TRYING MITOCHONDRIAL DNA CASES

Mitochondrial DNA (mtDNA) is different from nuclear DNA (nDNA) – the default forensic variety – on a number of counts. A carefully challenge to mtDNA evidence should home in on those differences, in addition to following the general guidelines above. What follows is a synopsis of the core substantive areas of forensic mtDNA 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. A sample motion to exclude mtDNA "match" evidence is provided as an appendix for reference.

I. MtDNA Biology

The mtDNA genome is distinct from the nDNA genome in terms of its location in the body, as well as its size, variability, method of inheritance, and genetic makeup. Perhaps the most significant distinction is the manner in which mtDNA is inherited: rather than recombining a random selection of alleles from each parent at every generation, mtDNA is passed wholesale from the mother to all of her offspring. The result is that the discriminatory power of mtDNA pales in comparison to its nuclear cousin, as all members of a given maternal line will share the same sequence, barring mutations, from generation to generation, future and past.[1]

Any attempt by the government to characterize a purported mtDNA "match" as evidence of identity is scientifically unsupportable on the basis of its method of inheritance alone, and that fact should be spotlighted to any judge or jury considering such evidence. Indeed, for this reason, scientists are in agreement that "Mitochondrial DNA typing does not provide definitive identification."[2]

MtDNA also differs with respect to its manner of storage in the body. Mitochondrial DNA is found outside the nucleus of a cell in energy-producing organelles called mitochondria, where one's entire mtDNA genome occurs in abundance – up to thousands of individual, complete copies[3] – in comparison to the singular set of nDNA found in the nucleus.[4] Because it is present in such dramatically higher quantities, mtDNA can often be detected in old or badly degraded samples, as well as bone, teeth, and hair, when nuclear DNA is not.[5]

---

[1] The mutation rate for mtDNA is, however, significantly higher than that of nDNA.

[2] Thomas J. Parsons & Michael D. Coble, *Increasing the Forensic Discrimination of Mitochondrial DNA Testing through Analysis of the Entire Mitochondrial DNA Genome*, 42(3) CROATIAN MED. J. 304, 304 (2001).

[3] The average has been estimated at approximately 500 copies for most cells. *See* M. Satoh & T. Kuroiwa, *Organization of Multiple Nucleoids and DNA Molecules in Mitochondria of a Human Cell*, 196 EXPERIMENTAL CELL RES. 137 (1991).

[4] The are technically two "copies" of nDNA in the nucleus of a cell – one from the mother and one from the father – but each "copy" comprises only half of the set required for forensic typing.

[5] JOHN M. BUTLER, FORENSIC DNA TYPING: BIOLOGY, TECHNOLOGY, AND GENETICS OF STR MARKERS 241 (2d ed. 2005).

While numerous copies of the mtDNA genome are present in each cell, that genome only accounts for about 0.25% of the total DNA content of a cell. MtDNA is far more abundant in quantity than nuclear DNA, but is far less so in information density. Indeed, an individual's mtDNA sequence has approximately 16,569 base pairs in total, as compared to 3.2 *billion* base pairs of nuclear DNA.[6]

MtDNA also differs from nuclear DNA insofar as most individuals are "heteroplasmic" with respect to their mtDNA genomes, meaning that individuals have more than one mtDNA genome. A person's mtDNA sequence can differ at various locations within the body, from tissue type to tissue type, or even within the same tissue.[7] Certain bodily tissues, such as hairs – often used for forensic mtDNA analysis – tend to exhibit higher levels of mtDNA variation than others.[8] Some research suggests that the occurrence of heteroplasmy in an individual increases with age, while other data appear to contradict that finding.[9] Nonetheless, while the cause or causes of mtDNA heteroplasmy are not fully understood, its existence has been widely observed and is not disputed.[10]

II. Forensic Application of MtDNA

Mitochondrial DNA profiling was originally developed outside of the criminal context as a research tool for population studies in various scientific disciplines, as its relative integrity across generations proved an effective means of deriving certain information about groups identified by a common mtDNA sequence.[11] Law enforcement, however, sought to adapt the practice of mtDNA profiling – originally found useful for the very fact that the same profile is shared by large groups of individuals – as an aid to criminal investigation, as an *inculpatory* tool. A collection of problems emerged from the improbable adaptation of a genetic feature known equally for its facility for identifying population groups and for its malleability within individuals, to the law enforcement effort to identify particular individuals as the source of evidence at a crime scene.

---

[6] BUTLER, *supra* note 5, at 242.

[7] *See* John Buckleton, Simon Walsh, & Sallyann Harbison, *Nonautosomal Forensic Markers*, *in* JOHN BUCKLETON, CHRISTOPHER M. TRIGGS & SIMON J. WALSH, EDS., FORENSIC DNA EVIDENCE INTERPRETATION 304 (2005) [hereinafter Buckleton et al. (2005)].

[8] Tomasz Grzybowski, *Extremely High Levels of Human Mitochondrial DNA Heteroplasmy in Single Hair Roots*, 21 ELECTROPHORESIS 548 (2000).

[9] Buckleton et al., *supra* note 11, at 305.

[10] *See* Buckleton et al., *supra* note 7, at 303; Peter D'Eustachio, *High Levels of Mitochondrial DNA Heteroplasmy in Human Hairs by Budowle et al.*, 130 FORENSIC SCI. INT'L 63, 63 (2002) ("Major unresolved issues include the molecular mechanisms responsible for the occurrence of heteroplasmy to different extents in different tissues...").

[11] *See, e.g.*, Douglas C. Wallace, *Mitochondrial Disease in Man and Mouse*, 283 Sci. 1482 (1999); Mark Shriver & Rick Kittles, *Genetic Ancestry and the Search for Personalized Genetic Histories*, 5 NATURE REV. GENETICS 611 (2004); Rebecca L. Cann et al., *Mitochondrial DNA and Human Evolution*, 325 NATURE 31 (1987).

The nature of mtDNA places rigid limits on its forensic adaptation, which a number of U.S. courts have acknowledged.[12] The relatively small mtDNA genome consists of two primary regions, one of which is a non-coding region and thus deemed suitable for forensic use.[13] The non-coding region of the mtDNA genome is known as the *control region*, which regulates replication of the mtDNA molecule.[14] The forensically viable control region is only 1125 base pairs long,[15] in contrast to the abundance of non-coding DNA that is dispersed throughout the 3.2 billion base pairs in the nuclear DNA molecule.

To distinguish one individual's mtDNA from another, forensic scientists look to specific locations within the control region that are known to be highly variable among humans, as they would with nDNA. In the context of mtDNA, however, only two such regions are currently used[16] for forensic identification due to their observed variability from person to person: "Hypervariable Region I" ("HVI") and "Hypervariable Region II" ("HV2"), which together encompass a total of only about 610 base pairs.[17] An individual's mtDNA "profile" is a list of the differences between the sequences observed in those regions, and those in a reference sequence known as the revised Cambridge Reference Sequence ("rCRS").

Those 610 bases, however, provide something far short of the discriminatory power of nuclear DNA, with its 13 (or more) distinct and genetically independent locations that have been developed to form the statistical basis of compelling forensic identification evidence. In fact, the 610 base pairs available for mtDNA analysis are tantamount to a "single genetic locus."[18] Most courts would balk at a prosecutor seeking to introduce nuclear DNA "match" evidence at a single genetic location, because the chance of two unrelated individuals matching by coincidence at a single location is in most cases exceedingly high. Defense attorneys should insist that the same standards be applied to mitochondrial DNA, rather than conceding to a standard that is lowered to accommodate the inherent limitations of that type of evidence.

---

[12] *See, e.g.*, *Vaughn v. State*, 646 S.E.2d 212, 214 (Ga. 2007) (observing that "mtDNA analysis is more applicable for exclusionary, rather than identification, purposes," but admitting evidence nonetheless); *Wagner v. State*, 864 A.2d 1037, 1045 (Md. App. 2005) ("MtDNA analysis provides significantly less ability to discriminate among possible donors than does nuclear DNA analysis and has been said to be a test more of exclusion than of identification."); *State v. Scott*, 33 S.W.3d 746, 756 (Tenn. 2000) ("Because it is not possible to achieve the extremely high level of exclusion provided by nuclear DNA, mtDNA typing has been said to be a test more of exclusion than one of identification.").

[13] It has been generally agreed among the forensic community that DNA typing should use only "junk" DNA, or regions on an individual's DNA molecule that do not encode for traits, as a protection of privacy interests as the revelatory power of DNA continues to expand.

[14] *Human Mitochondrial DNA – Amplification and Sequencing Standard Reference Materials* 1-2, NAT'L INST. OF STANDARDS AND TECH SPEC. PUB. NO. 260-155 (Sept. 2003).

[15] *See* Stephen Anderson et al., *Sequence and Organization of the Human Mitochondrial Genome*, 290 Nature 457, 457-65 (1981).

[16] There is a third highly variable region ("HV3") that has been studied, but which is not currently used by forensic laboratories.

[17] Mitchell M. Holland & Thomas J. Parsons, *Mitochondrial DNA Sequence Analysis: Validation and Use for Forensic Casework*, 11 FORENSIC SCI. REV. 21, 24 (1999).

[18] Department of Justice, *Mitochondrial DNA Analysis at the FBI Laboratory, Forensic Science Communications* (July 1999).

A. MtDNA Typing Problems

i. MtDNA Is Extremely Prone to Contamination

At the outset of typing process, mitochondrial DNA analysis presents a far higher risk of contamination than nDNA because of the need for many more stages of PCR amplification for a single sample. To conduct mtDNA analysis, technicians must make very small amounts of DNA visible, which entails amplifying a sample 10^{20} times more than is required of nDNA. The risk of contamination in mtDNA analysis is substantial, and has been acknowledged by law enforcement:

> The most critical potential source of error in mtDNA sequencing is contamination. If more than one individual's DNA is extracted and amplified, the sequencing results will reflect this mixture. In extreme cases the contaminating DNA can greatly exceed the DNA from the donor, and thereby yield a false positive result.[19]

To address this issue, the FBI has established a "contamination ratio" of 10:1, meaning that the FBI considers one part contamination per 10 parts mtDNA sample to be acceptable.[20] The FBI arrived at this figure through a single test of five samples, a size consistent with a 35% laboratory error rate.[21] Drs. Wilson and Budowle, both researchers in the FBI's laboratory division, have conceded that the FBI's method of assessing acceptable contamination rates in the laboratory is one that is not followed by any other DNA testing laboratory in the world. Defense attorneys should question whether the FBI has sufficient data to permit the use of the 10:1 contamination ratio.

The problems with mtDNA contamination are real, and well-documented, and should be included in a defense challenge to such evidence. A commercial laboratory in 2001 reported the results of a two-year study over thousands of trials to determine the effect of heteroplasmy, contamination, and other factors on laboratory test results.[22] With respect to contamination, the researchers found the presence of contaminants in 2.4% of cases.[23] The source of the contaminated DNA was not laboratory staff; the researchers determined the contaminants came from a source outside of the laboratory. In at least two cases, the contamination affected the interpretation of results.[24]

---

[19] Mark Wilson, Bruce Budowle, & Mitchell Holland, *Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science* (1993).

[20] *See* Constance L. Fisher et al., *Mitochondrial DNA: Today and Tomorrow, presented at the* Eleventh Annual Int'l Symposium on Human Identification, at 1 (2000) [hereinafter Fisher et al. (2000)].

[21] See Mark Wilson et al., *Extraction, PCR Amplification and Sequencing of Mitochondrial DNA from Human Hair Shafts*, 18 BioTechniques 662 (1995).

[22] *See* Terry Melton & Kimberlyn Nelson, *Forensic Mitochondrial DNA Analysis: Two Years of Commercial Casework Experience in the United States*, 42 CROATIAN MED. J. 298 (2001).

[23] *Id.* at 300.

[24] *Id.*

The debate in the scientific community concerning the acceptable level of contamination in mtDNA analysis is far from over. The absence of any independently determined standards other than the FBI's 10:1 ratio cannot satisfy either the general acceptance standard in *Frye* jurisdictions or the reliability requirement under *Daubert*.

ii. The Role of Heteroplasmy in Forensic Typing of MtDNA

Forensic analysts take account of the biological anomalies of mitochondrial DNA by applying a more flexible standard for declaring a match between two profiles. Whereas a "match" between nuclear DNA profiles is straightforward in non-mixture cases, it is less so with the idiosyncratic and heteroplasmic mtDNA. At the outset, a DNA analyst sequences the HVI and HVII regions of the evidence sample, and then does the same for the suspect's sample. Most laboratories, however – in accordance with their own protocols – do not automatically exclude the suspect as the source of the evidence sample if the two profiles objectively fail to match. Instead, most law enforcement agencies "will only definitively exclude a suspect if there are two or more base pair differences between the samples with no evidence of heteroplasmy, on the theory that one difference may be the result of heteroplasmy."[25]

A defense challenge to purported mtDNA "match" evidence might begin by highlighting the fluidity of definition that law enforcement agencies have adopted to accommodate the biological realities of that type of evidence. Law enforcement agencies maintain the untenable position that a match is always a match, and a non-match sometimes is too. A protocol that not only allows but *requires* interpretation of apparently exculpatory evidence in a manner that renders the same evidence inculpatory invites a new level of influence to analyst bias. Indeed, bias itself is institutionalized by a protocol that requires interpretation of non-matching forensic evidence samples as matching, when the latter interpretation fits the government theory of culpability.

More specifically, while law enforcement agencies will only declare an automatic exclusion in those cases in which two or more differences are observed between the evidence sample and the suspect sample, they will declare an "inclusion" (or "failure to exclude") under a variety of scenarios. For example, if an analyst concludes that the profiles being compared are identical at each of the bases in the HVI and HVII regions, the suspect will be deemed "included" as a possible contributor of the evidence sample. [26] If either the suspect or evidence sample displays heteroplasmy, the analyst will also not exclude the suspect as a possible source.[27] And, if the two profiles differ by a single base pair but neither profile appears heteroplasmic, the analyst will nonetheless not deem the suspect excluded as a possible contributor, and instead characterize the results as "inconclusive," notwithstanding the objectively different profiles.[28] Rather than sequence the mtDNA genome outside of the HVI

---

[25] Frederick A. Kaestle et al., *Database Limitations on the Evidentiary Value of Forensic Mitochondrial DNA Evidence*, 43 AMER. CRIM. L. REV. 53, 62 (2006).

[26] FBI Laboratory DNA Unit II, Mitochondrial DNA Sequencing Protocol (2004) [hereinafter FBI MtDNA Protocols (2004)], at § 11.3.3.

[27] *Id.*

[28] *Id.*

and HVII regions to validate the remarkable claim that objectively different profiles may come from the same source (either in HV3 or in the coding region), law enforcement agencies simply report that the suspect cannot be excluded as the source of the evidence sample.

To complicate things further, the exact criteria for distinguishing a "failure to exclude" from an actual exclusion vary from laboratory to laboratory with no standardization.[29] A defense challenge should spotlight the lack of agreement on this core question in any forensic identification analysis, namely, what constitutes a match between two specimens.

B. MtDNA Database Problems

After identifying a "match" on one of the theories above, a mtDNA analyst will next compare that sample to a database of profiles for the purpose of deriving its significance.[30] MtDNA also differs from nuclear DNA with respect to the integrity of the population databases relied upon for the statistical expression typically required prior to admission of DNA "match" evidence in U.S. courts, which is an area ripe for defense challenges.[31] As with all types of DNA evidence, the government generally must provide a reliable statistical expression of the likelihood that a purported "match" between profiles is a result of coincidence, rather than identity. In order to derive that statistic representing the likelihood that the evidence actually came from someone other than the suspect – without which DNA evidence is wholly devoid of meaning – a database of mtDNA profiles constituting a representative sample of the relevant population is essential.

An admissibility challenge to mtDNA evidence can focus principally on the well-documented shortcomings of the databases relied upon to derive the statistical expression that serves as a prerequisite to admission of the evidence. SWGDAM, under the direction of the FBI, maintains the only mtDNA database in the United States that is used for forensic analysis. Numerous problems have been observed in those databases, which serve as the sole basis for claims regarding the significance of purported "matches" between mtDNA profiles.

In order to assess the reliability of the FBI mtDNA database, a team of scientists selected the so-called "African-American" mtDNA sub-database to serve as the basis of a "thorough inspection."[32] Upon doing so, Dr. Bandelt and his colleagues found the SWGDAM database to be infected with "a number of major deficiencies."[33] Those and other deficiencies undermine the

---

[29] *Id.* (citing Statement of Dr. M. Thomas P. Gilbert, submitted in *United States v. Chase*, D.C. Super. Ct. Crim. No. F-7330-99 (July 9, 2004) (reviewing protocols for all major mtDNA testing laboratories and observing that "forensic laboratories come to no consensus as to how to interpret heteroplasmic sequences.... [T]he interpretation guidelines vary when determining what would be labeled as 'inconclusive' and what would be labeled as an 'exclusion.'").

[30] Technically, only the differences between the sample and the reference (CRS/Anderson) sequence are compared against the database profiles. Alice R. Isenberg & Jodi M. Moore, *Mitochondrial DNA Analysis at the FBI Laboratory*, 1 Forensic Sci. Comm. 1 (1999), *available at* http://www.fbi.gov/hq/lab/fsc/backissu/july1999/dnalist.htm.

[31] *See, e.g.*, *United States v. Porter*, 618 A.2d 629 (D.C. 1992) (requiring expression of statistical significance of DNA "match" as prerequisite to admission); *People v. Axell*, 235 Cal.App.3d 836 (2001) (same); [MORE CITES].

[32] H.J. Bandelt et al., *Problems in FBI MtDNA Database*, 305 SCIENCE 1402, 1403 (2004).

[33] *Id.*

ability of any of the mtDNA databases to serve as a reliable basis for any claims relating to the significance of mtDNA "match" evidence.

i. Entry and Typing Errors and the Lack of Sufficient Quality Control

Numerous problems have been identified relating to the accuracy of the specific profiles that comprise the SWGDAM database. Over the several years since the development of the SWGDAM mtDNA database, scientists have been critical of the accuracy of its contents based on observations of a collection of objective errors relating to the integrity of the individual profiles that populate that database.

Interpretation plays a critical role in typing an individual mtDNA genome, and the lack of sufficient quality control standards governing the initial typing of the profiles that now comprise the SWGDAM mtDNA database have infected its contents with an indeterminate number of errors. Molecular biologists Yao et al. identified "five major and common types of errors, namely, base shifts, reference bias, phantom mutations, base mis-scoring, and artificial recombination," which plague forensic mtDNA databases.[34] Following their analysis of existing databases, they described the need for quality control standards to be "urgent," and recommended "[e]xtreme caution…at all stages of data collection and proof-reading processes."[35]

In their analysis of the "African-American" sub-population database, Bandelt et al. "detected as many as five artificial combinations of totally unrelated mtDNA segments stemming from different samples, which suggest fatal sample mix-up in the lab or during data transcription."[36] Even following their report to the lab and a series of revisions by the FBI in response, "Several obvious clerical errors still remain in the revised database."[37] According to Bandelt et al., the remaining errors "could only be corrected through thorough resequencing of the original samples"[38] – a step which has never been attempted by the FBI, despite their knowledge of the persistent errors in these population databases that serve as the cornerstone of mtDNA forensic analysis.

If the individual profiles populating the SWGDAM mtDNA database do not accurately reflect the genomes of the individuals who contributed those samples, even the modest claims that law enforcement would hinge on that database are fundamentally invalid. A defense challenge should highlight the multiple sources of error in the typing and entry of mtDNA profiles into the SWGDAM database, as well as the documentation of persistent errors by disinterested scientists that remain unresolved.

ii. The MtDNA Databases Are Not Representative

---

[34] Yong-Gang Yao, Claudio M. Bravi, & Hans-Jürgen Bandelt, *A Call for MtDNA Data Quality Control in Forensic Science*, 141 FORENSIC SCI. INT'L 1, 1 (2004).

[35] *Id.* at 4.

[36] *Id.* at 1.

[37] *Id.*

[38] *Id.*

Paramount among the deficiencies of the SWGDAM mtDNA database that serves as the basis for the statistical information provided to juries is that, even if accurately typed, it is not representative of the populations it purports to represent. The entire SWGDAM mtDNA database contains only 5,071 profiles, which are divided across fourteen so-called "racial" sub-populations.[39] Some of those racial groups are represented by fewer than a hundred profiles, down to as few as eight. The profiles comprising these databases were gathered from a collection of blood banks, paternity testing outfits, scientific research groups, and even FBI agents.[40] The racial classifications are based not on genetic ancestry, but on the self-reporting of the individuals whose samples were collected. The samples are not geographically diverse or randomly selected, and the FBI does not claim otherwise. The SWGDAM mtDNA database is an example of a "convenience sample," obtained from only a handful of locations, where no effort was made to randomize the selection.[41] By definition, the databases are not representative of the genetic makeup of their respective populations.

In nuclear DNA analysis, forensic scientists derive typically miniscule random match probability statistics on the basis of a comparison of alleles at each of the thirteen standard STR locations.[42] The nDNA STR databases, and the validation studies thereof, were based on the premise that, particularly given the miniscule associated statistics, only statistically insignificant genetic linkage exists within the populations sampled. Accordingly, for nDNA, it was concluded that sampling from a small number of locations was acceptable.[43]

---

[39] The populations and the number of profiles within each are as follows:

Race	Number of Profiles
African Americans	1148
Apaches	180
Caucasians	1814
Chinese/Taiwanese	356
Egyptians	48
Guam	87
Hispanics	759
India	19
Japanese	163
Koreans	182
Navajos	146
Pakistan	8
Sierra Leone	109
Thai	52
TOTAL	5071

[40] Bruce Budowle et al., *Mitochondrial DNA Regions HVI and HVII Population Data*, 103 Forensic Sci. Int'l 23, 25 (1999).

[41] *Id.*

[42] *See* BUTLER (2005), *supra* note 5, at 502 ("Often the rarity of a calculated [nuclear] DNA profile goes beyond one in billions (10^9) or trillions (10^{12}) to numbers that are not frequently used because they are so large."); *id.* at 504, Table 21.3 (listing values such as quadrillion (10^{15}), quintillion (10^{16}), and googol (10^{100})).

[43] *See* NATIONAL RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 30 (1996) [hereinafter NRC II (1996)] (STR database consists of convenience samples from "blood banks, paternity-testing laboratories,

This assumption of randomness does not apply with respect to mtDNA sequences. Because mtDNA is maternally inherited and not recombinant, mtDNA profiles are not randomly distributed across the population. The distribution of a given mtDNA sequence is a function of nothing more than the migration of women. An individual and all of his siblings, in addition to their mother, and grandmother, and great-grandmother, and maternally related third-cousins, are expected, absent mutations, to share identical mtDNA profiles. Over generations, profiles stay intact or mutate to a very similar sequence. In addition, the high mutation rates characteristic of the HVI and HVII regions create unique variants, including more recently created ones that have not had time to spread from their location of origin. This creates geographical areas where certain haplogroups or haplotypes are prevalent, and other areas where those same haplogroups and constituent haplotypes are wholly or largely nonexistent.

The "frequency estimates" that are derived from these non-random mtDNA databases are something quite different from those associated with nuclear DNA, which purport to estimate the frequency of a particular profile in actual populations. To generate a statistical expression of the significance of a mtDNA match, on the other hand, a forensic analyst simply counts the number of occurrences of the profile in question in each of the sub-population databases, which is known as the "counting method."[44] In more cases than not, given that the database excludes the vast majority of actual mtDNA profiles, the "count" of a given profile is observed to be "zero" out of the number of profiles in a given sub-population database.

Where a profile is observed at least once, the conventional statistical calculation involves dividing the number of observations by the size of the database.[45] For example, if the profile were observed once in the African-American database, (n = 1148), the frequency would be reported as 1/1148 or 0.0008711. The analyst would then place a ninety-five percent confidence interval around that number as a margin of error in estimating the frequency in the larger population,[46] and the laboratory would report the upper-bound frequency. For an observed frequency of 0.0008711, the upper confidence limit is 0.004839, or 0.48%, and the laboratory would report that the suspect is one of less than one-half of 1% of individuals included as a potential contributor, and that 99.52% of all African-Americans are excluded.[47]

---

laboratory personnel, clients in genetic-counseling centers, law-enforcement officer, and people charged with crimes."); *United States v. Bridgett*, 120 Wash. L. Rptr. 1697, 1700 & n.12 (Aug. 11, 1992) (same).

[44] FBI MtDNA Protocols (2004), *supra* note 26, at § 11.1.

[45] Laboratories use a slightly different statistical calculation when the sequence is not observed in the database. *See* Holland & Parsons (1999), *supra* note 17, at 31-32.

[46] A ninety-five percent confidence interval means that, if a series of such margins of error were constructed in estimating the frequency of the sequence in the population, approximately ninety-five percent of them should include the true frequency of the sequence in the population. Alternatively stated, there is approximately a five percent chance that the margin of error does not contain the true frequency of the sequence in the population. *See* ROBERT S. WITTE, STATISTICS 215 (2d ed. 1985) [hereinafter WITTE (1985)]. As the sample size grows, the confidence interval will become narrower, indicating ninety-five percent confidence in a smaller range of possible values for the frequency . *Id.* at 216.

[47] Kaestle et al., *supra* note 25, at 64-65.

1. *Phylogeographic Studies Confirm That MtDNA Profiles Are Not Evenly Distributed*

Dozens of phylogeographic[48] studies have been performed to identify the geographic distribution of mtDNA haplotypes in countries all over the world, although these studies are extremely limited in the United States. These studies demonstrate that mtDNA is not randomly distributed and that different haplogroups (collections of ancestrally related mtDNA genomes) and haplotypes (individual mtDNA genomes) are concentrated within certain populations that vary geographically.[49] Scientists rarely encounter new nDNA haplotypes when studying new population subgroups, but the opposite is true for mtDNA. While certain haplogroups of mtDNA sequences are widely distributed throughout the population,[50] many exist only within certain geographic clusters.[51] Non-random MtDNA haplotype distributions also exists *within* geographic locations, because of often subtle linguistic, religious, or economic/caste distinctions.[52]

Distinctive mtDNA haplotype distributions are not limited to rare or ancient populations; today, different geographic regions demonstrate strikingly different mtDNA patterns.[53] For

---

[48] Phylogeography "is a field of study concerned with the principles and processes governing the geographic distributions of genealogical lineages, especially those within and among closely related species [and] deals with historical, phylogenetic components of the spatial distributions of gene lineages. In other words, time and space are the jointly considered axes of phylogeography onto which (ideally) are mapped particular gene geneaologies of interest." JOHN C. AVISE, PHYLOGEOGRAPHY: THE HISTORY AND FORMATION OF SPECIES 3 (Harvard Univ. Press 2000).

[49] *See, e.g.*, Dan Mishmar et al., *Natural Selection Shaped Regional MtDNA Variation in Humans*, 100 PROC. NATL. ACAD. SCI. 171 (January 7, 2003) ("extensive global population studies have shown that there are striking differences in the nature of the mtDNAs found in different geographic regions").

[50] *See* Martin Richards et al., *In Search of Geographical Patterns in European Mitochondrial DNA*, 71 AM. J. HUM. GENETICS 1168, 1170 (2002) [hereinafter Richards (2002)].

[51] While *phylogenetic* analysis – reconstructing genetic relationships within a population – has been conducted on many of the SWGDAM racial sub-databases, such studies only show, at most, that a particular database accurately reflects most of the haplogroups that exist in the relevant population, e.g., that the Caucasian database contains all major haplogroups in the Caucasian population. *See* Allard et al. (2002), *supra* note 23, at 8. Such studies do not, however, take into account the *geographical distribution* of the sequences within the population, and thus cannot be cited as evidence that a database accurately reflects the frequency of a profile in a particular geographic area. Only *phylogeographic* studies – those that focus on the spectrum and area-specificity of major haplogroups and the haplotypes within them – can accurately determine true frequencies. *See* J.C. Rando et al., *Phylogeographic Patterns of MtDNA Reflecting the Colonization of the Canary Islands*, 63 ANNALS HUM. GENETICS 413, 424 (1999) [hereinafter Rando et al. (1999)].

[52] *See, e.g.*, Michael Bamshad et al., *Genetic Evidence on the Origins of Indian Caste Populations*, 11 GENOME RESEARCH 994 (2001) (discussing economic and caste distinction); R. Dutta et al., *Patterns of Genetic Diversity at the Nine Forensically Approved STR Loci in the Indian Populations*, 74 HUM. BIOLOGY. 33 (2002) (same); D. Andrew Merriweather et al., *Mitochondrial DNA is an Indicator of Austronesian Influence in Island Melanesia*, 110 AM. J. PHYS. ANTHROPOL. 243 (1999) (linguistic distinctions); P. Rudan et al., *Anthropological Research of Hvar Islanders, Croatia – From Parish Registries to DNA Studies in 33 Years*, 28 COLLEGIUM ANTHROPOLOGICUM 321 (2004) (religious); Lev A. Zhivotvsky et al., *The Forensic DNA Implications of Genetic Differentiation Between Endogamous Communities*, 119 FORENSIC SCI. INT'L 269 (2001) (no obvious subdivision).

[53] *See, e.g.*, DAVID BALDING, WEIGHT-OF-EVIDENCE FOR FORENSIC DNA PROFILES 105-06 (2005) [hereinafter BALDING (2005)] ("*[M]aternally-related individuals might be expected to be tightly clustered, possibly on a fine geographical scale.* Reports of Fst estimates for mtDNA drawn from cosmopolitan European populations typically cite low values, reflecting the fact that this population is reasonably well-mixed, as well as the effects of high mtDNA mutation rates. However, researchers rarely are able to focus on the fine geographic scale that may be

example, a particular cluster of mtDNA sequences called haplogroup J is widely distributed in western and central Europe, but is rare in Iberia.[54] A sub-haplogroup of that cluster has been observed primarily in Britain, with one other occurrence from an ancestor in Italy.[55] A mutation that has an eight percent frequency within the Canary Islands has never been found outside the Islands.[56] One study related to the natives of Mozambique as compared to those in the Americas identified a considerable number of matches between Mozambique and American sequences from African haplogroups, including some sequences that had never been observed outside Mozambique, as well as others observed only in the American populations.[57]

MtDNA population genetic linkage in North America – discussed in detail in the next two sections – is also well documented in scientific research.[58] Whether the heterogeneous

---

relevant in forensic work, and there are some large Fst estimates at this scale.") (emphasis added); A. Brandstatter et al., *Mitochondrial DNA Control Control Region Sequences from Nairobi (Kenya): Inferring Phylogenetic Parameters for the Establishment of a Forensic Database*, 118 INT'L J. LEGAL MED. 294 (2004) (describing new forensic database containing sequences from Nairobi and finding that there were significant differences in mtDNA compositions of this new database and the African-American SWGDAM database, as well as of published sequences from Sierra Leone, Mozambique, and United States); Forster et al., *Continental and Subcontinental Distributions of MtDNA Control Region Types*, 116 INT'L J. LEGAL MED. 99-108 (2002); Frederika A. Kaestle & K. Ann Horsburgh, *Ancient DNA in Anthropology: Methods, Applications, and Ethics*, 119(S35) AMER. J. OF PHYSICAL ANTHROPOLOGY 92, 95 (2002) ("[M]itochondrial markers are also often geographically specific, and in some cases are limited in distribution to a single tribe (private polymorphisms)."); R. Kittles & S.O.Y. Keita, *Interpreting African Genetic Diversity*, 16 AFRICAN ARCHEOLOGY REV. 87-91 (1999) [hereinafter Kittles & Keita (1999)]; Pereira et al., *Prehistoric and Historic Traces in the MtDNA of Mozambique: Insights into the Bantu Expansions and the Slave* Trade, 65 AM. J. HUM. GENETICS 439-458 (2001); J.C. Rando et al., *Phylogeographic Patterns of MtDNA Reflecting the Colonization of the Canary Islands*, 63 Annals Hum. Genetics 413, 424 (1999) [hereinafter Rando et al. (1999)]; Antonio Salas et al., *The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade,* 74 Am. J. Hum. Genetics 454-65 (2004) [hereinafter Salas et al. (2004)]; Y.G. Yao et al., *Phylogeographic Differentiation of Mitochondrial DNA in Han Chinese*, 70(3) AM. J. HUM. GENETICS 635, 649 (2002).

[54] M.B. Richards et al., *Phylogeography of Mitochondrial DNA in Western Europe*, 62(3) ANNALS OF HUM. GENETICS 241, 255 (1998) (discussing J Haplogroup).

[55] *Id.* at 254 (discussing J1b1 Haplogroup).

[56] Rando et al. (1999), *supra* note 53, at 420, 424 (1999).

[57] Luisa Pereira et al., *Prehistoric and Historic Trace in the MtDNA of Mozambique: Insights into the Bantu Expansions and the Slave Trade,* 65 AM. J. HUM. GENETICS 439, 451, 452 (2001) ("There remain a large number of sequences from African haplogroups sampled in the Americas and Europe for which no match can be found in the current African database. This may be due in part to the fact that the main regions from where slaves were taken, such as Angola and the Slave Coast remain uncharacterized.") (citation omitted). *See also* Joseph Lorenz et al., *African-American Lineage Markers: Determining the Geographic Source of MtDNA and Y Chromosomes*, presented at 73rd Annual Meeting of the American Association of Physical Anthropologists, Tampa, FL, Apr. 15, 2004, available at <http://www.physanth.org> (discussing study suggesting that there is large proportion of unexamined, undocumented mtDNA variability among individuals indigenous to sub-Saharan Africa).

[58] *See, e.g.,* Jason Eshleman, Ripan S. Malhi, & David G. Smith, *Mitochondrial DNA Studies of Native Americans: Conceptions and Misconceptions of the Population Prehistoric of the Americas*, 12 EVOLUTIONARY ANTHROPOLOGY 7-18 (2003) (noting that while Haplogroup X is found in low frequency in Europe and Western Asia, Native American variant is significantly different, possessing mutation that distinguishes it from Old World versions); Lynn B. Jorde & Stephen P. Wooding, *Genetic Variation, Classification, and "Race,"* 36 NATURE GENETICS S28, S29 (Nov. 2004) ("[I]ndividuals tend to cluster according to their ancestry or geographic origin."); Ripan S. Malhi et al., *The Structure of Diversity Within New World Mitochondrial DNA Haplogroups: Implications for the Prehistory of North America*, 70(4) AM. J. HUM. GENETICS 905 (2002) (Native Americans have haplogroups whose frequencies vary greatly among Canada, United States, and Mexico); Esteban J. Parra, Rick A. Kittles et al., *Ancestral Proportions and Admixture Dynamics in Geographically Defined African Americans Living in South Carolina*, 114 AM. J. PHYSICAL ANTHROPOLOGY 118 (2001) [hereinafter Parra & Kittles (2001)]; Estaban J. Parra,

geographic distribution of mtDNA lineages reflects genetic clustering, inadequate sampling, or some combination of the two, it appears clear that the sampling of mtDNA profiles must take into account geographic heterogeneity and stratification in order to create representative databases for use in forensic typing. This concern applies to the various ethnic groups in the United States – whether African American, Hispanic, Asian, or otherwise (though less so for Caucasians).

a. African-American Population

The need to account for migration patterns is critical when considering the experience of African Americans living in the United States. The oldest mtDNA profiles stem from Africa, whose population displays great regional diversity and heterogeneity in mtDNA profiles.[59] In some regions, specific mtDNA profiles are common; in others, the same mtDNA profiles are rare or nonexistent.[60] Scientific studies in Africa repeatedly uncover more unknown and previously unexamined mtDNA sequences, and far more is left to learn about regional differences that exist both now and hundreds of years ago.

During the period of slavery in the United States, the forced migration of Africans to the New World brought these regional differences to the United States and led to significant regional differences in the ethnic and geographic ancestry of African Americans.[61] Various political, economic, and cultural factors associated with the implementation of slavery contributed to these regional differences. For instance, during the period of slavery in the South, plantation owners in South Carolina primarily grew rice. These owners sought West Africans who already knew how to grow rice and therefore imported enslaved Africans from the "Grain Coast" of Africa.[62] In contrast, in Virginia, plantation owners primarily sought to grow tobacco.[63] The area surrounding the tobacco farms was swampy, and with the swamps, mosquitoes and malaria were

---

Amy Marcini et al., *Estimating African-American Admixture Proportions by Use of Populations-Specific Alleles*, 63 AM. J. HUM. GENETICS 1839 (1998) [hereinafter Parra & Marcini (1998)]; Sarah A. Tishkoff & Kenneth K. Kidd, *Implications of Biogeography of Human Populations for "Race" and "Medicine,"* 36 NATURE GENETICS S21, S26 (Nov. 2004) [hereinafter Tishkoff & Kidd (2004)] (frequency of mtDNA haplogroups are unevenly distributed within and among geographic regions and "knowledge of ethnicity (not just broad geographic ancestry) and statistical tests of substructure are important proper design of case control association studies"). *Cf.* Terry Melton et al., *Diversity and Heterogeneity in Mitochondrial DNA of North American Populations*, 46 J. Forensic Sci. 46 (2001) (while arguing that North American population is homogeneous, identifies, without exploring, population of Hispanics in Pennsylvania who differed significantly from any other population in study).

[59] Rebecca L. Cann, Mark Stoneking, & Allan C. Wilson, *Mitochondrial DNA and Human Evolution*, 325 NATURE 31 (1987). *See also* PHILIP D. CURTIN, THE ATLANTIC SLAVE TRADE: A CENSUS (U. Wisc. Press 1969; Lovejoy 2d ed. 1994) [hereinafter CURTIN (1969)]. Curtin's calculations were later refined by David Northrup. DAVID NORTHRUP, THE ATLANTIC SLAVE TRADE (1994). *See also* E. Watson et al., *MtDNA Sequence Diversity in Africa*, 59 AM. J. HUM. GENETICS 437 (1996).

[60] *See, e.g.*, Terry Melton et al., *Extent of Heterogeneity in Mitochondrial DNA of sub-Saharan African Populations*, 42 FORENSIC SCI. INT'L 582, 588-89 (1997) (finding numerous haplotypes with SSO frequencies of greater than ten percent in particular African population and "substantial subpopulation heterogeneity" in "continental African populations"). The authors conclude that "control region sequencing would be a good alternative for forensic identifications in African or African-derived populations where there is uncertainty about whether subpopulations are present, at least until further populations are studied." *Id.* at 589.

[61] *See generally* Salas et al. (2004), *supra* note 53, at 455-56.

[62] Parra & Kittles (2001), *supra* note 58, at 19.

[63] PHILLIP D. MORGAN, SLAVE COUNTERPOINT: BLACK CULTURE IN THE EIGHTEENTH CENTURY CHESAPEAKE AND LOWCOUNTRY 33-44 (1998).

common.[64] Plantation owners sought enslaved Africans resistant to malaria and turned to the "Gold Coast" – modern day Ghana and Benin.[65] Similarly, because Portuguese and French slave traders were the primary slave traffickers in New Orleans, many of the enslaved Africans brought to Louisiana were from Angola.[66] Thus, the forced migration of enslaved Africans to the United States led to geographic variation in this country similar to that of regional African variation.

Once in the United States, the clusters of African Americans either remained in their geographical origins or migrated in distinct groups, as family members joined family members, friends followed friends, and neighbors encouraged neighbors to emigrate.[67] This patterned migration resulted in further geographic variation throughout the United States. For the most part, this took place during the "Great Migration" – roughly 1910 to 1930 – when African Americans in the rural south traveled north for better jobs in light of World War I and a boll weevil crop infestation in the South.[68] These migrations took predictable routes: African Americans from Mississippi, Alabama, and Louisiana largely followed the Mississippi River and migrated to the great cities of the Midwest, such as Detroit, Chicago, Cleveland, and Kansas City; and African Americans from the Carolinas and Virginia tended to travel up the coastline to Washington, D.C., Philadelphia, and New York.[69] Notwithstanding the effects of this large-scale migration, most African Americans have remained in the southern part of the United States, in the crescent-shaped region ranging from Washington, D.C. to Louisiana.[70] Today, scientists observe genetic variation among African Americans in different regions of the country based on their migration routes and the variable levels of mixing with European Americans in different parts of the United States.

Heterogeneity also exists in African-American mtDNA profiles as one moves westward across the country. African Americans living in the western United States tend to exhibit larger percentages of European and Native-American ancestry than those living in the South, Mid-Atlantic, and Midwest.[71] These phenomena may be the result of history, in that western

---

[64] *Id.* at 34-36.

[65] Fatimah L. Jackson, *Concerns and Priorities in Genetic Studies: Insights from Recent African-American Biohistory*, 27 SETON HALL L. REV. 951, 961-62 (1997); Parra & Marcini (1998), *supra* note 58, at 1839 (listing countries of Africa by economic region). This very same resistance makes African Americans whose ancestors come from the Gold Coast more susceptible to sickle cell trait and sickle cell disease. A. Muniz et al., *Sickle-Cell-Anemia and Beta-Gene Cluster Haplotypes in Cuba*, 49 AM. J. OF HEMATOLOGY 163 (1995); G. Pante-De Sousa et al., *Betaglobin Hapolytpes Analysis in Afro-Brazilians from the Amazon Region: Evidence for a Significant Gene Flow from Atlantic West Africa*, 26 ANNALS OF HUM. BIOLOGY 365 (1999).

[66] CURTIN (1969), *supra* note 59, at 83.

[67] *See generally* JAMES R. GROSSMAN, LAND OF HOPE: CHICAGO, BLACK SOUTHERNERS, AND THE GREAT MIGRATION (1991).

[68] *Id.* at 28-30.

[69] *Id.* at 112-13 (migration from Mississippi delta to Chicago); NICHOLAS LEMANN, THE PROMISED LAND: THE GREAT BLACK MIGRATION AND HOW IT CHANGED AMERICA 109-20 (1991) (migration from Carolinas and Virginia up East Coast).

[70] *See* http://www.census.gov/geo/www/mapGallery/images/black.jpg. (pictorial depiction of geographical distribution of African Americans in United States).

[71] See Parra & Marcini (1998), *supra* note 58, at 1845-47; Ranajit Chakraborty, *Gene Admixture in Human Populations: Models and Predictions*, 29 YEARBOOK OF PHYS. ANTHROP. 1-43 (1986); David C. McLean, Jr. et al., *Three Novel MtDNA Restriction Site Polymorphisms Allow Exploration of Population Affinities of African Americans*, 75 HUM. BIOLOGY 147 (2003).

territories and states had less restrictive social mores with respect to interracial relationships at the time of greatest migration. Additionally, the number of Native Americans surviving European settlement living in western states was significantly higher than in the east, which helps to explain the Native-American "admixture" in African-American mtDNA profiles.[72]

A further critical dimension of regional variation in mtDNA profiles is a result of the variation in the level of admixture between African Americans and other groups around the country. For example, while African Americans living in Charleston, South Carolina, possess about 6.5% of European maternal ancestry, this figure is much higher in Baltimore (14.94%), New York (9.11%), and Pittsburgh (9.9%).[73] To determine the frequency of a mtDNA sequence at a Charleston crime scene, for example, a forensic scientist might best use a database that takes into account the types of mtDNA profiles that exist in Charleston. As further illustration, Jamaican Americans, whose mtDNA is on average 12.93% derived from European ancestry, have quite different mtDNA profiles from African Americans in most American cities – information that should be known to the forensic scientist in electing to which mtDNA database to compare the questioned profile.[74] The forensic database does not account for, or reflect, these regional differences.

b. Hispanic Population

Regional variation is of equal concern with Hispanics. While the database has a Hispanic category, most geneticists agree that the term "Hispanic" is primarily a language-based categorization, not a genetic one.[75] Not surprisingly, then, individuals in the linguistic category "Hispanic" display tremendous amounts of genetic variation.[76] One cannot reasonably claim, for example, that Hispanics living in South Florida (largely Cuban and Puerto Rican in ancestry) are genetically representative of Hispanics living in California (largely Mexican in ancestry). Yet the design of the SWGDAM mtDNA database assumes that mtDNA profiles of Hispanic-Americans are randomly distributed. This failure to account for genetic diversity is particularly troubling given that the classification system *does* distinguish between Southeast and Southwest Hispanics in the *nuclear* DNA database, presumably to account for population substructure

---

[72] *See* http://www.census.gov/geo/www/mapGallery/images/americanindian.jpg (visual depiction of heavy Native American clustering in western part of United States); STELLA U. OGUNWOLE, THE AMERICAN INDIAN AND ALASKA NATIVE POPULATION: 2000, 4-6 (U.S. Census Bureau Feb. 2002) (noting that forty-three percent of American Indians lived in West, thirty-one percent lived in South, seventeen percent lived in Midwest, and nine percent lived in Northeast).

[73] Parra & Marcini (1998), *supra* note 58, at 1845. The admixture study reports *two* results from Philadelphia, based on two independent sample sets taken from patients in two separate hypertension studies. These sample sets exhibited significant differences in their percentage of admixture. *Id.* Thus, even within a single city, different groups of African Americans display significantly different mtDNA profiles.

[74] *Id.* at 1845-47.

[75] *See* C. Bonilla, M.D. Shriver et al., *Admixture in the Hispanics of the San Luis Valley, Colorado and Its Implications for Complex Trait Gene Mapping*, 68 ANNALS HUM. GENETICS 139, 140 (2004) (term "Hispanic" applies to individuals from several continents with "diverse cultural features and genetic backgrounds").

[76] *See id.* (reporting differences in admixture among Puerto Rican, Cuban, and Mexican groups, as well as within smaller region of San Luis Valley); J. Irwin et al., *Development and Expansion of High-Quality Control Region Databases to Improve Forensic MtDNA Evidence Interpretation.*, 1(2) FORENSIC SCI. INT'L: GENETICS 154 (2007) (showing significant regional differences between "Hispanic" populations).

within the "Hispanic" population.[77] The FBI's attempt to subcategorize the (more recombinant) nDNA database to account for substructure is laudable, but the lack of recombination in mtDNA inheritance makes geographic clustering all the more critical in designing a representative mtDNA population database.

c. *Native-American Population*

The SWGDAM database is strikingly inadequate with respect to Native American profiles. The database contains only two categories of Native-American mtDNA profiles, Apache and Navajo, which contain 180 and 146 mtDNA sequences, respectively. These SWGDAM sub-databases are incomplete and unrepresentative. Haplogroup D exists in Apache anthropological databases but is completely missing from the SWGDAM Apache database.[78] The frequency of Haplogroup X in studies in the academic literature is four-to-five times greater than in the corresponding SWGDAM database.[79] More generally, the collections of mtDNA sequences in the research literature – which are approximately one-third the size of the SWGDAM database – report entire haplogroups not present in the SWGDAM database, and significantly different percentages of the kinds of haplotypes represented in the SWGDAM database.[80]

d. *Asian Population*

The same issue of disproportionate representation of sub-populations is also reflected in SWGDAM's East Asian databases. The database fails to account for the proportional ancestry of East Asian Americans. The 753 individuals in the SWGDAM East Asian database are from China, Korea, Japan and Thailand, with almost half from China.[81] Significant disconnect exists between the SWGDAM database and the 2000 U.S. Census figures.[82] For example, based on

---

[77] *See, e.g.*, Bruce Budowle et al., *Population Data on the STR Loci D2S1338 and D19S433*, Forensic Sci. Comm. (July 2001), available at <http://www.fbi.gov/hq/lab/fsc/backissu/july2001/budowle2.htm>.

[78] R.S. Malhi et al., *Native American MtDNA Prehistory in the American Southwest*, 120 Am. J. Phys. Anthrop. 108, 113 (2003) [hereinafter Malhi et al. (2003)].

[79] *Id.* Additionally, the Navajo and Apache are not representative of the variation present in haplotypes/haplogroups among all North American Native Americans. Tribal groups in the United States share few haplotypes. *See* Malhi et al. (2002), *supra* note 60, at 914, Table 2 (estimating sharing at about 29%).

[80] Malhi et al. (2003), *supra* note 78, at 121-22.

[81] The primary published analysis of this database concerns only the Chinese samples and while the analysis suggests that the frequencies of the haplogroups in the dataset are similar to those in another Han Chinese dataset of 263 individuals, the authors' data reveals significant differences in almost all cases. Allard & Wilson et al., *Control Region Sequences for East Asian Individuals in the Scientific Working Groups on DNA Analysis Methods Forensic MtDNA Data Set*, 6 Legal Med. L11, L18 Fig. 2 (2004). Other studies also show significant genetic variation among and within Asian populations. *See, e.g.*, Kivisild & Tolk et al., *The Emerging Limbs and Twigs of the East Asian mtDNA* Tree, 19 MOLECULAR BIOLOGICAL EVOLUTION 1737 (2002) (other Asian populations not represented in the SWGDAM East Asian database have significantly different frequencies of mtDNA haplogroups than those in the database); Melton & Stoneking, *Extent of Heterogeneity in Mitochondrial DNA of Ethnic Asian Populations*, 41 J. FORENSIC SCI. 591-602 (1996) (same); Yao et al. (2002), *supra* note 53, at nn. 76-78 & accompanying text (combining all Han Chinese would be inappropriate).

[82] *SEE* TERRANCE J. REEVES & CLAUDETTE E. BENNETT. WE THE PEOPLE: ASIANS IN THE UNITED STATES, Pub. No. CENSR-17, U.S. Census Bureau, Dep't of Commerce 1 & Table 1 (2004), available at http://www.census.gov/prod/2004pubs/censr-17.pdf (listing major Asian groups in U.S., many of which are not included in SWGDAM Asian databases).

the 2000 Census, 18.3% of the U.S. Asian population is Filipino, while there are no Filipinos in the SWGDAM database.[83] Similarly, the percentage of Chinese and Korean individuals in the SWGDAM database is double that of the Census, while the database's percentage of Asian Indians is one seventh of their true percentage in the population in 2000. Simply put, the SWGDAM database is not representative of the distribution of Asian American source populations.[84]

Perhaps most significantly, a study of the Han Chinese revealed dramatic regional differences in haplogroup frequencies among a population that constitutes ninety-three percent of the Chinese population and nearly twenty percent of the world's population.[85] Researchers examined 263 unrelated Han Chinese samples taken from six different provinces. They observed that, while certain haplogroups made up almost twenty percent of the population in a certain province, the haplogroup was nonexistent in a different province. Ultimately, the clustering in particular provinces was so pronounced that the authors concluded that an East Asian database, or even "Northern Han" and "Southern Han" databases, would grossly underestimate the frequency of certain groups of sequences that themselves are highly common in surrounding regions.[86]

e. Caucasian Population

[83] *Id.* at 4 & Fig.1.

[84] A chi-square analysis comparing the SWGDAM and 2000 Census frequencies of Asian subpopulations (converted to sample sizes in both cases) rejects the hypothesis that the SWGDAM database is a random sample of the Census Asian Populations, with an extremely significant p value of less than 10^{-20}

Asian Population	2000 Census	SWGDAM Database
Asian Indian	16.1%	2.4%
Cambodian	1.7%	0
Chinese	23.8%	45.7%
Filipino	18.3%	0
Hmong	1.7%	0
Japanese	7.8%	20.9%
Korean	10.6%	23.3%
Laotian	1.7%	0
Pakistani	1.6%	1%
Thai	1.1%	6.7%
Vietnamese	10.9%	0
Other Asian	4.7%	0

[85] Yao et al. (2002), *supra* note 53, at 635.

[86] *See id.* at 649 ("The comparison of the regional Han mtDNA samples revealed an obvious geographic differentiation in the Han Chinese, as shown by the haplogroups-frequency profiles. . . . Hence, the grouping of different Han populations into just "Southern Han" and "Northern Han" or the use of one or two Han regional populations to stand for all Han Chinese . . . does not appropriately reflect the genetic structure of the Han.") (citations omitted).

While not as obviously problematic, the Caucasian database is also flawed in its apparent failure to account for non-random distribution of ancestral haplogroups. Those who argue that SWGDAM's Caucasian database is representative point to the fact that its samples include 44.2% of the H Haplogroup, which appears in approximately the same proportion in certain Western European countries.[87] But the percentage of "H" varies widely outside a handful of countries in Western Europe, such as countries in Scandinavia, Eastern Europe, and parts of West Europe such as France, Northern Germany, and Scotland – areas where, of course, many American families originated.[88]

In sum, the SWGDAM database appears to misrepresent the regional genetic diversity of mtDNA profiles to ignore studies showing tremendous intra-group diversity in the various macro-ethnic categories represented in the database. The database combines internally heterogeneous groups under broad rubrics without a demonstration that the deviation from homogeneity would have negligible consequences in reporting mtDNA match significance. If the database is not representative, it does not serve its intended purpose of providing a random selection of the relevant population from which reliable sequence frequency estimates may be calculated. In turn, if the database does not produce reliable frequency estimates, its admissibility under the prevailing rules for admission of scientific evidence would seem to be questionable at best.[89]

III. The Role of Ethnicity in Trying the MtDNA Case

The ethnic background of the accused may play a critical role in informing the appropriate defense strategy in a mtDNA case, given the significant population substructuring described above and its influence on the distribution and frequency of a given mtDNA 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 become knowledgeable about 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. If, for example, a defendant's ancestral origins are not represented in the mtDNA database but he is a member of a significant community of

---

[87] *See* Allard et al. (2002), *supra* note 23, at 5-6.

[88] *See* Wojciech Branicki, Ksenia Kalista et al., *Distribution of MtDNA Haplogroups in a Population Sample from Poland*, 50 J. FORENSIC SCI. 732, 733 (2005) (H Haplogroup observed in 37.8% of samples in population from Southern Poland); Vincent Dubut & L. Chollet, *MtDNA Polymorphisms in Five French Groups: Importance of Regional Sampling*, 12 EUR. J. HUM. GENETICS 293-300 (2004) (within France alone, frequency of H varies between 35% and 50% in two separate communities in Brittany); Ana M. Gonzalez & Antonio Brehm, *Mitochondrial DNA Affinities at the Atlantic Fringe of Europe*, 120 J. PHYSICAL ANTHROPOLOGY 391-404 (26.3% in Norway; 34% in England; 36.4% in Northern Germany; 38.5% in France; 42.2% in Galicia); Malyarchuk & Grzybowski, *Mitochondrial DNA Variability in Bosnians and Slovenians*, 67 ANNALS HUM. GENETICS 412-25 (2003) (frequency of H haplogroup is 24% in Finland, 26.8% in Scotland, and 45% in Poland). *See also* Luisa Pereira et al., *Evaluating the Forensic Informativeness of MtDNA Haplogroup H Sub-Typing on a Eurasian Scale*, 159(1) FORENSIC SCI. INT'L 43, 50 (2006) (use of SNPs to more closely examine haplogroups demonstrates significant inter-relatedness below the haplogroup level and suggests that "phylogenetic dissection of mtDNA haplogroups is revealing gradients previously hidden on the Eurasian scale").

[89] That the SWGDAM database is not representative also arguably invalidates the FBI's use of a confidence interval to extrapolate from the database to a sub-population. Use of such margins of error presupposes random distribution in the population. WITTE (1985), *supra* note 54, at 214.

immigrants of common origin in his place of residence, a strong argument can be made that the lack of a "match" to his mtDNA 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.

IV. 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 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 mtDNA 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. Indeed, the use of experts is limited only by the imagination of defense counsel.

V. MtDNA Treatment in the Courts

A number of challenges have been made to the admissibility of mtDNA evidence as an inculpatory tool, but generally without success thus far. Many courts have acknowledged that "mtDNA analysis is more applicable for exclusionary, rather than identification, purposes,"[90] but have nonetheless opted to admit the evidence on grounds that the discriminatory power and other limitations of mtDNA evidence are questions of weight, rather than admissibility.[91] In all, appellate courts in at least eleven states,[92] as well as one federal district court, [93] have found the

---

[90] *Vaughn v. State*, 646 S.E.2d 212, 214 (Ga. 2007).

[91] *See, e.g.*, *id.* at 215 ("The conflicting expert opinions on the [mtDNA] test results go to the weight rather than the admissibility of the testimony"); *People v. Ko*, 757 N.Y.S.2d 561, 563 (App. Div. 2003) ("mitochondrial DNA analysis has been found reliable by the relevant scientific community; issues regarding contamination go to the weight to be given such evidence); *State v. Pappas*, 776 A.2d 1091 (Conn. 2001) (holding that "issues regarding contamination are important and may bear on the weight of mtDNA evidence in a particular case, but that those issues do not undermine the admissibility of the results of the mtDNA sequencing process") (internal citation omitted).

[92] *Wagner v. State*, 864 A.2d 1037 (Md. App. 2005) (finding mtDNA "inclusion" evidence properly admitted); *State v. Council*, 515 S.E.2d 508, 518 (1999) (finding underlying science of mtDNA reliable and "inclusion" evidence properly admitted); *State v. Underwood*, 518 S.E.2d 231, 240 (N.C. 1999) (holding mtDNA testing sufficiently reliable to warrant its admissibility into evidence); *State v. Scott*, 33 S.W.3d 746, 756 (Tenn. 2000) (holding mtDNA properly admitted without admissibility hearing); *Adams v. State*, 794 So.2d 1049, 1064 (Miss. App.2001) (holding science of mtDNA sequencing adequately proven); *State v. Pappas*, 776 A.2d 1091, 1110 (Conn. 2001) (finding no error in admitting mtDNA evidence); *People v. Holtzer*, 660 N.W.2d 405, 411 (Mich. 2003) (holding use of mtDNA for identification of defendant admissible under test for novel scientific evidence); *Magaletti v. State*, 847 So.2d 523, 528 (Fla. Dist. Ct. App. 2003) (holding use of mtDNA analysis to prove identity satisfied *Frye*); *People v. Ko*, 757 N.Y.S.2d 561, 563 (2003) (upholding trial court's admission of mtDNA evidence). Admission of mtDNA evidence also has been upheld in a number of unpublished appellate decisions. *See State v. Smith*, 100 Wash.App. 1064, 2000 WL 688180 (Wash.Ct.App.2000); *State v. Ware*, 1999 WL 233592 (Tenn. Crim. App.1999); *Sheckells v. Texas*, 2001 WL 1178828 (Tex. Ct. App. 2001).

[93] *United States v. Coleman*, 202 F.Supp.2d 962, 970-71 (E.D. Mo.2002) (denying defendant's motion to exclude mtDNA and finding it reliable, helpful to the jury, and not unduly prejudicial);

results of mtDNA testing admissible under various evidentiary standards. Fresh admissibility challenges are critical, however, in light of the wealth of new and ongoing research that calls the reliability of the mtDNA forensic databases, and the inculpatory claims hinging upon them, seriously into question.