

TRYING Y-STR DNA CASES

Y chromosome DNA, from which Y-STR forensic markers are derived, differs from traditional nuclear DNA on a number of counts that should be central to a defense challenge. Just as with mtDNA, a well-planned challenge to Y-STR DNA evidence should home in on those differences in addition to following the general guidelines for DNA cases generally, outlined above. As with Section [] on mtDNA, what follows is a synopsis of the core substantive areas of forensic Y-STR DNA that can be developed through admissibility challenges under *Frye*, *Daubert*, or other evidentiary standards governing expert testimony and scientific evidence, or as a basis for challenging the expert directly through cross-examination. Again, a sample motion to exclude Y-STR DNA “match” evidence is provided as an appendix for reference.

I. Y-STR Basics

Y-STR DNA profiles are derived from the Y chromosome, which is passed down largely intact from a father to all male offspring, from generation to generation without changing (barring mutation).¹ The mutation rate of Y chromosome DNA is dramatically lower than that of mtDNA, which makes a given Y-STR profile extremely stable over many generations²: every male – fathers, sons; brothers; uncles; first-, second-, third-, and fourth-cousins; and all widely dispersed male relatives – in a paternal lineage will share the same Y-STR profile. Just as maternal lineages can be tracked with mtDNA, paternal lineages can be tracked with Y chromosome markers.³ Similar to mitochondrial DNA, a Y-STR profile can be thought of as a “single genetic locus,” as contrasted with the numerous independent loci available for traditional DNA typing. Because Y chromosome DNA does not undergo recombination at each generation, the discriminatory power of Y chromosome DNA also pales in comparison to traditional nuclear DNA.

Y chromosome DNA is found in the nucleus of the cell, along with the autosomal (non-sex) chromosomes from which traditional STR profiles are derived. The Y chromosome is found only in males, which limits its application, but also makes it particularly attractive to law enforcement in cases involving mixtures of male and female DNA, or cases where a male profile is sought where only female DNA is detected with traditional testing. Typing kits have been developed that will only respond to markers present on the Y chromosome, which allows law enforcement to isolate male DNA for forensic analysis that might otherwise have been overwhelmed by the presence of female DNA, or gone undetected altogether. As such, Y-STR DNA evidence is introduced most commonly in sexual assault cases.⁴

¹ JOHN M. BUTLER, FORENSIC DNA TYPING: BIOLOGY, TECHNOLOGY, AND GENETICS OF STR MARKERS 201 (2d ed. 2005).

² M. Kayser et al., *Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, As Revealed by Direct Observation in Father/Son Pairs*, 66 Am. J. Hum. Genetics 1580–1588 (2000) (observing the average mutation rate of Y chromosome DNA to be approximately 0.2% per generation).

³ BUTLER, *supra* note 1, at 201.

⁴ *Id.* at 201-03.

A core set of locations on the Y chromosome comprises a “minimal Y-STR haplotype” that has served as the basis of forensic applications since 1997.⁵ More recently, in 2003, SWGDAM recommended the use of the minimal haplotype loci in combination with two additional Y-STR markers known as DYS438 and DYS439, which have since been adopted in the major commercial kits that are used for Y-STR analysis.⁶ Most law enforcement agencies in the United States, including the FBI, outsource Y-STR testing to private laboratories, where testing is conducted using one of these kits.

II. The Limited Discriminatory Power of Y-STR

The minimal discriminatory power of Y-STR stands in stark contrast to traditional nuclear (autosomal) DNA. Y-STR DNA rarely changes over time – which also distinguishes it from mitochondrial DNA – and Y-STR profiles are shared by large, unknown numbers of distant relatives.⁷

That Y-STR profiles are exactly the same among distantly related males down centuries of patrilineal lines is well known. For example, Y-STR DNA analysis was used to investigate the assertion that Thomas Jefferson had fathered children with Sally Hemings. The Jefferson/Hemings investigation underscored two scientific truths: The first is that Y chromosome DNA can be used to exclude some men from large groups of males joined together by common paternal ancestry. The second truth is that, among the males in the group of common paternal ancestry, Y chromosome DNA *cannot* be used to distinguish one individual from another. Put another way, Y chromosome DNA is not a good tool to identify a particular person as the contributor of Y-STR DNA. For Jefferson/Hemings scholars, the result was that Y-STR DNA was useful to determine that children of Ms. Hemings were fathered either by President Jefferson *or* a male relative of his, but could not be used to definitively conclude that President Jefferson himself was the father.⁸

While Y-STR DNA has proved a powerful *exclusionary* tool, its prowess as an inculpatory device something short of remarkable. Scientists are in agreement that Y-STR analysis is a precise mechanism to exclude persons as possible contributors to DNA evidence, but is a mere blunt object in a quest to identify an individual person. Far short of identifying any particular individual as the source of a Y-STR profile, the observation of a profile that is consistent between an evidence sample and a suspect does no more than reduce the population of possible contributors to the defendant plus “all patrilineal related male relatives and an unknown

⁵ *Id.* at 207.

⁶ *Id.*; see J.J. Mulero et al., *Development and Validation of the AmpFISTER YFiler PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System*, 51(1) *J. Forensic Sci.* 64 (2006); J.G. Shewale et al., *Y-Chromosome STR System, Y-PLEX 12, for Forensic Casework: Development and Validation*, 49(6) *J. Forensic Sci.* 1278 (2004).

⁷ BUTLER, *supra* note 1, at 206.

⁸ See, e.g., Eugene A. Foster et al., *Jefferson Fathered Slave's Last Child*, 396 *Nature* 27 (1998); Turi E. King et al., *Thomas Jefferson's Y Chromosome Belongs to a Rare European Lineage*, 132 *Am. J. Physical Anthropology* 584 (2007).

number of unrelated males as being the donor of the evidence sample.”⁹ Consequently, “the observation of a match with Y-STRs does not carry the power of discrimination and weight into court as an autosomal STR match.”¹⁰ A defense challenge to Y-STR evidence should spotlight its marked deficiency at this central forensic task of distinguishing one individual from another.

III. Frequency Estimates and Y-STR Population Databases

As with mtDNA, a collection of deficiencies with available Y-STR population databases also provides fertile ground for a defense challenge. As with any DNA “match” evidence, the ability to assign significance to the match hinges on the reliability of the databases that serve as the basis for claims regarding the likelihood that a match is the result of coincidence, rather than identity.

A. Commercial Y-STR Databases Are Not Representative of the Population They Purport to Represent

It is fundamental that a DNA population database must comprise a randomly sampled, representative subset of the larger population it purports to represent, in order to provide a reliable estimate of the frequency of particular profiles in that population. The widely used “YFiler” database, for example – relied upon by law enforcement agencies relying on the YFiler typing kit for Y-STR analysis – is not randomly sampled. Despite the fact that scientific sampling is widely understood and employed in numerous scientific disciplines (e.g., political polling), the commercial company that compiles and maintains the YFiler database ignored the requirements of scientific sampling in constructing its database. As such, the YFiler database cannot serve as the basis of any claims relating to the frequency at which any Y-STR profile occurs in the population, which raises questions regarding any “match” evidence relying on that database for its statistical component.

B. Y-STR Profiles Cluster Regionally and Only Properly Sampled *Local* Databases Can Serve As the Basis for Meaningful Frequency Estimates

Unlike autosomal DNA, Y-STR profiles cluster geographically as they follow migration and settlement patterns. Even within discrete racial groups, scientific studies have shown that there are statistically significant differences in the frequency of Y-STR profiles depending upon which geographic locations are sampled.¹¹ Throughout the scientific literature, there exists

⁹ BUTLER, *supra* note 1, at 214.

¹⁰ *Id.* at 213-214.

¹¹ See e.g., Carolina Bonilla et al., *Admixture in the Hispanics of the San Luis Valley, Colorado, and Its Implications for Complex Trait Genemapping*, 68 *Annals of Hum. Genetics* 139 (2004) (reporting wide variation in genetic profiles of various ethnic groups falling under cultural rubric of Hispanic); M. Hedman et al., *Analysis of 16 Y STR Loci in the Finnish Population Reveals a Local Reduction in the Diversity of Male Lineages*, 142 *Forensic Sci. Intl.* 37 (2004) (particular sixteen-locus Y-STR profile is shared by thirteen percent of Finnish population); L Roewer et al., *Online Reference Database of European Y-Chromosomal Short Tandem Repeat (STR) Haplotypes*, 118 *Forensic Sci. Comm.* 106 (2001) (most frequent minimal haplotype is observed in three percent of continental European population); Michael E. Weale et al., *Armenian Y Chromosome Haplotypes Reveal Strong Regional Structure Within an Single Ethno-National Group*, 109 *Hum. Genetics* 659 (2001) (finding significant regional stratification of Y-STR DNA profiles and observing that London Armenian sub-sample was insufficient to describe genetic variation);

“considerable evidence for geographical substructure at the Y chromosome.”¹² Thus, some Y-STR profiles are very common in certain geographic locales, while wholly absent from others. The national YFiler database does not account for regional clustering of Y-STR profiles, and therefore does not provide a statistical estimation of the frequency of any particular profile in any population.

As a result of the substantial geographic substructuring of Y-STR DNA profiles, scientists have expressed “particular concern [over] the sampling of multiple populations and their assembly into global databases.”¹³ When Buckleton et al. surveyed the available scientific literature in 2005, they found “no report in the literature yet of how to interpret Y chromosome haplotypes accounting for population subdivision.”¹⁴ As a result, they reported that “further investigation into how to compensate for population subdivision at the Y chromosome is warranted urgently,” and urged that “it is imperative that every effort should be made to use appropriate local databases” when attempting to estimate the frequency of a given Y-STR profile.¹⁵

In other words, even if Y-STR population databases were randomly sampled in accordance with well-established scientific principles, they would only serve as a reliable basis for frequency estimates if they were sufficiently local to reveal patterns in geographic substructuring. Either shortcoming should serve as an independent bar to admission of Y-STR evidence in a criminal trial, for its failure to establish a reliable statistic representing the likelihood that an observed match between two profiles is the result of coincidence. A defense challenge to Y-STR DNA evidence – whether pretrial or before the jury – should focus on both fatal shortcomings of the available Y-STR databases: the absence of scientifically sound, random population sampling, and the absence of a local database representing the specific population where the defendant resides and/or the alleged crime took place.

IV. The Role of Ethnicity in Trying the Y-STR Case

The ethnic background of the accused may play a critical role in informing the appropriate defense strategy in a Y-STR DNA case, given the significant population substructuring described above and its influence on the distribution and frequency of a given Y-STR DNA profile in both the region in which the defendant resides and the region or regions from which the population databases were collected. At minimum, a defense attorney must be

M.T. Zarrabeitia et al., *Significance of Micro-Geographical Population Structure in Forensic Cases*, 117 Int'l J. Legal Med. 302 (2003) (studying Y chromosome profiles in Cantabria region of Spain and finding substantial overstatement of evidential strength frequently results from use of population databases collected on too broad geographical scale); Tatiana Zerjal et al., *The Genetic Legacy of the Mongols*, 72 Am. J. Hum. Genetics 717 (2003) (eight percent of 2100 males from Central Asia region closely matching area of Genghis Khan's former Mongol Empire had unique Y chromosome lineage).

¹² JOHN BUCKLETON, CHRISTOPHER M. TRIGGS, & SIMON J. WALSH, *FORENSIC DNA EVIDENCE INTERPRETATION* 324 (2005).

¹³ *Id.*

¹⁴ *Id.*

¹⁵ *Id.*

aware of the historical and contemporary migration patterns from the defendant's ancestral origin to both the geographic location where he resides, and the location or locations from which the database samples were drawn. Just as with mitochondrial DNA, if a defendant's ancestral origins are not represented in the Y-STR database but he is a member of a significant community of immigrants of common origin in his place of residence, a strong argument can be made that the lack of a "match" to his Y-STR profile in a database collected from a random set of individuals from elsewhere in the country is devoid of meaning, offering no information about the frequency of that profile in the relevant geographic area, and is therefore inadmissible under the controlling evidentiary standard.

By way of example, consider a defendant from San Miguel in eastern El Salvador, who immigrated to Washington, D.C. along with many of his countrymen, in the last twenty or thirty years.¹⁶ As such, the Salvadoran population of the D.C. region is a new community, having emerged within a single generation; in fact, this phenomenon of recent immigration helped to earn the District of Columbia the title of "immigrant gateway."¹⁷ In the absence of an opportunity for intermingling with established local communities, the Salvadoran population in D.C. can be shown to be genetically insulated and directly representative of its source population.

The defense can further argue that the Salvadoran community of the D.C. region is unique in a number of respects, due not only to the recency of its emergence, but due also to its uncommon genetic ancestry. That is to say, the Salvadoran source population is unlike more common sources of Hispanic immigrants to the United States, such as Mexico or Puerto Rico. In El Salvador, the native Indian population remained largely intact, despite Spanish conquests.¹⁸ In Mexico, by contrast, the majority of inhabitants have been classified as "mestizos," who are genetically traceable to a mixture of European and African ancestry, and the ancestral proportions of Puerto Ricans is heavily European and West African.¹⁹ Thus, Salvadoran immigrants are, as a group, genetically distinct from other Hispanics residing in the United States.

¹⁶ See, e.g., CARLOS B. CORDOVA, *THE SALVADORAN AMERICANS* 69 (2005) ("Since 1979, the influx of Salvadoran immigrants to the United States has risen at a high rate."); *LATINAS IN THE UNITED STATES: A HISTORICAL ENCYCLOPEDIA* 135 (Vicki L. Ruiz & Virginia Sanchez Korrol eds., 2006) ("Central American immigration increased exponentially [in the 1980s], quintupling the Salvadoran population [in the United States.]").

¹⁷ See Marie Price et al., *The World Settles In: Washington, DC, As an Immigrant Gateway*, 26 *Urban Geography* 61, 63 (2005) ("Unlike the more established urban immigrant destinations, the District of Columbia is not built upon a rich history of immigration and has only recently become an immigrant destination. Thus there are few historically ethnic immigrant neighborhoods or enclaves."); see also *The Rise of New Immigrant Gateways* (Brookings Feb. 2004).

¹⁸ See *EL SALVADOR: A COUNTRY STUDY* 67 (Richard A. Haggerty ed., 1990) ("observers have estimated that much of the Salvadoran population in the 1980s could be said to possess an Indian racial background").

¹⁹ See, e.g., Carolina Bonilla et al., *Ancestral Proportions and Their Association with Skin Pigmentation and Bone Mineral Density in Puerto Rican Women from New York City*, 115 *HUM. GENETICS* 57 (2004); Leonora Buentello-Malo et al., *Genetic Structure of Seven Mexican Indigenous Populations Based on Five Polymarker Loci*, 15 *AM. J. HUM. BIOLOGY* 23 (2003).

Further, the Salvadorans now residing in the D.C. metropolitan area have a markedly different ancestry than their countrymen who immigrated to other major U.S. destinations, such as Los Angeles and other parts of California. The population of Salvadorans that immigrated here came predominantly from the eastern part of El Salvador, from rural communities insulated from urbanized centers such as San Salvador in the west, where presumably the majority of genetic mixing would occur.²⁰ The Salvadoran immigrants who relocated to the California, on the other hand, came largely from major metropolitan areas in western El Salvador.²¹ The defendant and his fellow countrymen now calling the District of Columbia home remain genetically akin to the narrow subset of Salvadoran natives occupying that particular, insulated region in the east of El Salvador, unlike even other Salvadorans now living elsewhere in the United States.

Further amplifying the insulation and uniqueness of the genetic fabric from which the defendant and the other Salvadoran residents of this region are cut, immigrating families tend to follow family members who have gone before them.²² In other words, families tend to follow families, which results in tight clustering of genetic material particularly in “immigrant gateways” such as the District of Columbia. This effect is even more dramatic when the gateway is a new one, where the newcomer population has not yet had the opportunity to mingle genetically with other, more established populations in the region.

Arguments along these lines can be further substantiated with Census data and other research relating to migration and settlement patterns, and then contrasted with information relating to the source populations of the Y-STR database that served as the basis for the statistical requirement of the “match” evidence. With respect to the “Hispanic” branch of the YFiler database, it can be argued that neither eastern Salvadorans nor the Hispanic population of Washington, D.C. are represented. On that basis, any statistical conclusions based on a comparison of an eastern Salvadoran Y-STR profile to a database of genetically distinct “Hispanics” can be argued to be unreliable.²³

V. Defense Experts

If an admissibility challenge fails, a defense attorney may want to consider retaining her own expert, following the same considerations outlined in Section [___] above. Generally the

²⁰ See CORDOVA, *supra* note 16, at 78 (“This population [including that which immigrated to Washington, D.C.] is mainly rural, or coming from provincial Salvadoran cities and towns.”); *id.* (“large numbers of persons from the eastern part of El Salvador relocated in metropolitan centers in the East Coast” of United States, including Washington, D.C.).

²¹ *Id.* (“large numbers of urban dwellers and those with more education have relocated in the Los Angeles and San Francisco metropolitan areas. In San Francisco, for example, many people are from San Salvador, Sonsonate, and other major [western] provincial cities.”)

²² *Id.* at 78 (“These new immigrants arrived in the United States as a result of already established ethnic and family networks”).

²³ For a more comprehensive argument, see [APPENDIX; Sample MOTION IN LIMINE TO EXCLUDE EXPERT TESTIMONY CONCERNING Y-STR DNA “MATCH” EVIDENCE WHERE THE GOVERNMENT IS UNABLE TO PROVIDE A RELIABLE, GENERALLY ACCEPTED ESTIMATE OF THE LIKELIHOOD THAT THE “MATCH” IS COINCIDENTAL].

same strategic considerations apply: if the government called a forensic scientist, the defense may want to call a traditional scientist with expertise in the underlying science to challenge the claims put forth by the government expert if the necessary points cannot be sufficiently developed on cross. In a Y-STR case, the defense may additionally want to call, for example, another expert on the migration patterns of the defendant's ancestors to show that his haplogroup is not properly represented by the database the government relied upon for its statistical representation. Again, the use of experts is limited only by the imagination of defense counsel.

VI. Y-STR Treatment in the Courts

A number of challenges have been made to the admissibility of Y-STR evidence as an inculpatory tool, but generally without success thus far. Y-STR "inclusion" evidence has been admitted in at least nine states.²⁴ Fresh admissibility challenges are critical, however, in light of the wealth of new and ongoing research that calls the reliability of the Y-STR forensic databases, and the inculpatory claims hinging upon them, seriously into question.

²⁴ See *Curtis v. State*, 205 S.W.3d 656 (Tex. App. 2006) (finding Y-STR "inclusion" evidence sufficiently reliable under *Daubert*); *State v. Unsworth*, No. L-03-1189, No. L-04-1165 (Ohio App. Sept. 2, 2005) (admitting Y-STR evidence under *Daubert*); *State v. Sanders*, No. CR-2000 2900 (Ariz. Super. Ct. Dec. 16, 2003) (admitting Y-STR evidence, but limiting statistical characterization to number of occurrences of profile in database); *State v. Russell*, No. 05-1-02485-2 (Wa. Super. Ct. Jan. 2006) (finding Y-STR admissible under *Frye* with no need for new admissibility hearing); *State v. Avila*, No. 02CF1862 (Ca. Super. Ct. Feb. 17, 2005) (finding Y-STR from Y-PLEX kit and statistics based on ReliaGene database admissible under *Frye*); *State v. Temple*, No. 02040491 (Minn. Dist. Ct. Apr. 14, 2005) (finding Y-STR admissible under *Frye*); *State v. Polizzi*, 924 So.2d 303 (La. App. 2006) (admitting Y-STR without challenge); *Shabazz v. State*, 592 S.E.2d 876 (Ga. App. 2004) (same).