morphological differences.

⁶ Samuel Levy et al., *The Diploid Genome Sequence of an Individual Human*, PLOS BIOLOGY 5(10):e254 (2007).

⁷ See International Human Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 NATURE 860 (Feb. 15, 2001); J. Craig Venter, et al., *The Sequence of the Human Genome*, 291 SCIENCE 1304 (Feb. 16, 2001). See also International Human Genome Sequencing Consortium, *Finishing the euchromatic sequence of the human genome*, NATURE 431(7011): 931 (Oct. 21, 2004). For genome sequence and information, see NCBI Human Genome Resources, <http://www.ncbi.nlm.nih.gov/genome/guide/human/> (last visited June 19, 2009).

More recently, attention has turned to the relatively small amount of genetic material at which individuals differ.⁸ The roughly 0.5% of divergent sequence corresponds to about 15 million nucleotides of DNA at which any two individual humans may differ. Some of this sequence may account for commonly observed differences such as height, eye color, skin pigmentation and so forth; other divergent DNA sequence may have no outwardly observable effect, and may therefore be useful chiefly as a marker to differentiate individuals from one another. It is the latter type of genetic markers that are used in forensic DNA identification.

To appreciate the way in which such forensic markers are used, it is necessary to understand in general terms the organization of the human genome. Each of the 3 billion nucleotides comprising the human genome can be classified as either coding or non-coding. Coding DNA corresponds roughly to genes, which in general usually means discrete DNA sequences that are ‘read’ to produce functional molecules such as proteins.⁹ Coding DNA accounts for less than 2% of the human genome.¹⁰ The other 98% is non-coding DNA, which was at one time called “junk DNA” due to its apparent lack of function.¹¹ Within these non-coding regions, a large number of repeating sequences can be found. One class of repeats called short tandem repeats (STRs) occur

⁸ See, e.g., the Personal Genome Project, <http://www.personalgenomes.org/> (last visited July 3, 2009). This project, led by Dr. George Church of Harvard Medical School, aims to sequence the genomes of a large number of individuals, enabling comparisons between them.

⁹ See Michael Seringhaus and Mark Gerstein, *Genomics Confounds Gene Classification*, AMERICAN SCIENTIST (Nov.-Dec. 2008) 466 (explaining in detail how the definition of a gene itself is evolving).

¹⁰ See ENCODE Project Consortium, *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*, 447 NATURE 799 (Jun. 14, 2007). Estimates of the number of genes in humans varies and is frequently revised as new discoveries are made; at the time of this writing, there are 21,370 known protein-coding genes and a further 5,732 RNA-coding genes in human, although the predicted number of actual genes ranges as high as 50,000. See Ensembl genome browser 54 – H. sapiens – Assembly and Genebuild, http://www.ensembl.org/Homo_sapiens/Info/StatsTable (last visited July 12, 2009).

¹¹ More recently, however, this term has fallen out of favor as scientists have begun to uncover novel functions in these vast non-coding regions. See, e.g., W. Wayt Gibbs, *The Unseen Genome: Gems Among the Junk*, SCIENTIFIC AMERICAN 29 (Nov. 2003).

when a pattern of two to ten nucleotides repeat and are adjacent to one another.¹² Each span having a different number of repeats is a different ‘allele’ of the STR.

Such clusters of short repeated sequences can arise, and very occasionally change in number, through DNA replication or recombination errors; for the most part however, if an individual has n copies of a given repeat at a particular location, those n copies were inherited from a parent, and the same number will be passed on to a child.

STRs offer a convenient way to differentiate between individuals through markers that are, biologically speaking, essentially meaningless. At a given STR locus,¹³ one individual may have five copies of the repeated unit, whereas others might have three, four, or six copies. Because many thousands of STRs exist in the human genome, many with this sort of limited, integer-based allelic variation, it is possible to genetically ‘fingerprint’ any individual solely by characterizing the number of repeats at their various STR loci.

To uniquely pinpoint an individual, however, it is not necessary to examine all STR loci in the genome, or anywhere near that number. In the UK, forensic DNA labs characterize just 10 STR loci, and in the US, the FBI uses a set of 13. This paper will argue that these numbers are too low, but for now the important point is that even a small set of loci carries tremendous resolving power, and indeed has accounted for all forensic DNA identification up to this point.

Law enforcement agencies began identifying suspects via DNA forensics considerably before the announcement of the draft human genome sequence in 2001. This is because the process for amplifying and characterizing STRs is quite straightforward, has been common practice for several decades, does not depend upon knowing the full genome sequence, and in fact involves no DNA sequencing at all.

¹² For instance, ATCGATCGATCG is simply the tetranucleotide sequence ATCG repeated three times.

¹³ Locus (plural: loci) commonly refers to a specific span or region of DNA sequence.

Once an STR is selected, differentiating individuals based upon their number of repeats at that site amounts to taking DNA samples from each person concerned, amplifying the repeated region via PCR,¹⁴ then assessing its length in each sample via electrophoresis.¹⁵ The result will be a number, corresponding to the number of nucleotides present in the repeated region. For instance, if a given person has 6 copies of the repeat ATGG at a certain locus, then the total length of the repeated region at that locus will be 24 nucleotides (6x4); an individual with 7 copies would return a value of 28, and so on.

This length assessment is carried out at each STR locus, of which each individual has two distinct alleles, one on each chromosome.¹⁶ The resulting collection of numbers – two integer numbers per STR, times 13 STRs – is that person's genetic fingerprint.

As should by now be evident, this fingerprint has little, if any, biological significance. It is simply a collection of numbers characterizing the lengths of a handful of repeated sequences arbitrarily selected to serve as markers. It is not dependent upon, and indeed ignores entirely, the vast majority of information in the genome. This is an important point, since although every genetic sample contains a colossal amount of information – including phenotypic characteristics, disease propensities, and everything else necessary to blueprint an entire human being – the system of DNA forensics as it currently exists rests entirely upon the selection and use of a few STR markers.

¹⁴ Polymerase Chain Reaction. This technique, which revolutionized molecular biology in the early 1980s, involves amplifying a specific region of DNA by several million-fold through thermal cycling and the use of short molecules of DNA called primers. Today, the technique is readily performed by machines in the lab, and allows quick and easy amplification of STRs, along with many other regions of interest.

¹⁵ Electrophoresis is a well-established laboratory technique that separates charged molecules (such as DNA or denatured proteins) by size and shape, by drawing them through a gel matrix with an electric charge. Linear DNA molecules of various lengths migrate at known rates in an electric field, and as such, after staining the gel to reveal how far each band of DNA has travelled, the nucleotide length of a DNA sample can be discerned without the need for DNA sequencing.

¹⁶ Recall that each individual carries two copies of each autosomal chromosome, and two sex chromosomes. If a given STR locus is located on chromosome 8, for instance, then the individual will have two loci to examine: one on each copy of chromosome 8. These two alleles need not be, and often are not, the same.

Inheritance

In each normal somatic cell, a person carries 46 chromosomes: 2 copies of each of the 22 autosomal chromosomes, and 2 copies of the sex chromosome. For each chromosome, one copy from was inherited from the mother and the other from the father. Thus, in each chromosome pair, they have one maternal chromosome and one paternal chromosome.¹⁷

The same principles that Gregor Mendel elucidated through his study of pea plants¹⁸ also apply to human reproduction. Namely, human alleles display segregation (each sex cell receives one chromosome from each pair) and independent assortment (which of the two chromosomes a given gamete receives – i.e., whether it is maternal or paternal in origin – is random).

These principles can be applied to STR alleles. Barring replication errors or other mutations, which occur at a low rate, parental STRs will be passed on to children via this mechanism of independent assortment.

Consider the example of a particular STR locus on a particular autosomal chromosome. The father has 2 copies of this chromosome, as does the mother. If the father has 4 and 6 repeats at this STR, and the mother has 3 and 3, what might their child carry? From the father, the child will receive a single chromosome carrying either 4 or 6 repeats; from the mother, the child will necessarily receive a single chromosome carrying 3 repeats. Thus, 50% of the children should have 4 and 3, and 50% should have 6 and 3.

¹⁷ To illustrate this, consider the sex chromosomes. Because women are XX, all of a woman's egg cells should carry one X chromosome: When partitioning the chromosomes, the mother has only X to give. The sex of the child, then, is up to the father (XY); half of his sperm should carry X and the other half should carry Y. Women received an X from their mother and an X from their father; men received an X from their mother and a Y from their father.

¹⁸ Crossing and breeding pea plants in the mid-19th century, Mendel showed that certain traits (such as flower color or plant height) are passed on as discrete entities, and do not appear blended in offspring. See Mendel, J.G. 1866. *Versuche ueber Pflanzenhybriden. Verhandlungen des naturforschenden Vereines in Bruenn* 4 ABHANDLUNGEN, 3–47. Cited by Robert C. Olby (1997) on <http://www.mendelweb.org/MWolby.html> (last visited Mar. 16, 2009).

If an impostor child arrives who is found to have 1 and 3 repeats at this locus, we can be reasonably certain that he arose from a different father.

Adding additional STRs would rapidly complicate this hypothetical. Instead, simply note that loci in close proximity to one another (adjacent on the same chromosome) are more likely to be passed on together, whereas those far apart or on different chromosomes are passed on essentially independent of one another.

These principles of segregation and assortment will become important in the discussion of familial searching in Section IV.

Summary: The Genome, from a Forensic Standpoint

In general, the human genome is a vast chemical information repository, which scientists have chosen to represent with an equally vast character string composed of just four different letters. The fanfare surrounding the 2001 release of the draft human genome sequence centered on the fact that researchers had uncovered the common genetic blueprint of all humans. However, to a forensic DNA scientist wishing to differentiate between individuals, our shared DNA blueprint is of very little use. Instead, forensic analysts pay attention to the small fraction of genomic sequence at which individual humans consistently differ. Within this set, they focus on short repeated sequences with no apparent biological or medical meaning. Having settled on a group of just 13 STRs, technicians measure the length of each repeated region, and proceed to identify individual humans by just 26 integer numbers – corresponding to the length of their repeated sequences, on both chromosomal copies, at 13 STR sites.

II. ORIGIN, CONTENT AND FUNCTION OF DNA DATABASES

American forensic DNA databases: Origin and current organization

The seeds for forensic DNA identification were sown in 1980, when American geneticists Ray White and Arlene Wyman identified a variable-length polymorphism in humans, called an RFLP.¹⁹ True forensic DNA typing arrived in 1985, when UK scientist Sir Alec Jeffreys first proposed the use of such polymorphisms to identify individuals in criminal cases.²⁰ This technique was soon adopted in America, and by 1988 it had already appeared in a Florida appellate case.²¹ The adoption of PCR techniques enabled swift and accurate characterization of newly-discovered STRs.²²

Although DNA typing began as a means to perform an identity test on a pair of human genetic samples – normally, one taken from a crime scene and the other from a suspected offender – law enforcement agencies soon became interested in storing profiles in a computer database for later comparison. In the U.K., the Forensic Science Service started the National DNA Database (NDNAD) in 1995.²³ Its American counterpart, the

¹⁹ Restriction Fragment Length Polymorphism, which means that DNA cleavage by restriction enzymes – at sites flanking what would today be called an STR – yielded fragments of variable length in different individuals. See Arlene R. Wyman & Ray White, *A Highly Polymorphic Locus in Human DNA*, 77 PROC. NAT'L ACAD. SCI., 6754 (1980).

²⁰ See Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence*, 95 CAL. L. REV. 721, 731 (2007); National Institutes of Standards and Technology (NIST), Chemical Science and Technology Laboratory, *Brief History of Forensic DNA Typing*, available at <http://www.cstl.nist.gov/strbase/ppt/intro.pdf> (last accessed July 5, 2009). See also James Sturcke, *DNA Pioneer Alec Jeffreys: Drop Innocent From Database*, THE GUARDIAN (London), Apr. 15 2009; Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 BROOK. L. REV. 13, 40 (2001).

²¹ Murphy, *supra* note 20, at 731 (citing *Andrews v. State*, 533 So. 2d 841 (Fla. Dist. Ct. App. 1988)).

²² For a brief discussion of PCR, see *supra* note 14.

²³ Forensic Science Service (UK), *The National DNA Database (NDNAD) Fact Sheet* (Feb. 2006), available at <http://www.forensic.gov.uk/pdf/company/foi/publication-scheme/communications/NDNAD.pdf> (last accessed July 7, 2009).

FBI Combined DNA Index System (CODIS) database,²⁴ was authorized by Congress in 1994²⁵ and officially launched in 1998.²⁶ It has expanded rapidly since.²⁷

The FBI CODIS database is a central repository containing DNA profiles (numeric STR length records) for individuals. It consists of three tiers of databases: the National DNA Index System (NDIS) maintained by the FBI, a collection of State DNA Index Systems (SDIS) maintained by each state, and the Local DNA Index Systems (LDIS) administered by local police departments or sheriffs' offices.²⁸ Federal, state and local member labs can search NDIS directly;²⁹ in addition, CODIS conducts an automated weekly search on all DNA profiles in NDIS.³⁰

After a DNA sample is collected, it is sent to the FBI, which analyzes the sample and includes the resulting profile in CODIS.³¹ Local and state agencies can also process their own samples, and upload DNA profiles to LDIS and SDIS databases according to their own laws and regulations; however, certain parameters must be met in order to upload locally computed profiles to NDIS.³² This is important, because NDIS relies to a large extent on state submissions for DNA samples from convicted felons: of the more than 1 million felonies committed in the U.S. each year, less than 10% of these are

²⁴ CODIS actually refers to the software the FBI uses to administer the database; however, it has come to be used as a shorthand to refer to the database itself. See Murphy, *supra* note 20, at 739 n74.

²⁵ 42 U.S.C. § 14131 *et seq.*

²⁶ National Institutes of Standards and Technology (NIST), *supra* note 20.

²⁷ In 1998, CODIS included only nine states. Federal offenders were added in 2000, beginning with violent felons, then all felons, and more recently all felony arrestees (see Section IV, *infra*). CODIS today includes 178 labs, and its software is used in all 50 states and labs in 30 foreign nations. See Jeffrey Rosen, *Genetic Surveillance for All*, SLATE, (Mar. 17, 2009), available at <http://www.slate.com/id/2213958> (last accessed July 9, 2009).

²⁸ DNA.gov: Levels of the Database, <http://www.dna.gov/dna-databases/levels/> (last accessed July 2, 2009) (describing the local, state and national levels of DNA databases in the US).

²⁹ Request for specific searches go to the custodian of the national DNA database. See Rosen, *supra* note 27.

³⁰ DNA.gov: Combined DNA Index System (CODIS), <http://www.dna.gov/dna-databases/codis/> (last accessed July 12, 2009). See also Rosen, *supra* note 27 ("Every Monday at 9 a.m., the national database automatically conducts two searches, looking for matches between the DNA of convicted offenders and the DNA at crime scenes. The automatic search also compares all crime-scene DNA samples with one another in search of serial criminals.").

³¹ 42 U.S.C. § 14135a(b).

³² For instance, a forensic profile must have a usable result at no fewer than 10 CODIS loci. *Id.*

federal.³³ CODIS stores state-contributed DNA profiles anonymously – that is, names and other identifying information corresponding to each profile are kept at the state level.

CODIS contains two basic types of profiles: (1) offender profiles, which are samples taken from convicted offenders (and other individuals, as will be discussed in Section IV below); and (2) forensic profiles, drawn from crime-scene samples (blood, semen, saliva and other biological material containing useful DNA).³⁴ A match between forensic profiles suggests that a common offender has committed crimes at different scenes, whereas a match between a forensic sample and an offender sample may suggest a suspect for a previously unsolved crime.³⁵

As of this writing, the CODIS database contains over 7 million offender profiles and more than 260,000 forensic profiles, making it the largest such database in the world.³⁶

An Odds Game: The 13 CODIS markers

A complete DNA profile³⁷ contains 26 numerical scores,³⁸ corresponding to the nucleotide lengths at 26 alleles of the 13 different STR loci on both chromosomal copies.³⁹

³³ Henry T. Greely, Daniel P. Riordan, *et al.*, *Family Ties: The Use of DNA Offender Databases to Catch Offenders' Kin*, 34 J.L. MED. & ETHICS 248, 251 (2006).

³⁴ See Murphy, *supra* note 20, at 738; DNA.gov: Combined DNA Index System (CODIS), <http://www.dna.gov/dna-databases/codis/> (last accessed July 12, 2009). There is something of an art to DNA recovery from crime scene samples, a topic that unfortunately is beyond the scope of this paper. See, e.g., John W. Bond & Christine Hammond, *The value of DNA material recovered from crime scenes*, 53 J. FORENSIC SCI. 797 (2008) (examining the use of NDNAD for crime-solving, and presenting the rates of recovery of usable DNA from various in situ sources, such as cigarette ends, blood and chewing gum. The authors' main conclusion is that, for all DNA sources, the conversion of DNA recovered from the scene to a useful NDNAD profile is significantly dependent upon the accreditation level of the crime scene examiner performing the recovery.).

³⁵ See Murphy, *supra* note 20, at 738.

³⁶ Federal Bureau of Investigation (FBI), CODIS Statistics Clickable Map, <http://www.fbi.gov/hq/lab/codis/clickmap.htm> (last accessed July 13, 2009). The database has also grown rapidly, roughly doubling every two years: In November 2005, it contained 2.5 million offender and 125,000 forensic profiles. See Greely, *supra* note 33, at 251 n.16.

The FBI CODIS markers are a set of 13 STRs spread across 12 of the 22 autosomal chromosomes. All 13 markers were specifically chosen from stretches of so-called “junk DNA,” non-coding DNA not thought to be “associated with any known physical or medical characteristics.”⁴⁰

Importantly, all thirteen CODIS loci are not created equal: an individual is more likely to match a stranger at some CODIS loci than others, and even at a single locus, some alleles are more common than others.

At each STR locus, only a finite number of alleles exist at an appreciable frequency within the population: this number of alleles ranges from 6 to 21, depending on the locus.⁴¹ A locus with only 6 possible alleles means that two individuals are much more likely to match randomly at that locus – that is, to share the same number of repeats by chance – than they are to match at a locus with 21 possible alleles.

Moreover, even within the set of possible alleles at a given locus, each allele is not equally common within the population. Thus, it may be much more common to see

³⁷ Unlike offender and arrestee samples, some forensic samples are taken from degraded DNA found at crime scenes and consequently may contain useful data from only a subset of the 13 STRs.

³⁸ “Each CODIS profile contains 52 characters representing 13 genetic locations, with two results per location and two digits for each result.” Rosen, *supra* note 27.

³⁹ As described in Section I, these lengths vary according to the number of repeats present at each STR, and are normally assessed via PCR-based amplification and electrophoretic separation.

⁴⁰ *U.S. v. Kincade*, 379 F.3d 813 at 818 (9th Cir. 2004) (en banc) (citing H.R.Rep. No. 106-900(I) at *27). See also Solomon Moore, *In a Lab, an Ever-Growing Database of DNA Profiles*, THE NEW YORK TIMES (May 11, 2009) (“By collecting such “junk DNA,” [as the CODIS markers,] which have no known genetic characteristics, police agencies can argue that they are not violating privacy rights. “There is nothing that would reveal any medical or physical characteristics at all,” said Jennifer C. Luttman, who runs the offender-DNA collection program.”).

⁴¹ See Bruce Budowle, F. Samuel Baechtel & Ranajit Chakraborty, *Partial matches in heterogeneous offender databases do not call into question the validity of random match probability calculations*, 123 INT. J. LEGAL MED. 59, 62 (2009) (Table 1, recounting the number of alleles that occur with a frequency greater than 0.01); Bruce Budowle, B. Shea, et al., *CODIS STR loci data from 41 sample populations*, 46 J. FORENSIC SCI. 453 (2001). Professor Greely refers to a range of 7-23 alleles at STR loci, but does not reference this figure. See Greely, *supra* note 33 at 250. In actuality, because the human population continues to reproduce and our DNA accrues new mutations, there is no single definitive answer to the total number of alleles at any STR locus; a new allele can arise in the population at any time. Hence, Budowle’s approach – including only those alleles that have achieved some prevalence in the population – seems the most sensible, and yields the figures reported here.

five repeats at a given locus than it is to see three, or eight.⁴² (If that were so, an individual with five repeats would match many more people at that locus by chance, simply by virtue of carrying a more common allele.)

To complicate things further, the CODIS loci and their alleles mutate at different rates. Initially, a single standard mutation rate⁴³ was assumed to apply all STR loci, but more recently mutation rates have been shown to vary between loci.⁴⁴ Also, it has now become evident that even within a single STR locus, certain alleles are more prone to mutation than others.⁴⁵

Despite the limited number of alleles observed at each STR, combinations of all 13 loci quickly become complex: two randomly selected Americans are likely to share no more than 2 or 3 alleles.⁴⁶ The chance of two unrelated people matching both alleles at all 13 loci is infinitesimal, the odds usually estimated as upwards of one in several billion.⁴⁷ Identical twins, however, will share both alleles all 13 loci, and first-degree relatives will share at least half.⁴⁸ (This is the foundation of familial DNA searching, discussed in detail below.)⁴⁹

⁴² See, e.g., Thomas M. Reid, Michael L. Baird *et al.*, *Use of sibling pairs to determine the familial searching efficiency of forensic databases*, 2 *Forensic Sci. Int'l: Genetics* 340, 340 (2008) (discussing the 'kinship matching' approach, which takes into account the population frequency of matching alleles to compute relatedness).

⁴³ Here, mutation rate refers to the rate of spontaneous change in repeat number at a given STR locus. This change occurs between generations. The standard mutation rate (μ) was assumed to be ~ 0.001 , or one in one thousand. That is, it was assumed that 1/1000 children would inherit through mutation a different number of repeats than either parent had at any one locus. The new number of repeats was usually ± 1 repeat; spontaneous insertions or deletions of more than one repeat are less common. Note that in most cases, gain or loss of a single repeat through mutation will yield an allele already present in the population – just not present in the parents.

⁴⁴ Robert E. Wenk, *Testing for parentage and kinship*, 11 *Curr. Opin. Hematology* 357, 359 (2004).

⁴⁵ See, e.g., Michael Edwards & Robert W. Allen, *Characteristics of mutations at the D5S818 locus studied with a tightly linked marker*, 44 *Transfusion* 83 (2004).

⁴⁶ Greely, *supra* note 33 at 250.

⁴⁷ Greely, *supra* note 33 at 250 (giving the odds as "one in several hundred billion"); Department of Energy, Human Genome Project Information: DNA Forensics, at http://www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml (last accessed July 16, 2009) (giving the odds as "one in a billion").

⁴⁸ Parent-child matches will share at least 13/26 alleles, and at least one allele per locus, a rare distribution pattern; on average, they will share 15.7 alleles. Siblings can theoretically share anywhere from 0-26 alleles, but on average will share 16.7. 13 of these come from a 50% chance of each sib inheriting the

Statistically, CODIS regularly returns highly significant matches using these 13 loci. However, it is worthwhile to note that CODIS does not use its 13 loci to their full potential.

First, by matching in a binary fashion (simply returning a ‘yes’ or ‘no’ at each locus), CODIS squanders valuable population genetics data. The system does not currently take into account the relative frequency of each allele in the population – thus, a hit based on two profiles matching at the most common allele is treated the same as a hit based on two profiles matching at a very rare allele, even though in the latter case, the two samples are more likely to come from the same source.⁵⁰

Second, CODIS does not report allelic distribution patterns. Comparing DNA profiles from parents and child, one of each pair of alleles will match – that is, at least 13 of 26 alleles will match, with a distinctive pattern: matching at least one of every pair. The odds of matching 13/26 alleles in general is roughly 3%; however, the odds of matching 13/26 alleles in this particular pattern is less than 1 in 2000.⁵¹ Hence, parents and children share a particular inheritance pattern than CODIS could highlight – but was designed to ignore.

Chromosomal linkage is another useful factor not exploited by CODIS. Two of the 13 CODIS loci are located on chromosome 5,⁵² and each individual has two separate copies of chromosome 5. If two DNA profiles match, CODIS is not set up to report

same allele from a parent; additional alleles come from 1) parents possibly having 2 copies of an allele, or 2) parents sharing an allele. For an average pair of Caucasian siblings at 13 STR loci, there is 1 locus with no shared alleles, 7 loci with one shared allele, and 5 loci with both alleles shared. See Greely, *supra* note 33 at 250, 252; Section I *supra* (discussion of inheritance patterns).

⁴⁹ See Section IV, *infra*.

⁵⁰ To be fair, this is presumably because when CODIS was launched, allelic population distributions were not well known. Of course, even currently accepted allelic distributions are taken from the database as it now stands; these distributions could shift dramatically if suddenly all Americans were profiled and added to the database.

⁵¹ See Greely, *supra* note 33 at 252.

⁵² These loci are D5S818 and CSF1PO. See National Institutes of Standards and Technology, *supra* note 20.

whether the same two alleles co-localize to the same physical chromosome in each sample.⁵³

These shortages aside, the FBI has used 13 loci with success to identify individuals from crime scene samples through pairwise matches, and at very high levels of statistical significance.

III. DNA DATABASES: LEGAL FOUNDATIONS

This Section sets forth the legal landscape in which CODIS operates, by summarizing both the statutory authorization for DNA collection and databasing, and its approval by the courts on Fourth Amendment grounds.

Statutory Authorization

The DNA Identification Act of 1994 created CODIS and provided funding to law enforcement for DNA collection,⁵⁴ thus setting the stage for DNA collection and analysis on a national scale. The DNA Analysis Backlog Elimination Act of 2000 limited the compulsory DNA collection to those convicted of a “qualifying” federal offense,⁵⁵ which included selected felonies such as murder, kidnapping and some sex crimes.⁵⁶ However, this list of “qualifying” offenses has since then steadily grown. The Patriot Act of 2001

⁵³ This analysis is arguably of limited use when the odds of a match at all 13 loci are so long. However, this analysis would be highly valuable for partial-match or ‘familial’ searching, because parents and children share different allelic conservation patterns than siblings – and neither group normally matches at all 13 loci. Of course, at increasing distances between markers on the same chromosome, normal DNA recombination during sexual reproduction effectively de-links even such ‘linked’ markers.

⁵⁴ P.L. 103-322, 108 Stat. 2065 (1994) (codified at 42 U.S.C. §§ 14131-14134).

⁵⁵ 42 U.S.C. §14135a(a)(1)(B).

⁵⁶ P.L. 106-546, §3, 114 Stat. 2726, 2729-30 (2000).

added terrorism-related crimes to this category,⁵⁷ and the Justice for All Act of 2004 further added all violent crimes, all sexual abuse crimes and all felonies.⁵⁸

The most substantial regulatory shift came earlier this year. The DNA Fingerprinting Act of 2005 authorized the Attorney General, acting through Department of Justice (DOJ) regulations, to collect DNA from arrestees and detainees, including non-U.S. persons.⁵⁹ This policy was implemented by the DOJ in December 2008, and took effect in January 2009.⁶⁰ This new policy will be discussed further in Section IV, below.

The CODIS statutory framework also sets forth criminal and financial penalties for the misuse or sharing of DNA samples and data,⁶¹ as well as restrictions on sample retention. As it now stands, CODIS must expunge the DNA profile of anyone whose conviction is overturned, or who was arrested and whose charges were later dismissed.⁶² As a condition of access to the national database, states must also expunge such records.⁶³ However, these provisions do not address retention of DNA profiles from those who have successfully completed their sentences.⁶⁴ As an interesting side note, the U.K. has been much more aggressive in retaining profiles in its own national DNA database, although its policies must soon change in this regard.⁶⁵

⁵⁷ P.L. 107-56, §503, 115 Stat. 272, 364 (2001).

⁵⁸ P.L. 108-405, §203(b), 118 Stat. 2260, 2270 (2004) (codified at 42 U.S.C. §14135a(a)(2)).

⁵⁹ Tit. X, P.L. 109-162, 119 Stat. 2960 (codified at 42 U.S.C. §14135a(a)(1)); The Adam Walsh Child Protection and Safety Act of 2006, Section 155, also extended the scope of testing to all arrestees. PL 109-248.

⁶⁰ Department of Justice, DNA Sample Collection and Biological Evidence Preservation in the Federal Jurisdiction, 73 FEDERAL REGISTER 74,932 at 74,935 (2008).

⁶¹ “A person who knowingly discloses a sample or result described in subsection (a) of this section in any manner to any person not authorized to receive it, or obtains or uses, without authorization, such sample or result, shall be fined not more than \$250,000, or imprisoned for a period of not more than one year. Each instance of disclosure, obtaining, or use shall constitute a separate offense under this subsection.” 42 U.S.C. § 14135e(c), as amended by the 2004 Justice for All Act, H.R.5107.

⁶² 42 U.S.C. § 14132(d)(1)(A); *see also* Federal Bureau of Investigation, Expungement of DNA Records, <http://www.fbi.gov/hq/lab/html/expungement.htm> (last visited May 16, 2009).

⁶³ 42 U.S.C. § 14132(d)(2)(A).

⁶⁴ *See Kriesel*, 508 F.3d at 952 (“once they have [an individual’s] DNA, police at any level of government with a general criminal investigative interest . . . can tap into that DNA without any consent, suspicion, or warrant, long after his period of supervised release ends.”).

⁶⁵ Britain has aggressively expanded its own DNA database, NDNAD. For example, NDNAD includes DNA from arrestees charged with even minor crimes, and its policies provide no means of expungement,

DNA Collection is a Fourth Amendment Search

The Fourth Amendment protects the rights of the people to be secure against “unreasonable searches and seizures.”⁶⁶ A seizure threatens property while a search threatens personal privacy.⁶⁷ Thus, to challenge a government action under the Fourth Amendment, it is necessary to determine whether the accused action constitutes a search or seizure, and if so, whether that search or seizure is “reasonable.”

It is “settled law” that compulsory DNA collection and indexing constitutes a search under the Fourth Amendment.⁶⁸ (Although it could also be viewed as a seizure, courts have generally shied away from this characterization.⁶⁹)

even in the case of acquittal. As a result, as of March 2008 “857,000 people in the British database, or about one-fifth, [had] no current criminal record.” Solomon Moore, *F.B.I. and States vastly expand DNA databases*, THE NEW YORK TIMES, (April 18, 2009). In December 2008 the European Court of Human Rights in Strasbourg unanimously decided that the indefinite retention of DNA samples and profiles, even when people are not charged or found not guilty of an alleged crime, breached Article 8 of the European Convention on Human Rights. *S. and Marper v. The United Kingdom*, 30562/04 [2008] ECHR 1581 (4 December 2008). As of this writing the U.K. has yet to enact regulatory changes in response to this ruling. This has led to enhanced criticism from commentators, including Sir Alec Jeffreys, the founder of forensic DNA analysis. See James Sturcke, *DNA Pioneer Alec Jeffreys: Drop Innocent From Database*, THE GUARDIAN (London) (Apr. 15 2009).

⁶⁶ U.S. Const. Amend IV. (“The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.”).

⁶⁷ See, e.g., *Texas v. Brown*, 460 U.S. 730, 747 (1983) (“The [Fourth] Amendment protects two different interests of the citizen - the interest in retaining possession of property and the interest in maintaining personal privacy. A seizure threatens the former, a search the latter.”)

⁶⁸ See, e.g., *U.S. v. Amerson*, 483 F.3d 73, 77 (2d. Cir. 2007) (“It is settled law that DNA indexing statutes, because they authorize both a physical intrusion to obtain a tissue sample and a chemical analysis to obtain private physiological information about a person, are subject to the strictures of the Fourth Amendment. The extraction and analysis of plaintiffs’ blood for DNA-indexing purposes constitutes a search”) (citations omitted, internal quotation marks omitted).

⁶⁹ See *U.S. v. Kincade*, 379 F.3d 813, 821 n.15 (9th. Cir 2004) (en banc) (“The compulsory extraction of blood for DNA profiling unquestionably implicates the right to personal security embodied in the Fourth Amendment, and thus constitutes a search within the meaning of the Constitution.” (citing *Skinner v. Ry. Labor Executives' Ass'n*, 489 U.S. 602, 616 (1987) (“We have long recognized that a compelled intrusion into the body for blood to be analyzed for alcohol content must be deemed a Fourth Amendment search.”)); cf. *Kincade*, 379 F.3d at 873, (“[I]t is important to recognize that the Fourth Amendment intrusion here is not primarily the taking of the blood, but seizure of the DNA fingerprint and its inclusion in a searchable database.”) (Kozinsky, J., dissenting).

Thus, in order for DNA collection to be constitutionally permissible, it must be a reasonable search; and “the reasonableness of a search is determined by assessing, on the one hand, the degree to which it intrudes upon an individual's privacy and, on the other, the degree to which it is needed for the promotion of legitimate governmental interests.”⁷⁰

Standards for evaluating reasonableness differ, however, according to the circumstances of the search. The most stringent, the probable cause standard, applies to traditional law enforcement activities such as arrests or searches of residences, and must generally be provided in the form of a warrant.⁷¹ A somewhat lower standard, reasonable suspicion, governs Terry stops⁷² and other less intrusive searches.⁷³ Finally, in so-called “special needs” or “suspicionless” searches, courts will assess reasonableness by weighing the “totality of the circumstances,” the lowest standard; this achieved by a general balancing test “assessing, on the one hand, the degree to which [the search] intrudes upon an individual’s privacy and, on the other, degree to which it is needed for the promotion of legitimate governmental interests.”⁷⁴ This test is also called the “general reasonableness” test.

Importantly, the Supreme Court has applied this lowest-stringency Fourth Amendment analysis to cases involving convicts, and recently to individuals serving post-conviction punishment such as probation⁷⁵ or parole,⁷⁶ without finding any special

⁷⁰ *U.S. v. Knights*, 534 U.S. 112, 118-19 (2001) (citation omitted, internal quotation marks omitted).

⁷¹ *Payton v. New York*, 445 U.S. 573, 586 (1980) (“It is a basic principle of Fourth Amendment law that searches and seizures inside a home without a warrant are presumptively unreasonable.”) (Internal quotation marks omitted).

⁷² See *Terry v. Ohio*, 392 U.S. 1 (1968) (holding that a seemingly random pat-down for weapons by police was permissible under the Fourth Amendment if grounded in “specific reasonable inferences.”).

⁷³ See *Alabama v. White*, 496 U.S. 325, 330 (1990) (stating that “[r]easonable suspicion is a less demanding standard than probable cause”).

⁷⁴ *Knights*, 534 U.S. at 118-19. Searches of this type can occur where justified by a routine administrative purpose, or where a “special need” beyond the need for law enforcement makes the more stringent analyses impracticable. See *Griffin v. Wisconsin*, 483 U.S. 868, 873 (1986).

⁷⁵ *Knights*, 534 U.S. 112.

⁷⁶ *Samson v. California*, 547 U.S. 843 (2006).

needs justifying this diminished standard. Thus, offenders are afforded a diminished expectation of privacy that allows circumvention of the special needs test, and also cuts in the government's favor in the general reasonableness test.

Virtually all courts that have heard the issue have upheld compulsory DNA collection against Fourth Amendment challenges. The Supreme Court has yet to hear such a case, but Circuit courts have unanimously upheld the 2004 version of the DNA collection law – that is, the version authorizing collection from convicts.⁷⁷ In the majority of these cases, courts have directly applied a general reasonableness or “totality of the circumstances” analysis without first establishing that a special needs situation applied.⁷⁸ By contrast, the Second and Seventh Circuits have to date relied upon the “special needs” test in upholding DNA collection.⁷⁹ This distinction between special needs and general reasonableness may be moot, however: The Sixth Circuit has suggested that DNA collection from parolees and those on probation would still be upheld regardless which test is applied.⁸⁰

Nonetheless, the Fourth Amendment remains the main grounds by which DNA collection is contested; for instance, courts have roundly rejected Fifth Amendment substantive due process challenges to the constitutionality of DNA collection.⁸¹

⁷⁷ *U.S. v. Weikert*, 504 F.3d 1 (1st Cir. 2007); *U.S. v. Amerson*, 483 F.3d 73 (2d Cir. 2007); *U.S. v. Kriesel*, 508 F.3d 941 (9th Cir. 2007); *U.S. v. Banks*, 490 F.3d 1178 (10th Cir. 2007); *Wilson v. Collins*, 517 F.3d 421 (6th Cir. 2006); *U.S. v. Hook*, 471 F.3d 766 (7th Cir. 2006); *U.S. v. Kraklio*, 451 F.3d 922 (8th Cir. 2006); *U.S. v. Castillo-Lagos*, 147 Fed. App'x. 71 (11th Cir. 2005).

⁷⁸ See *Weikert*, 504 F.3d at 3 (“We interpret the Supreme Court’s decision in *Samson v. California* to require that we join the majority of the circuits in applying a totality of the circumstances approach to the issues in this case, rather than the special needs analysis”) (Citation omitted, internal quotation marks omitted); *Kriesel*, 508 F.3d 941; *Kincade*, 379 F.3d at 839-840.

⁷⁹ *Amerson*, 483 F.3d at 73, 78-79 (where the Second Circuit in 2007 read *Samson* narrowly and declined to directly apply the general balancing test to a case involving DNA collection from people on probation); *U.S. v. Hook*, 471 F.3d 766, 772-74 (7th Cir. 2006).

⁸⁰ See *Wilson*, 517 F.3d at 427, n. 4 (“Even if we were to apply the more stringent special-needs test, there is no reason to believe the ultimate result would be different.”). See also *U.S. v. Conley*, 453 F.3d 674, 677-81 (6th Cir. 2006).

⁸¹ See *Rise v. Oregon*, 59 F.3d 1556, 1563 (9th Cir. 1995) (holding that the extraction of blood from an individual in a medically acceptable manner – despite the individual’s lack of opportunity to object – does not implicate the Due Process Clause.); see also, *U.S. v. Hugs*, 384 F.3d 762, 768-69 (9th Cir. 2004) (DNA testing as a condition of supervised release is not unconstitutionally vague and so meets

Similarity to Fingerprinting

An obvious comparison suggests itself, between DNA collection and fingerprinting. Taking traditional fingerprints is similar to DNA sampling in that both seek to establish identity by impartial means (one phenotypic, the other genotypic); both are minimally invasive,⁸² and both are routinely used by law enforcement to aid in identification.

This similarity has not escaped courts,⁸³ and has been used to justify DNA collection from a progressively broader base of individuals.⁸⁴ Highlighting this similarity in its May 2009 decision in *U.S. v. Pool*,⁸⁵ the U.S. District Court for the Eastern District of California quoted the following passage from a 1932 Second Circuit case:

“Finger printing seems to be no more than an extension of methods of identification long used in dealing with persons under arrest for real or supposed violations of the criminal laws. It is known to be a very certain means devised by modern science to reach the desired end, and has become especially important in a time when increased population and vast aggregations of people in urban centers

procedural due process requirements); *Johnson v. Quander*, 370 F.Supp.2d 79, 89-93 (Dist.DC. 2005) (DNA Act does not violate substantive or procedural due process under the Fifth Amendment), *aff'd*, 440 F.3d 489, 503 (D.C.Cir. 2006) (Fifth Amendment challenge to DNA Act is “without merit and do[es] not warrant separate discussion”).

⁸² DNA can be sampled by cheek swab or blood test, both of which have been held to be a minimal intrusion. *See Skinner v. Railway Labor Executives’ Ass’n*, 489 U.S. 602, 625 (1989) (blood tests are safe, common and “do not constitute an unduly extensive imposition on an individual’s privacy and bodily integrity.”). *See also U.S. v. Kriesel*, 508 F.3d 941, 948 (9th Cir. 2007) (“Consequently, the additional privacy implications of a blood test collecting DNA, as opposed to a cheek swab . . . do not significantly alter our analysis.”).

⁸³ *See, e.g., Banks v. U.S.*, 490 F.3d 1178, 1193 (10th Cir. 2007) (stating that the FBI CODIS database “operates much like an old-fashioned fingerprint database (albeit more efficiently.)”; *Nicholas v. Goord*, 430 F.3d 652, 671 (2nd Cir. 2005) (the intrusion of privacy effected by DNA sampling is similar to the intrusion and maintenance of fingerprint records).

⁸⁴ *See Jones v. Murray*, 962 F.2d 302, 307 (4th Cir. 1992) (“The governmental justification for this form of identification, therefore, relies on no argument different in kind from that traditionally advanced for taking fingerprints and photographs, but with additional force because of the potentially greater precision of DNA sampling and matching methods.”)

⁸⁵ *U.S. v. Pool*, CR S-09-0015 EJG GGH (E.D.Cal., May 27, 2009).

have rendered the notoriety of the individual in the community no longer a ready means of identification.”⁸⁶

Summary of Legal Standards

Compulsory DNA collection stands on firm legal footing, at least as originally envisioned and implemented: As a system to collect samples from convicted felons and, more recently, from those on various forms of supervised release. But as the next section will show, new expansion of the database to include arrestees recently faced new challenges. Moreover, technical advances, new search techniques and concerns over sample retention remain largely unexamined by courts, and could also curb law enforcement ability to recover DNA from an ever-wider swath of the population.

IV. NEW CHALLENGES: WILL THE LEGAL FOUNDATION HOLD?

This Section highlights several ways in which the forensic DNA database system has outgrown its original mandate or soon will, creating a set of problems that will require either radical statutory repair or a fresh judicial take on DNA collection under the Fourth Amendment.

Including Arrestee DNA

Among the most dramatic and immediate shifts in DNA database policy is the recent move to include those arrested for, but not yet convicted of, a crime.⁸⁷

⁸⁶ *U.S. v. Kelly*, 55 F.2d 67, 69 (2nd Cir. 1932).

⁸⁷ Arrestee profiles were originally not permitted in CODIS. See 42 USC § 14132 (2000). Now, however, states may upload any profile collected in a manner consistent with its own laws. See 42 USC § 14132 (2009), and the Federal government also indexes arrestees. See discussion below, in this Section.

California, which already presides over the largest state DNA database in the country – with profiles from over 1.2 million convicts – announced its inclusion of arrestee profiles beginning in January 2009, a move expected to double its annual database growth.⁸⁸

This year, the FBI has joined California and 15 other states⁸⁹ in collecting DNA samples from individuals awaiting trial.⁹⁰ The FBI expects that this move will sharply increase database growth, from 80,000 new entries added per year to 1.2 million new entries per year by 2012.⁹¹

The sudden influx of arrestees into federal DNA databases is a mixed blessing for forensic scientists. Database expansion is useful from a searchable-records standpoint: The more people that are profiled in the database, the less the odds that analysts must resort to exotic and legally dubious techniques such as familial searches to return a match.⁹² At the same time, however, such expansion oversteps the original mandate of DNA databases – to record genetic markers from convicted offenders, on the dual theories that 1) they are likely to reoffend, and 2) their diminished expectation of privacy legitimizes the search.

Arrestee DNA: Treatment in the Courts

State court response to this expansion has so far been mixed. In 2006, the Minnesota Court of Appeals struck down a state law that authorized DNA collection

⁸⁸ Jeffrey Rosen, *Genetic Surveillance for All*, SLATE, Mar. 17, 2009, available at <http://www.slate.com/id/2213958> (last accessed July 9, 2009).

⁸⁹ Nearly a third of all states now have laws authorizing arrestee DNA sampling. These are: Alaska, Arizona, California, Hawaii, Kansas, Louisiana, Maryland, Michigan, Minnesota, New Mexico, North Dakota, South Dakota, Tennessee, Texas, Vermont and Virginia, now have laws authorizing arrestee DNA sampling. See Denver District Attorney's Office, DNA Database Cases, http://www.denverda.org/DNA/DNA_Arrestee_Database_Cases.htm (last accessed July 15, 2009).

⁹⁰ Solomon Moore, *F.B.I. and States vastly expand DNA databases*, THE NEW YORK TIMES, (April 18, 2009). Available at: http://www.nytimes.com/2009/04/19/us/19DNA.html?_r=2&hp (last visited April 22, 2009).

⁹¹ *Id.*

⁹² See subsection discussing familial search, *infra*.

from arrestees on Fourth Amendment grounds.⁹³ However, in 2007 the Virginia Court of Appeals reached the opposite result, upholding DNA collection upon arrest on the theory that it is similar to fingerprinting.⁹⁴

Although Circuit courts have upheld DNA collection from convicts, parolees and individuals on probation, until summer 2009 no federal court had yet extended this approval to arrestees in a Fourth Amendment challenge. Indeed, in *U.S. v Kriesel* (2007), the Ninth Circuit explicitly stated that its approval of DNA collection in that case did not extend to arrestees.⁹⁵

On May 27, 2009, in a case of first impression for the federal courts, the U.S. District Court for the Eastern District of California applied the “totality of the circumstances” test to uphold mandatory DNA collection from individuals arrested upon probable cause for felony criminal charges.⁹⁶ In its decision, the court emphasized the importance of a finding of judicial probable cause,⁹⁷ and inventoried the various ways in which a defendant’s liberty is routinely restricted after such a finding and before trial.⁹⁸

⁹³ *In the Matter of the Welfare of C.T.L.*, 722 N.W.2d 484 (Ct. App. Minn. 2006) (Minnesota Court of Appeals invalidated a statute that authorized DNA collection after a judicial finding of probable cause that arrestee had committed a crime).

⁹⁴ *Anderson v. Commonwealth*, 274 Va. 469 (VA Sup. Ct., 2007) (holding that the taking of defendant's DNA sample upon arrest was analogous to fingerprinting upon arrest, was not an unlawful search under the Fourth Amendment, and that no additional finding of individualized suspicion was required to do so).

⁹⁵ *U.S. v. Kriesel*, 508 F.3d 941, 948-49 (9th Cir. 2007) (This 2007 Ninth Circuit decision emphasized “that our ruling today does not cover DNA collection from arrestees or non-citizens detained in the custody of the United States, who are required to submit to DNA collection by the 2006 version of the DNA Act”).

⁹⁶ *U.S. v. Pool*, CR S-09-0015 EJJ GGH (E.D.Cal., May 27, 2009) (hereinafter *U.S. v. Pool*). In its 20 page opinion, the court holds that after a judicial or grand jury determination of probable cause has been made for felony criminal charges against a defendant, no Fourth Amendment or other Constitutional violation is caused by a universal requirement that a charged defendant undergo a “swab test,” or blood test when necessary, for the purposes of DNA analysis to be used solely for criminal law enforcement identification purposes.

⁹⁷ *Id.* at *13 (The court’s holding “does not authorize DNA sampling after citation or arrest for infractions or misdemeanors, as in these cases there will be no *judicial* finding of probable cause soon after the arrest or citation, or no *grand jury* finding before or after the arrest. It does not authorize police officials to perform DNA sampling prior to a judicial finding of probable cause . . . [I]t is the finding of probable cause on criminal charges which allows the court to set release conditions similar to those of probation and parole, which is the underpinning of the court’s holding in this case.”) (emphasis in original, citation omitted).

⁹⁸ *Id.* at *7-8.

Moreover, the court noted that after such a finding, individuals have a “diminished expectation of privacy in [their] own identity,” and that DNA identification as a law enforcement tool is simply a “technological progression” from traditional fingerprints and photographs, both of which are “part of the routine booking process upon arrest.”⁹⁹ Given this, the court stated “the decision to impose the DNA testing requirement on pre-trial detainees or releasees seems clearly warranted, if not compelling.”¹⁰⁰

It is unclear how or whether federal appellate courts will treat this issue, and as before, the Supreme Court remains silent on the fair reach of DNA collection statutes and regulations.

Other Expansion Concerns

In addition to arrestees, in some cases innocents such as laboratory workers, victims of crime, or voluntary DNA donors may also be profiled and indexed in DNA databases, sometimes without their knowledge. Although CODIS does not solicit the inclusion of such records, a number of states and localities maintain ‘offline’ DNA databanks of samples taken from people never charged with a crime.¹⁰¹ These local databases “have literally no oversight and regulation and yet are pushing the boundaries farther than anyone could imagine.”¹⁰²

Because it is not authorized at the federal level, and instead represents what some commentators term “rogue databases,”¹⁰³ this form of DNA database expansion will not be dealt with further in this paper. However, it remains a meaningful problem, particularly for those personally affected – knowingly or otherwise. Moreover, it is clear

⁹⁹ *Id.* at *8 (citing *Napolitano v. U.S.*, 340 F.2d 313, 314 (1st Cir. 1963)).

¹⁰⁰ *Id.* at *8.

¹⁰¹ Ellen Nakashima, *From DNA of Family, a Tool to Make Arrests*, WASHINGTON POST (Apr. 21, 2008), at A01.

¹⁰² *Id.*

¹⁰³ *Id.*

that under current judicial reasoning, such inclusion at the very least cannot be supported in the same way as that of convicts, parolees or arrestees.

The following Subsection examines a de facto expansion of the database: the inclusion of family members via familial DNA searching.

Familial Search

Because of the special inheritance patterns that link siblings, parents, children and other close relatives, certain law enforcement agencies are expanding their use of DNA databases to scour for offenders' kin.

Familial DNA search has received national media coverage,¹⁰⁴ as well as recent treatment in law reviews.¹⁰⁵ This Subsection will analyze briefly this new use of existing DNA databases from a technical and legal standpoint.

Familial DNA searching is relatively simple to perform, and in principle involves only changing the threshold on a database search. Whereas an exact hit requires that all 26 alleles match at all 13 CODIS markers – signifying exact parity between two profiles – searches can also be performed at a lower stringency, reporting profiles that match at some fraction of the 26 alleles.¹⁰⁶ Near-matches may be useful in the case of degraded

¹⁰⁴ See, e.g., Jaelyn O'Malley, *Police Consider Familial DNA*, RENO GAZETTE-JOURNAL (Jan. 18, 2009); Maura Dolan & Jason Felch, *Tracing a Crime Suspect Through a Relative*, L.A. TIMES (Nov. 25, 2008); Maura Dolan & Jason Felch, *California Takes Lead on DNA Crime-Fighting Technique*, L.A. TIMES (Apr. 26, 2008); Richard Willing, *Suspects Get Snared by a Relative's DNA*, USA TODAY (June 8, 2005) at A1; CBS "60 Minutes", *A Not So Perfect Match: How Near-DNA Matches Can Incriminate Relatives of Criminals*, CBS News (Jul. 15 2007), available at <http://www.cbsnews.com/stories/2007/03/23/60minutes/main2600721.shtml> (last visited June 20, 2009).

¹⁰⁵ See Greely, *supra* note 33; Daniel J. Grimm, *The Demographics of Genetic Surveillance: Familial DNA Testing and the Hispanic Community*, 107 COLUM. L. REV. 1164 (2007); Lina A. Hogan, *Fourth Amendment – Guilt by Relation: If Your Brother Is Convicted of a Crime, You Too May Do Time*, 30 W. NEW ENG. L. REV. 543 (2008); Erica Haimes, *Social and ethical issues in the use of familial searching in forensic investigations: insights from family and kinship studies*, 34 J. L. MED. & ETHICS 263 (2006).

¹⁰⁶ For a general overview of this technique, see Frederick R. Bieber *et al.*, *Human Genetics: Finding Criminals Through DNA of Their Relatives*, 312 SCIENCE 1315 (2006); David R. Paoletti *et al.*, *Assessing the Implications for Close Relatives in the Event of Similar But Nonmatching DNA Profiles*, 46 JURIMETRICS J. 161 (2006); Eva Steinberger & Gary Sims, *Finding Criminals Through the DNA of their*

DNA, where some markers are missing or improperly recorded from one sample, although these searches are still targeted to match profiles from the same individual. However, because first-degree relatives match, on average, at half their alleles or more, a low stringency search with a crime-scene sample can sometimes return the offender's close-matching relatives in the database.

Such searches typically return many thousands of results.¹⁰⁷ But winnowing these down by geographical location or other means can provide law enforcement with new leads. According to U.K. statistics, about 10% of the time a partial match identified via familial DNA searching provides a useful investigative lead;¹⁰⁸ this partial match is later confirmed via an exact match to the suspect's own DNA.¹⁰⁹ As such, the chance of wrongful convictions is small.¹¹⁰ However, as a de facto (and silent) expansion of DNA databases, the technique does raise privacy concerns.

In the U.K., law enforcement agencies have been performing deliberate familial DNA searches since 2004, and have solved at least 18 cases in this way.¹¹¹ For example, the so-called "shoe rapist," who eluded authorities for decades, was ultimately identified via his sister's DNA obtained on a drunk-driving arrest;¹¹² in another case, a drunken man dropped a brick from an overpass, killing a truck driver; a partial-profile DNA

Relatives – Familial Searching of the California Offender DNA Database, 31 Prosecutor's Brief 28, 30 (2008).

¹⁰⁷ Recall that the odds of matching 13 of 26 alleles is 3% in the general population. *Supra* note 51 and related discussion. Matching 3% of the ~7 million records in CODIS today would return 210,000 matches.

¹⁰⁸ See Rosen, *supra* note 27.

¹⁰⁹ In males, a partial match can also be reinforced by Y-STR typing, which examines STR loci found on the Y chromosome.

¹¹⁰ Nakashima, *supra* note 100, quoting Prof. Henry Greely.

¹¹¹ See, e.g., Maura Dolan & Jason Felch, *Tracing a Crime Suspect Through a Relative*, L.A. TIMES (Nov. 25, 2008); Rosen, *supra* note 27.

¹¹² The shoe rapist attacked women in the U.K. and thereafter stole their high heels. The perpetrator's sister was arrested on DUI charges some 20 years after these crimes, and her DNA sample was automatically searched against cold case records. A close match was returned that led police to her brother. "When he was arrested, his DNA was a perfect match and police found more than 100 stiletto heels hidden under a trap door." See Nakashima, *supra* note 101.

search of the database turned up his brother and led to a confession.¹¹³ Familial DNA searching is just beginning to be used in the United States: In Kansas, the BTK Killer was snared in part by his daughter's DNA, recovered from a pap smear sample at her college;¹¹⁴ in Louisiana, a rapist was caught by a chance partial DNA match with another rapist's *victim* in another county, who turned out to be his own sister.¹¹⁵ While neither U.S. case involved an open-ended database search to find the relative in question, both nonetheless highlighted the investigative value of familial DNA, and combined with U.K. success, sparked investigative interest stateside.

Denver District Attorney Mitchell R. Morrissey is a particularly active champion of familial DNA searching in the U.S, and has been instrumental in obtaining approval for the technique.¹¹⁶ Pursuing a Colorado rapist in 2005, Morrissey discovered three close matches in CODIS; each located in a different state.¹¹⁷ Because the FBI administers CODIS to keep identifying information with the contributing states, the D.A. was unable to match names to the profiles. Frustrated, he contacted CODIS administrators, who refused to authorize familial search in general, on the theory that it might cause public

¹¹³ This case received a good deal of media coverage. *See, e.g.*, Matthew Falloon, *DNA Traps Brick Thrower Who Killed Lorry Driver*, THE GUARDIAN (London), (Apr. 20, 2004); Shaoni Bhattacharya, *Killer convicted thanks to relative's DNA*, NEW SCIENTIST (Apr. 20, 2004), available at <http://www.newscientist.com/article/dn4908-killer-convicted-thanks-to-relatives-dna.html> (last visited July 15, 2009).

¹¹⁴ Late in the BTK investigation, police had a suspect, and also had crime scene DNA samples. However, rather than try to recover a DNA sample from their suspect, they accessed his college-aged daughter's pap smear sample, on file with a health clinic, and found a partial match. The killer's full confession negated the need to introduce this DNA evidence at trial, however. *See* Nakashima, *supra* note 101.

¹¹⁵ *See* Nakashima, *supra* note 101.

¹¹⁶ *See* Maura Dolan & Jason Felch, *Tracing a Crime Suspect Through a Relative*, L.A. TIMES (Nov. 25, 2008) ("The FBI and California law enforcement officials long resisted the approach, fearful of inciting legal opposition and a public backlash. They yielded only after aggressive lobbying by prosecutors, who pointed to some dramatic successes . . . It took a dogged three-year campaign by a Colorado prosecutor to bring down the barriers to familial searching in the United States."); Denver D.A.'s office, DNA Resource, http://www.denverda.org/DNA/DNA_INDEX.htm (last accessed July 17, 2009).

¹¹⁷ Although CODIS at the time prohibited deliberate trolling for partial matches, the system did permit low-stringency searches designed to match at least 10 of 13 CODIS markers (that is, 20 of 26 alleles) – this functionality was included mainly to uncover matches to degraded DNA samples, although it occasionally would return partial matches to complete profiles. *See* Seth Axelrad, *State Regulations on Low Stringency / Familial Searches of DNA Databases*, American Society of Law, Medicine & Ethics, available at http://www.aslme.org/dna_04/reports/axelrad1.pdf (last accessed July 9, 2009).

backlash that would endanger the expansion of the DNA database – including upcoming plans to add arrestees.¹¹⁸ In June 2006, Morrissey contacted senior FBI directors directly to complain, and in response the FBI instituted an interim policy permitting states to perform partial-match searches.¹¹⁹

Even with the FBI go-ahead, states still had the final word as to whether to release the names of partial matches to out-of-state inquiries. California, the holdout on Morrissey's list, eventually bucked to pressure and not only granted Morrissey access to his long-sought suspect, but in a surprising turn, instituted a full-fledged familial DNA search policy of its own.¹²⁰ That policy, the most aggressive in the nation, not only permits California law enforcement agencies to follow up on 'accidental' familial DNA hits, but also to employ low-stringency searching deliberately to troll for family members in its own databases.¹²¹

Maryland is currently the only state to have banned familial DNA searching.¹²² Massachusetts and New York expressly authorize partial DNA searches via statute, but neither state regularly conducts them.¹²³ Other states seemingly have neither explicit legislative endorsement nor prohibition of the technique, leaving it to the discretion of the crime lab.¹²⁴

¹¹⁸ Then-CODIS Director Callaghan is quoted in several news reports as having responded with this concern. *See, e.g.*, Dolan & Felch, *supra* note 104 ("The FBI feared that racing ahead to familial searching could prompt a backlash and endanger database expansion").

¹¹⁹ *Id.*

¹²⁰ California Dep't of Justice, Division of Law Enforcement, Information Bulletin No. 2008-BFS-01, DNA Partial Match (Crime Scene DNA Profile to Offender) Policy (2008).

¹²¹ The California policy also includes various safeguards and limitations on the use of partial-match DNA searching. For instance, such searches are limited to violent-crime cases with single-source DNA samples (i.e., no mixed samples containing multiple individuals' DNA), where all other investigative leads have been exhausted. *Id.*

¹²² *See* Nakashima, *supra* note 101.

¹²³ *See* Duncan Carling, *Note, Less Privacy Please, We're British: Investigating Crime with DNA in the U.K. and the U.S.*, 31 HASTINGS INT'L & COMP. L. REV. 487 (2008).

¹²⁴ Interview with Dr. Carll Ladd, Ph.D., head of DNA forensics, Connecticut State Crime Lab, 278 Colony St., Meriden CT (Fri. Mar. 27, 2009). Dr. Ladd stated that while no regulation prohibits his lab from conducting partial stringency searches, it is currently unclear whether to do so, whether to report any inadvertently obtained partial matches, and more importantly, how courts would deal with the use of such evidence.

The FBI permits states to conduct familial DNA searches under its interim policy, but does not openly conduct them itself. In March 2007 the FBI held a symposium to consider reform of this interim policy, but ultimately decided not to permit familial searches,¹²⁵ since the Bureau would be “more comfortable with congressional authorization to conduct familial searches.”¹²⁶ Indeed, CODIS software is not designed for such requests, and will often toss out the vast majority of promising familial DNA matches and return instead random matches at the same threshold stringency.¹²⁷ As a result, Colorado – under Morrissey’s guidance – has implemented its own software to conduct familial DNA searches in that state.

And what of the Denver D.A.’s crusade? All three of Morrissey’s suspects – catalysts for watershed reform on familial search in the U.S. – turned out to be duds: random partial matches from unrelated individuals.¹²⁸

Familial DNA: Technical Considerations

From a technical and statistical standpoint, it is easy to see why familial DNA searches are often ineffective or misleading. Designed only to return matches at a certain threshold number of loci, CODIS ignores the hallmark inheritance patterns that characterize inheritance relationships. Partial matches returned may be parent-child, or sib-sib, but more often still are likely to be random individuals with no meaningful genetic relationship.¹²⁹ Even expanding the number of CODIS loci from 13 to 15 would not permit true sib-pairs to be distinguished with confidence from false matches.¹³⁰

¹²⁵ See Rosen, *supra* note 27.

¹²⁶ Nakashima, *supra* note 101.

¹²⁷ See Dolan & Felch, *supra* note 104. This is in part because the software does not distinguish inheritance patterns or allelic rarity in reporting results: in particular, a pair of profiles with a certain combination of rare alleles is much more likely to point to a true familial relationship. See discussion in Section I, *supra*.

¹²⁸ Dolan & Felch, *supra* note 104.

¹²⁹ Such searches will still return a large number of false positives, even if allelic rarity is considered. This was neatly illustrated by a 2008 research study. Using a mock offender database containing 12,292 profiles, Reid and coworkers sought through two methods to identify known (true) sibling pairs and

Performing partial stringency searches in large databases can uncover interesting, if not particularly meaningful, trends. For instance, researchers studying the Arizona offender DNA database in 2008 found an “unusually” high number of partial profile matches within that collection – suggesting either that Arizona’s DNA databank was strangely rich in siblings, or that some problem existed with the search algorithm.¹³¹ FBI researchers published a pointed response, arguing that because DNA databases represent mere subsets of the population, they are expected to deviate from genetic equilibria.¹³² Unexpected spikes in matches may occasionally occur, but these are evidence only of the ‘birthday scenario,’¹³³ not of any fault in the underlying mathematics.

exclude false positive matches. This study tests both profile matching based on shared alleles and kinship matching, which also takes into account the population frequency of the shared alleles. While both methods were effective at (eventually) identifying true sib pairs, they both produced an unacceptably high number of false positives alongside the true match. Thus, siblings are always easy to identify in partial-match-stringency searches; in other words, a partial search of an offender database will likely return siblings, if any exist, as well as other unrelated individuals who nonetheless share similar DNA profiles at the CODIS loci. See Thomas M. Reid, Michael L. Baird *et al.*, *Use of sibling pairs to determine the familial searching efficiency of forensic databases*, 2 FORENSIC SCI. INT’L: GENETICS 340 (2008).

¹³⁰ See Chang E. Pu & Adrian Linacre, *Systematic evaluation of sensitivity and specificity of sibship determination by using 15 STR loci*, 15 J. FORENSIC & LEGAL MED. 329 (2008). (Even using 15 STR loci does not permit all true sib pairs to be distinguished from non-sib-pairs. The best approach for distinguishing them, absent parental DNA, is a combination of “combined sibship index” and “two-allele-sharing locus” measures.)

¹³¹ Of 65,493 profiles in the Arizona database in 2008, 122 pairs of individuals matched at 9 of 13 STR loci, 20 pairs matched at 10 of 13, and 1 pair each matched at 11 and 12 loci. In past, some have attempted to use this statistical anomaly to question the underlying “random match” probability calculations used in DNA matching. See Laurence D. Mueller, *Can simple population genetic models reconcile partial match frequencies observed in large forensic databases?*, 87 J. GENETICS 101 (2008) (positing that the AZ offender database either contains between 1000-3000 sibling pairs, or is an exceedingly rare statistical anomaly).

¹³² Bruce Budowle, F. Samuel Baechtel & Ranajit Chakraborty, *Partial matches in heterogeneous offender databases do not call into question the validity of random match probability calculations*, 123 INT. J. LEGAL MED. 59 (2009). Budowle and colleagues argue that because offender databases are highly heterogeneous in population terms, yet also contain duplicate profiles and profiles of close relatives, they are expected to violate the basic assumptions of independence and therefore deviate from expected probabilistic norms. High-level partial matches will occur at some rate by chance: For instance, a 9/13 STR match between a Caucasian and an African American in the AZ database. There are 715 possible combinations of 9 loci out of the 13 CODIS loci. Because the population in the db is skewed, deviations from probabilistic ‘expectation’ should be expected in CODIS; this does not call into question the standard probabilistic mathematics used in determining DNA matches, however.

¹³³ The birthday scenario is a well-known statistical brainteaser that asks what the odds are that any random person shares the same birthday as you. The odds are normally about 1/365. However, in a room of 23 people there is a ~50% chance that any two will share the same birthday; the odds climb to 99.9% for a room of 75 people. This ‘paradox’ hinges upon the fact that in the first case, you have limited the

The power of familial DNA matches will benefit to a degree from specialized search software, such as that recently tested in Colorado, and it would benefit also from any expansion of the number of STR loci that CODIS profiles.¹³⁴ However, given the current set of 13 CODIS loci and rapidly expanding DNA database coverage, familial search seems poised to yield a staggering number of false positive leads.

Familial DNA: Legal Concerns

The California policy on familial searching is an order of the executive branch,¹³⁵ and to date no U.S. court has dealt with familial DNA searching.

The practice raises substantial Fourth Amendment concerns. The databasing of offenders' DNA profiles (whether they be arrestees, parolees, or convicts) has to date always turned upon the diminished expectation of privacy assigned to these individuals, which in turn justifies the application of the "totality of the circumstances" test for constitutionality of the search.¹³⁶

However, if a diminished expectation of privacy is a necessary prerequisite for inclusion in the searchable database, why should law enforcement be able to end-around this issue and search in such a way that the database effectively includes close relatives of offenders?

This raises a subtle distinction: In reaching the aforementioned decisions, courts have characterized the 'search' as the actual taking of the DNA sample (by cheek swab or blood test) as opposed to queries against the resultant profile in the database. Once the

possible solutions to a single date; in the latter cases, the number of possible solutions grows as the population of the room increases. In this case, the FBI argues, Mueller was not looking for specific partial matches, but rather *any* partial match. This subtle difference accounts for a substantial result.

¹³⁴ See Pu & Linacre, *supra* note 130. See also Dolan & Felch, *supra* note 104 (quoting Lance Gima, California's top forensic scientist, who stated that although familial searching would provide new leads to police, "the search for relatives would be a longshot because many unrelated people share genetic markers. He said he hoped the state's decision would spur technology to improve the accuracy of such searches.").

¹³⁵ See California Policy, *supra* note 120.

¹³⁶ For a full discussion, see Section III, *supra*.

government has a DNA profile and is authorized to store it,¹³⁷ it may be difficult to restrict government use of it unless courts find that familial searching is beyond the mandated scope for such database records,¹³⁸ or tack back to find that queries against the record themselves constitute a search.

At present, database searches that extend to family members may be permitted on the simple theory that, as far as the relatives are concerned, “nothing has been seized from them and they have not been searched.”¹³⁹ Thus, Professor Greely of Stanford Law School finds “no general legal barrier to the use of family forensic DNA.”¹⁴⁰

Professor Jeffrey Rosen of George Washington University Law School has argued that “[t]he strongest legal argument against familial searches is that they're not what Congress intended when it set up the [DNA] database.”¹⁴¹ Rosen cites the two main reasons behind the 2004 9th Circuit (en banc) decision *U.S. v Kincade*,¹⁴² namely the diminished expectation of privacy, and the state’s interest in ensuring reform over recidivism. “Familial searches can't be justified by either rationale. The family members of offenders have done nothing to reduce their expectation of privacy, and the state is investigating new crimes, not stopping repeat offenders.”¹⁴³

California’s policy on familial DNA searching – at present, the only such policy in existence – has set an informal precedent that familial DNA searching should be done only for serious crimes and when other leads have been exhausted. However, even if this

¹³⁷ That is, the person has not been cleared or released, and has not filed a request to purge pursuant to 42 U.S.C. § 14132(d)(1)(A).

¹³⁸ Here, possible traction may be afforded by the fact that deliberate partial-stringency searches could perhaps be classed as a different type of ‘use’ of the stored DNA data.

¹³⁹ Greely, *supra* note 33, at 257.

¹⁴⁰ *Id.* at 258.

¹⁴¹ Rosen, *supra* note 27.

¹⁴² 379 F.3d 813. *See supra* note 40 and related discussion.

¹⁴³ Rosen, *supra* note 27.

standard meets judicial scrutiny, local labs are currently free to conduct partial searches as they see fit – Colorado, for instance, has used the technique to solve car break-ins.¹⁴⁴

Database Bias

Familial DNA searching is also potentially politically worrisome, particularly from the standpoint of racial bias.

Dr. Frederick Bieber, co-author of the seminal *Science* piece on familial DNA searching¹⁴⁵ and a medical geneticist at Brigham and Women’s Hospital in Boston, has opined that familial searching of offender databases would not be useful “if close relatives didn’t commit crimes.”¹⁴⁶ He points to a “familial clustering in crime,” citing a 1999 study that found “46% of prison inmates had at least one close relative who had been incarcerated.”¹⁴⁷

One might argue that it is one thing to note such statistics, and another to employ them to direct investigations. Any investigative technique that posits beginning with the relatives of convicts as first-order suspects would surely run counter not only to American ideals of civil liberty, but also to constitutional protections against bills of attainder and “corruption of blood.”¹⁴⁸

As several scholars have pointed out, familial searching aggravates inherent racial bias in the CODIS database.¹⁴⁹ Because African Americans make up a disproportionately large fraction of felony arrests in the United States, Professor Greely hypothesizes that

¹⁴⁴ See Dolan & Felch, *supra* note 104.

¹⁴⁵ Bieber, *supra* note 106.

¹⁴⁶ Dr. Frederick Bieber as quoted in Nakashima, *supra* note 101.

¹⁴⁷ *Id.*

¹⁴⁸ U.S. Const. Art. III, § 3, cl. 2 states that “The Congress shall have Power to declare the Punishment of Treason, but no Attainder of Treason shall work Corruption of Blood, or Forfeiture except during the Life of the Person attainted.” Note, however, that this protects relatives from conviction on account of their familial relationships; it does not explicitly protect them from suspicion or even investigation on the same account.

¹⁴⁹ See David H. Kaye and Michael E. Smith, *DNA identification databases: Legality, legitimacy and the case for population-wide coverage*, WISC. L. REV. 413 (2003); Greely, *supra* note 33; Murphy, *supra* note 20; Grimm, *supra* note 105.

they account for a similarly large fraction of records in the CODIS database.¹⁵⁰ By applying familial DNA searches, the fraction of the total African American population that would be included in CODIS (either directly or indirectly) would be roughly four to five times larger than the corresponding fraction of Caucasians or Hispanics.¹⁵¹ (A student note has argued that due to the commonness of large families among U.S. Hispanics, that population might also be overrepresented in databases ‘expanded’ by familial DNA searches.¹⁵²)

Now that arrestees are included in CODIS, the numbers will move even further in this direction: On average, 28.5% of African American men, 16% of Hispanic men and 4.4% of Caucasian men are likely to be imprisoned on a felony charge at some point in life.¹⁵³ Extending this to include the first-degree relatives of these people means a very substantial fraction of the African American population will be ‘searchable’ in CODIS, compared to a small fraction of the Caucasian population. If all arrestees are included, not simply felony arrestees, Kaye and Smith point out that the end result may be functionally indistinguishable from a universal DNA database for African Americans only.¹⁵⁴

It is not clear what legal issues may arise from this disparate representation by race in CODIS databases; courts have not yet addressed the issue. Professor Greely has discounted equal protection arguments based on disparate impact, but nonetheless states that this disparity “does seem fundamentally unfair.”¹⁵⁵ He suggests that the issue may have more political than legal traction.¹⁵⁶

¹⁵⁰ Greely, *supra* note 33 at 258.

¹⁵¹ Roughly 17% versus 4%. Greely, *supra* note 33 at 258-259.

¹⁵² See Grimm, *supra* note 105.

¹⁵³ See Kaye & Smith, *supra* note 149, at 452.

¹⁵⁴ *Id.* at 455-456.

¹⁵⁵ Greely, *supra* note 33, at 259.

¹⁵⁶ *Id.* at 259-260.

Other Issues: Phenotypic Profiling, Gems in the Junk, and Sample Retention

Given the vast amount of information contained in our genome – sufficient to blueprint an individual human being, complete with medical traits, hereditary predispositions to disease and outward identifying characteristics – it may seem surprising that forensic use of DNA has thus far been limited to a handful of essentially meaningless sequence repeats. There is no technical reason that this need be so: After all, given the time and resources to sequence and analyze an offender's genome, law enforcement agencies could conceivably prepare, among other things, highly detailed phenotypic descriptions of suspects.

Once the purview of science fiction, extraction of meaningful tidbits from our ocean of genomic data is becoming more common and more feasible. The drivers for this development are twofold. First, drastic reductions in sequencing cost and concomitant leaps in sequencing speed are rapidly making personal genome sequencing a reality.¹⁵⁷ Second, through academic and medical research into the meaning of individual genetic traits and markers – chiefly, their correlation with phenotypic, disease or other states – islands of meaning are cropping up.

Private companies such as 23andMe already offer personal genetic profiling services based on a mail-in cheek swab;¹⁵⁸ medical workers have been providing genetic testing for disease predisposition (and associated counseling) for some years now. Patent rights for genetic tests involved with cancer predisposition are at the center of a lawsuit filed May 12, 2009, involving the ACLU and biotech company Myriad Genetics.¹⁵⁹ And

¹⁵⁷ See, e.g., Avak Kahvejian, John Quackenbush & John F Thompson, *What would you do if you could sequence everything?*, 26 NAT. BIOTECH. 1125 (2008).

¹⁵⁸ 23andMe Homepage, <http://www.23andme.com/> (last visited July 19, 2009).

¹⁵⁹ American Civil Liberties Union: ACLU Challenges Patents on Breast Cancer Genes, <http://www.aclu.org/freespeech/gen/brca.html> (last visited July 19, 2009).

futurists have long worried what might happen if our genes one day divulged information about our propensity for crime.¹⁶⁰

So far, courts have specifically noted the “junk” status of the CODIS STRs in upholding DNA collection and profiling under the Fourth Amendment.¹⁶¹ However, recent scientific work has shown that STR data (albeit at many more than 13 loci) is sufficient to gain a great deal of information about an individual, even identifying their geographic birthplace to within a few hundred kilometres.¹⁶²

With this new technology now a reality, privacy concerns about government access to our genomic information is no longer science fiction.

Another concern hinges upon DNA sample retention. Under current policies, the FBI retains not only the computerized DNA profile, but also the biological tissue sample, for all individuals included in CODIS. The government’s ability to perform additional follow-up tests on this sample, such as Y-STR typing, may overstep the bounds under which constitutional authorization was granted to take the sample in the first place.

Such “CODIS creep”¹⁶³ is a prime risk of DNA databases: Because only a tiny sample of tissue is needed to generate a CODIS entry, samples originally taken for that purpose could easily be stored and re-used for another, presumably with little or no judicial oversight. Given the wealth of health-related and other valuable data buried in those very genomic samples, this is understandably a prime worry of privacy advocates today.¹⁶⁴ As such, Kaye & Smith have advocated destroying tissue samples immediately

¹⁶⁰ See, e.g., FuturePundit, *British DNA Crime Database Ups Crime Clearance Rates* (Apr. 19, 2005), <http://www.futurepundit.com/archives/002727.html> (positing that in the near future, “[c]riminals will be genetically classified by their propensity to commit various types of crimes.”)

¹⁶¹ See Section II, *supra*.

¹⁶² John Novembre, Toby Johnson *et al.*, *Genes mirror geography within Europe*, 456 NATURE 98 (Nov. 6, 2008) (analyzing some 500,000 variable markers among 3,000 European individuals and generating with “surprising accuracy” a map of their origins).

¹⁶³ See Rosen, *supra* note 27.

¹⁶⁴ *Id.*

upon communication of CODIS profiles to the government, and even suggested dedicated machines to accomplish both tasks in a single step.¹⁶⁵

Forensic DNA scientists prefer STR loci precisely because they do not convey meaningful phenotypic information.¹⁶⁶ However, neither a geneticist nor a data-mining expert would be particularly surprised to learn that personal information can be extracted from STR loci given a large enough data set. Thus, the real check on improper use of DNA profiles must lie not in the characteristics of the underlying data, but in the strict regulation of the means by which law enforcement agencies access and use the records. There may be no such thing as “junk DNA,” but any data can be made non-invasive through careful processing. By recording only the lengths of STR loci, and using this set of numbers as identifiers solely to match individuals against their own stored records, the opportunity for misuse of DNA data is virtually nil.¹⁶⁷

If courts are to continue to uphold the forensic collection and query of DNA profiles based on STR data, law enforcement agencies must at some point reassure the judiciary (and the public) that the information contained therein serves no purpose beyond identification – and if it does, that records are not accessed in such a way as to access that information.

V. TECHNICAL AND POLICY REFORM

The steady advance of scientific understanding, DNA sequencing techniques and computer technology demands that lawmakers regularly revisit policies concerning DNA sampling and storage. Certain developments, specifically familial searches, rapid

¹⁶⁵ See Kaye & Smith, *supra* note 149, at 437.

¹⁶⁶ Interview with Dr. Carll Ladd, *supra* note 124.

¹⁶⁷ Misuse in the sense of invading privacy. Other forms of exploitation, such as planting lab-generated DNA samples to frame individuals, is frighteningly simple. See, e.g., Dan Frumkin, Adam Wasserstrom, et al., Authentication of forensic DNA samples, *FORENSIC SCI. INT. GENET.* (17 June 2009).

expansion and the availability of useful phenotypic data from genomic information, threaten to push the current CODIS system beyond its original mandate. This section suggests two main reforms: First, dramatically expanding the number of profiled STR loci, and second, expanding the DNA database to include all U.S. citizens.

Technical Reform: Expanding the CODIS Loci

Familial DNA searching raises important concerns: Not only because it subjects relatives to scrutiny, but also because in most cases such searches will result in entirely unrelated individuals being investigated for crimes they did not commit. Assuming that familial DNA searching will continue to gain acceptance in the U.S., one relatively simple reform could dramatically increase its usefulness while sharply curbing its impact on unrelated people: Increasing the number of CODIS markers.

There is no special reason why CODIS must use 13 variable markers, while U.K. forensic databases use 10 and *Nature* researchers profiling genetic origins use 500,000. Many thousands of suitable STR loci are situated in so-called “junk DNA” regions, and could easily be added to the CODIS set.

Such an addition is necessary. Although matching 26 alleles at 13 loci is a robust metric for assessing pure identity, their familial or partial-search counterpart is woefully inadequate. Simply too many chance matches of, say, 9 of 13 loci will exist in DNA databases to even remotely pinpoint siblings; worse still, CODIS software routinely discards information useful in identifying true relatives when returning search results.

Increasing the number of profiled loci will achieve a substantial increase in performance of familial DNA searching, and only intensify the already statistically powerful foundation of the exact match. Moreover, with 50 or 100 loci profiled and allelic rarity data considered, it would be conceivably possible to recover even second- and higher-degree relatives from a familial search, all the while limiting false positives.

There is no ‘correct’ number of loci to profile, but much like the apocryphal Bill Gates quotation that 640k of memory should be enough for anyone,¹⁶⁸ technological requirements do change over time. Although 10 or 13 STR loci were more than enough to yield a one-in-several-billion sample-to-sample match twenty years ago, database expansion into the millions and the push for familial searching has strained this handful of markers.

Denver D.A. Morrissey has stated that the 13 CODIS markers are “an extremely powerful tool,” and that declining to employ them for familial DNA search would be like building “a Porsche [and driving it] like a Pinto.”¹⁶⁹ The D.A. has it backwards: familial search is driving a Pinto like a Porsche, and if authorities are set upon doing so they should soon upgrade their motor pool.

Expanding the number of CODIS markers is not some far-off ideal: it is in order and past due. That academic researchers are profiling a half million markers at a time in population studies hints at the plummeting costs involved for modern day DNA profiling.

Virtually no cost will be involved in identifying STRs to use: Thousands have already been characterized and are searchable online.¹⁷⁰ Indeed, the cost of profiling additional STR markers in the laboratory should be marginal as well, once standard forensic DNA kits are re-issued to include the correct PCR primers. The most convenient number of loci to profile might in fact be 96, which corresponds to the number of wells in a standard laboratory microtiter plate: the same type commonly used by biologists, geneticists and accepted by lab machinery worldwide.¹⁷¹

¹⁶⁸ Which Gates claims never to have said; *see* Wired Magazine, Did Gates Really Say 640k Is Enough For Anyone? <http://www.wired.com/politics/law/news/1997/01/1484> (last accessed July 18, 2009).

¹⁶⁹ Morrissey quoted in Nakashima, *supra* note 101.

¹⁷⁰ National Institutes of Science and Technology, STRbase, <http://www.cstl.nist.gov/strbase/> (last visited July 2, 2009); some STRs have already been analyzed and recommended for forensic use. *See* <http://www.cstl.nist.gov/strbase/newSTRs.htm>

¹⁷¹ *See, e.g.*, Fisher Scientific: Semi-skirted, 96-well PCR plates, <http://www.fisherlab.info/landscape.cfm?getid=82> (last accessed July 10, 2009).

DNA forensic profile matches have gained broad law enforcement support because they offer a powerful window into the genomic underpinning of each unique human being. But as new techniques are added and the database continues to grow, CODIS should increase its profiled markers to keep pace.

Policy Reform: In Support of a Universal DNA Database

A universal DNA database is at once a *bête noire* for privacy advocates, and the obvious answer to the problems of racial bias and familial DNA searching facing U.S. DNA databases today.

The proposal is not new. In 2002, Professor Akhil Amar of Yale Law School argued publicly in support of a universal U.S. DNA database, which would be “a godsend to innocent convicts.”¹⁷² The following year, Kaye & Smith detailed a proposal for such a database.¹⁷³

The idea has received a fair amount of criticism. Professor Rosen calls the idea “utopian,” arguing that sample retention, the lack of nuanced privacy regulations, and the inability to ensure that the government would access the database only to solve crimes makes this proposal unfeasible.¹⁷⁴ Some scholars recognize that more coverage in DNA databases is desirable, but stop short of advocating universal inclusion.¹⁷⁵

In fact, Kaye & Smith address sample retention directly by arguing that the government should not retain tissue samples after an STR profile is generated.¹⁷⁶ Indeed,

¹⁷² Akhil Reed Amar, *A Search for Justice in our Genes*, THE NEW YORK TIMES (May 7, 2002) at A31, available at <http://www.law.yale.edu/news/4279.htm> (last accessed July 8, 2009).

¹⁷³ See Kaye & Smith, *supra* note 149.

¹⁷⁴ See Rosen, *supra* note 27.

¹⁷⁵ Mark A. Rothstein & Sandra Carnahan, *Legal and Policy Issues in Expanding the Scope of DNA Data Banks*, 67 BROOK. L. REV. 127, 129-130 (2001) (stating that “DNA data banks of increasingly broad scope would help to solve more crimes” but stating also that “public policy” requires that only convicted felons and sex offenders be included). Public policy is fickle: Note that within several years of this article, CODIS policies had already expanded to include substantially more people than this.

¹⁷⁶ See Kaye & Smith, *supra* note 149, at 437, 440 n. 96.

an argument for a universal DNA database should be sharply distinguished from a call for widespread “genetic surveillance,” as some privacy advocates put it.¹⁷⁷

Instituted correctly, a universal DNA database would resemble a collection of Social Security numbers much more than a medically useful genomic snapshot.

As stated in the previous section,¹⁷⁸ the factor that most determines the potential invasiveness of a DNA profile is not its source (in genetic material), but rather the form in which it is stored and accessed. STR profiling at the 13 CODIS loci distills the almost limitless variability contained in the human genome to a single 52-digit number.¹⁷⁹ These numbers have the potential to uniquely identify every human on the planet, and carry the concomitant forensic benefit of being routinely found in crime scene evidence, shed skin cells, even on discarded cigarette butts.

So long as the original tissue sample is not retained, or at the very least, not later subjected to different analyses – a problem which, though sinister, need not accompany the universal STR database itself – the opportunity for misuse of this data is virtually nonexistent.

A universal database renders moot any need for familial DNA searching and its accompanying aggravation of racial bias. If everyone is already searchable as an exact match, nothing is gained by extending this search to relatives. Moreover, there is no possibility that certain communities will draw an undue share of police interest on the grounds of their inclusion in the database alone.

Counterintuitively, expanding the database to include everyone actually removes some of the pressure to expand the CODIS markers, much of which stems from the inadequacy of using 13 markers for familial DNA search. However, expanding the list of

¹⁷⁷ Tania Simoncelli, ACLU Science Advisor, as quoted in Nakashima, *supra* note 101; Prof. Rosen also uses this term. *See* Rosen, *supra* note 27.

¹⁷⁸ *See supra* note 166, and related discussion.

¹⁷⁹ *See supra* note 38. Conceptually, a DNA profile of this sort is quite similar to an SSN: both are ideally ‘assigned’ at birth, remain unchanged throughout natural life, and are uniquely tied to each individual.

markers remains a wise step: We do not ask that the government struggle with five-digit SSNs for a population of 300 million, and similarly there is no reason to restrict to 13 STRs a profile that could just as easily assess many times that number.

Perhaps most obviously, a universal database will also increase apprehension rates, decrease repeat offenses and deter crime more effectively than the original or current CODIS model. As Kaye and Smith point out, “a convicted-felon database is of no help in deterring or investigating felonies committed by persons not previously convicted of a felony”¹⁸⁰ and similarly, an arrestee database is of no use deterring those who have not yet been arrested. But even a database including all arrestees – which CODIS is on track to become – will nonetheless exclude those responsible for a “major proportion” of all felony offenses: 44% of people arrested and prosecuted for a serious felony have never been arrested on felony charges, and roughly one third have no arrest record at all.¹⁸¹ Current DNA database coverage positions forensic DNA as a tool for deterring and investigating primarily repeat offenses: an occurrence whose very existence a nationwide database should markedly diminish.

Fourth Amendment concerns facing the universal database can perhaps be addressed by casting the database not as a law enforcement tool – as indeed it might not primarily be – but instead as a recordkeeping and identity management center, such as those used by the Social Security Administration. Instead of compelling sample collection from individuals with a diminished expectation of privacy, sample collection via painless, non-invasive cheek swab could be a required step for any number of societal hurdles: obtaining a driver’s license, for instance, or a Social Security Number.¹⁸² The

¹⁸⁰ See Kaye & Smith, *supra* note 149 at 451.

¹⁸¹ *Id.* at 451 n.134.

¹⁸² Compelling government interests in identification can be envisioned beyond those for law enforcement: For instance, state Department of Motor Vehicles offices might have an interest in identifying remains from road accidents; the Social Security Administration has an obvious interest in identification, which could extend to missing persons, disoriented or mentally challenged individuals, and so forth. These

Supreme Court has permitted regular searches where justified by a routine, administrative purpose;¹⁸³ such special circumstances have led the court to apply “general reasonableness” or “totality of the circumstances” analyses, the same standards that have been used to uphold current DNA collection under CODIS.¹⁸⁴

As Professor Amar points out, expanding the database will also eliminate innocents from suspect lists – a compelling government interest that should weigh on the side of sample collection.¹⁸⁵

A universal DNA database without tissue sample retention and would enhance the positive aspects of the current CODIS system: Deterring crime and recidivism, speeding resolution of crimes involving recorded persons, potentially exonerating the falsely accused, and permitting identification of missing persons. At the same time, such an expansion would diminish or negate entirely many of the negative aspects that currently plague CODIS: Racial bias, inefficient, misleading and legally dubious familial DNA searching, database creep, and the threat of unwarranted genetic profiling based on phenotypic data.

Once the database is in place, Fourth Amendment challenges to DNA profiling would shift from upfront sample inclusion to *in situ* sample recovery. Courts should permit crime scene samples to be profiled and searched against the universal database,

ideas arose during the author’s discussion with Prof. Hank Greely, Stanford Law School (Friday June 19, 2009).

¹⁸³ See *Griffin*, 483 U.S. at 873, *supra* note 74.

¹⁸⁴ See *Samson*, 547 U.S. at 848, *supra* note 76.

¹⁸⁵ See *U.S. v. Sczubelek*, 402 F.3d 175, 185 (3rd Cir. 2005) (“A DNA database promotes increased accuracy in the investigation and prosecution of criminal cases. It will aid in solving crimes when they occur in the future. Equally important, the DNA samples will help to exculpate individuals who are serving sentences of imprisonment for crimes they did not commit and will help to eliminate individuals from suspect lists when crimes occur The interest in accurate criminal investigations and prosecution is a compelling [government] interest that the DNA Act can reasonably said to advance.”). Moreover, this is an area where the government could use some prodding. See Shaila Dewan, *Prosecutors Block Access to DNA Testing for Inmates*, THE NEW YORK TIMES (May 17, 2009).

while at the same time prohibiting suspicionless monitoring such as random spot-checks of an individual's travels via shed skin cells or similar means.¹⁸⁶

As Professor Amar has suggested, such a database would demand careful regulation and political oversight.¹⁸⁷ And although a detailed regulatory and constitutional treatment of this proposal is beyond the scope of this paper, it does on its face seem to present, at least, fertile ground for further discussion. After all, compared with the current instantiation of CODIS – a racially skewed and steadily expanding partial-coverage database with full sample retention and poor regulation on follow-up genetic assays, that is steadily marching toward widespread adoption of familial search – an STR-only, minimally intrusive universal database looks fairly tame.

VI. CONCLUSION

The CODIS forensic DNA database is expanding steadily, and now includes arrestees and various foreign detainees in addition to its original corpus of convicted felons. New technologies and techniques such as familial search are stretching the power of the original 13 CODIS loci, a reference set that should promptly be expanded to assay several-fold more DNA markers. Moreover, the threat of increasingly invasive phenotypic inquiries into stored genomic tissue samples highlights the need for a sample-free, records-only DNA database. This database should be universal, thereby negating the need for familial search and erasing in a stroke the racial bias that mars and continues to encroach upon CODIS today.

¹⁸⁶ Courts have dealt with similar issues concerning fingerprints. In *Palmer v. State*, 679 N.E.2d 887 (Ind. 1997), the Indiana Supreme Court found that acquisition of defendant's fingerprints during trial and without a warrant does not amount to a forbidden seizure under the Fourth Amendment since "fingerprints are an identifying factor readily available to the world at large." *Id.* at 891. However, this is likely a minority view. See Kaye & Smith, *supra* note 149, at 442 n.105.

¹⁸⁷ See Amar, *supra* note 172 ("Building a universal database would encourage the creation of a broad political coalition with incentives to protect privacy interests from being eroded over time.")

Instituting such a database demands reconsideration of the judicial reasoning that has thus far upheld CODIS, a line of cases justifying DNA collection under the loose “general reasonableness” standard of Fourth Amendment jurisprudence on the grounds that offenders of various types have a diminished expectation of personal privacy.

Privacy advocates rightly balk at the widespread mining of individuals’ DNA, either by government or private interests. But STR profiles must be sharply distinguished from full genomic DNA – as opposed to 3 billion-letter text strings chock-full of juicy biological tidbits, they are, at present, simply 52-character numerical strings without phenotypic correlation.

There may be no such thing as “junk DNA,” but there is such thing as non-invasive data – and uniform STR profiles come very close to this ideal. Consequently, such profiles are ideally suited for use as unique identifier tags to replace or to supplement SSNs, and carry the added benefit of being left behind at crime scenes.