

Peer Review and the Validity of the FBI's Validation Studies of P.C.R. Amplification and Automated
Sequencing of Mitochondrial DNA for Forensic Use.

by

William M. Shields, Ph. D.

Department of Environmental and Forest Biology

State University of New York

College of Environmental Science and Forestry

Syracuse, New York 13210 USA

Running Head: Mitochondrial DNA Validation

Abstract: Mitochondrial DNA sequencing was introduced for the first time in North America in a criminal case by the FBI laboratory as evidence of identity in a capital murder trial in August of 1996 (State of Tennessee v Paul Ware). The use of this novel, and potentially very useful, technique is predicated on a series of papers reporting on the forensically useful characteristics of mtDNA and a set of validation studies that claimed that methods and a protocol had been developed that could sequence forensic samples reliably and permit valid inferences to be drawn from the occurrence of identical sequences in known and evidence samples. Here, I review the publications and compare their recommendations, the actual data presented, and the conclusions drawn in those papers with the testimony given in court by the FBI scientist/authors of those papers. I conclude that there are significant discrepancies between the two sources of information about the foundation for the reliable use of these techniques. The validation studies are weak, rather than rigorous, challenges to the technology, with many critical tests being based on samples too small to permit valid inferences. For example, the recent documentation of significant mutation rates and resultant levels of heteroplasmy indicate that contra earlier assertions, maternal relatives will not always show identical mtDNA sequences and that different tissues from a single individual (e.g., hair versus blood) can produce different mtDNA sequences as well. This problem was not anticipated during the initial use of this technique, so earlier testimony was incorrect. Similarly, while all agree that contamination is a crucial issue for mtDNA analysis, the studies performed to date do not permit firm conclusions regarding the frequency and level of contamination necessary to invalidate sequence comparisons. In sum, we do not yet know how often we might expect errors in mtDNA comparisons, including both false positives and false negatives, owing to these two phenomena. We are at a similar stage of

uncertainty about the meaning of a match. If the FBI observes an mtDNA match, they currently report how often the same sequence is seen in their database(s) of sequences. In my opinion, this provides a potentially biased estimate of the likelihood of matching because the databases are small. I have analyzed those databases, using pair-wise comparisons, and show that globally, the expected frequency of control region matching ranges from 1 in 114 to 1 in 468 in the different ethnic databases. These estimates support the published conclusion that mtDNA typing is not nearly as discriminating as standard nuclear typing at multiple loci, but contradicts much of the testimony on this issue. The contrasts and contradictions among the published recommendations, the data gathered, the conclusions drawn, and the testimony lead to a conclusion that despite editorial peer review of the validation papers, the validation process has just begun for this promising new technique. More data must be gathered and additional studies completed to fine tune the forensic protocols and to insure that the technique can be used reliably in forensic circumstances.

KEYWORDS: mitochondrial, DNA, validation, serology, mtDNA, heteroplasmy, contamination, sequencing, database, match frequency.

INTRODUCTION

Whether the legal standard is Frye or its recent federal successor Daubert, novel scientific techniques or novel uses of established scientific techniques only produce evidence judged legally admissible in court proceedings after they have been studied enough to be judged scientifically valid and reliable. The primary tool permitting technology transfer from a basic science or medical context to the forensic arena is the “validation” process. Here validation means nothing more than the process of performing experiments of different kinds that challenge the technology’s ability to do what it purports to do. In the case of DNA typing this means demonstrating through rigorous tests that the typing of evidentiary and known samples can validly produce reliable evidence of identity. In simplest form, this requires demonstration that an individual’s DNA is reasonably constant over time and tissues, and is typed repeatedly as identical, that such DNA differs among individuals and that differences can be distinguished accurately and repeatedly using the pristine (e.g., known blood) and potentially contaminated and degraded samples (e.g., evidence samples left at a crime scene) associated with forensic cases. Finally, to understand the potential meaning of a reported match between evidence and known samples, some reliable estimate of the probability of a random match and the likelihood of technical or human error must be estimated and presented to a judge or jury making decisions about guilt or innocence (1).

It is my belief that every scientist recognizes the importance of validation studies in technology transfer. This is as true of the forensic scientist as it is of the medical geneticist or evolutionary biologist. That’s why we all have to learn some statistics or at least consult with competent statisticians. One would hope that before any technology is offered in the courtroom, rigorous validation that documents the capacity and limitations of the system will have been completed and incorporated into the use and

presentation of such evidence. Consistent with this view is the statement by Budowle et al. (5, p. 83) about mitochondrial DNA (mtDNA) that:

“For a new genetic typing technique to gain acceptance and use in forensic serology, it must be genetically informative, permit analysis of a large number of samples, be reproducible, and be properly validated in a forensic context.”

That the proponents of forensic mtDNA typing at the FBI agree with this principle is indicated by Mark Wilson’s statement in an affidavit dated July 16, 1997 and filed in the case of Tennessee v Ware, where he states (p. 4):

“Mitochondrial DNA typing underwent years of research and development prior to being introduced into forensic casework. It was initially envisioned as a potential forensic marker in 1988. After extensive research projects conducted by the FBI Laboratory and other laboratories over approximately a four year time span, it was implemented into criminal casework in June 1996. Extensive validation studies were completed prior to the implementation of the technique. These studies were published in peer-reviewed scientific journals.”

I am convinced that this affidavit overstates the breadth and rigor of the validation studies actually completed. I intend to illustrate this conflict between the evidence and testimony here as a caution and case study about the proper or improper interpretation and presentation of novel technology in the forensic arena.

I have acted as a consultant and expert witness on DNA typing issues in numerous courtrooms since 1990. Although I have worked with prosecutors and have testified for the State, I usually work with the defense bar. My review of forensic DNA technology has convinced me that the FBI

laboratory, whatever it's failings(2), if any, is the premier organization generating the ideas, validation research, and eventually the practical transfer of technology from numerous disciplines for use in forensic settings. This has been especially true of DNA typing. Witness the fact that their methods, databases, computer programs, and slant on things genetic are the foundation for most, if not all, of the public (and many private) labs doing forensic DNA analyses in the United States. The research and development of such technologies takes place at their research and training center in Quantico, Virginia and only then is it transferred to the working labs in Washington DC.

Here, I will review the research and validation studies they have performed for forensic mtDNA typing as a case study which I hope will illustrate what I believe are the potential and actual failings of forensic science in validating this new technology. I will also discuss some of the potential reasons why this failure might have occurred. I plan to do this by reviewing quotes and data from published papers and court testimony by the FBI and other forensic experts engaged in using the technology. I believe such a review illustrates what the forensic scientists say should have been done, shows what actually has been done, and finally illustrates what they report has been done, and how they present such evidence in courtrooms. I believe that a lack of concordance about crucial issues at these levels indicates a potentially serious problem with the validation of new forensic technology.

Mitochondrial DNA in Forensic Applications

Mitochondrial DNA has become an issue in forensic science because of its unique and forensically useful properties. Because it is found outside the nucleus and occurs in higher copy number than nuclear DNA it can be extracted from tissue with little or no typeable nuclear DNA. This includes single hair shafts (1), bones, teeth, and other tissues that have been exposed to the elements, in graves for example, even for considerable periods of time (e.g., 3, 4 and reviewed in 5). It is maternally

inherited, and owing to its asexual transmission, until recently (e.g., 6, 7) was thought to be homoplasmic (that is the same sequence would be found in any individual over time and tissue type) in almost all normal individuals (e.g., 5, 8-9). If all of this were true, this would allow one to use living maternal kin to provide evidence of identity for remains that have been lost to view for considerable time or remains from victims of mass disaster. For example, the Armed Forces DNA Identification Laboratory (AFDIL) has used it to help identify the remains of soldiers from the Vietnam war, as well as victims of plane crashes, and other mass disasters (e.g., 10). One recent newsworthy example of its use in the investigation of historical remains was the “identification” of Tsar Nicholas II and his family, and servants by comparison of mtDNA sequences to both living and long dead maternal relatives of the Tsar and Tsarina (11, 12).

Amplification and automated sequencing of the hypervariable regions (HV1 and HV2) of the non-coding displacement region (D-loop) or control region (CR) in human mitochondrial DNA has been proposed as a technique that could provide evidence about the identity of crime victims, especially skeletal remains, as well as perpetrators of crime who leave biological material at a crime scene not suited for standard DNA typing, especially hair shafts without roots (5, 8, 13-14). Both sorts of information often would be useful in criminal investigations and subsequent court proceedings and so it would be a forensically useful technique.

On the basis of this early promise and preliminary research, then, the FBI decided to generate a forensic protocol and complete the necessary validation studies to bring mtDNA typing to court (15-16). Upon completion of these “validation” studies, and publication of their protocol in May, 1996, they brought such evidence to court for the first time in August of 1996 in Chattanooga, Tennessee in the capital murder trial of Paul Ware. Since then, additional evidence has been analyzed by the FBI and

has been offered in more than a dozen cases (as of late 1997), including other capital crimes where the ultimate consequence can be considered, without hyperbole, a life or death matter.

Given the stakes, and the fact that the original validation studies were published in peer reviewed journals, *Biotechniques* (15) and the *International Journal of Legal Medicine* (16), one would presume that the validation and the resulting protocol were themselves valid with respect to the logic and statistical design of crucial experiments acting as the foundation for their forensic use. Such an assumption is consistent with the statement by FBI research scientists that:

“However, it can be anticipated that the techniques developed to detect mtDNA polymorphisms will be subjected to a thorough legal scrutiny upon their introduction to the courts, and rightly so. It is the responsibility of those engaged in research directed at the application of the mtDNA typing methods to forensic samples to ensure that these approaches have been comprehensively validated and can be applied legitimately to the types of specimens encountered in the forensic laboratory.” (5, pp.90-91)

One would also presume that when the forensic scientists who designed and completed such validation studies entered the courtroom, their testimony would reflect all of the information gathered during their studies. In my opinion, one would be wrong on both counts.

THE GUIDELINES

In 1993, the FBI presented an overview and a set of “Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science” in their Crime Laboratory Digest (8, based in part on ref. 5) in which they noted several major assumptions and made several recommendations for validating mtDNA sequencing. As they (p. 68) noted:

“An important feature of mtDNA, which simplifies DNA sequencing, is its monoclonal nature. ... for practical purposes all copies of an individual’s mtDNA sequence are identical.”

Similarly, their colleagues in the Forensic Science Service in the UK (14, p. 86) noted:

“A major concern in the use of mtDNA sequences for forensic purposes is potential heterogeneity within an individual. However a high degree of sequence homogeneity has been demonstrated in somatic cells [20,21] and although heteroplasmy (i.e., a mixture of mutant and wild type mtDNAs in the same cell) has been reported in humans this only appears to be associated with certain neuromuscular diseases and has not been observed in the non-coding region (12, 22).”

This “validated” assumption of homoplasmy from Higuchi et al.’s (3) early reports was still being asserted in 1995 in the FBI validation papers (15-16) and by other non-forensic experts like Poulton (9, p 224) who noted:

“Heteroplasmy (the presence of both mutant and wild-type mtDNA in the same individual) is common in mtDNA disease, while homoplasmy (a population that is ~ 99.9% identical [Monnat and Loeb, 1985] is the norm in controls.”

The section of the FBI Guidelines paper (8, p 72-73) entitled, “Validation”, notes:

“A number of specific considerations should be addressed before mtDNA sequencing is used in forensic casework. They are as follows:”

1. Database sharing among laboratories requires that the sharing labs produce identical results with same mtDNA regardless of the many different methods being used (e.g., automated versus manual sequencing, or automated sequencing with dye terminators or dye primers).
2. “Maternal inheritance of mtDNA must be experimentally verified if it is to be used as a means of establishing identity in forensics. As maternal inheritance has been well documented, this verification should be minimal, encompassing a small number of families and showing that sequence information is identical from maternally related individuals.”
3. “A number of tissue types from the same individual should be used for DNA extraction and sequencing to show that an individual’s mtDNA sequence is identical from tissue to tissue, regardless of the somatic origin of the extracted mtDNA. Studies with human hairs should show that different hairs from the same individual yield identical mtDNA sequences.”
4. Environmental insult studies, such as exposure of samples to sunlight, soil, etc. as well as hair dye and bleaches, should show that insults allow reliable sequencing and known positive controls should be used with every amplification and sequencing run.

Sources of Error

One source of sequencing error would be the technical limits of the automated sequencer and their associated chemical kits. Forensic scientists have acknowledged this and suggest that the best

method to deal with the inherent error associated with automated sequencing (about 1% of bases sequenced incorrectly or ambiguously) is to sequence the same sample again, either with an independent extraction, by sequencing both complementary sequences, or both. Sequencing twice would reduce the intrinsic error rate to about .01% or 1 in 10,000 (13).

In a section of the Guidelines entitled, "Contamination", Wilson et al. (8, p 73) note:

"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. This fact would not be detected without the appropriate negative controls. Reagent blanks and negative controls should be employed with each extraction and amplification."

Finally, many forensic scientists have noted that once a match between the mtDNA sequence from an evidence sample and the sequence of a known individual is declared, one must provide some notion of matching probabilities. For example, Holland et al., of AFDIL (10, p 551) state:

"It is important to note that mtDNA sequence analysis, or SSO typing of the mtDNA control region is only useful as an identification tool if a relevant sequence database is generated, and the integrity of maternal inheritance is validated."

The FBI in their seminal paper (8, p 73) note that:

"Once a mtDNA type has been established, a determination of the frequency of this type in the general population could be presented in order for a jury to assess the weight of the evidence in a courtroom trial."

“As the mitochondrial genome is inherited in its entirety without recombination, a complete mtDNA sequence is the equivalent of a single genetic locus.”

“Therefore, at this time the frequency of a mtDNA type within the population is calculated by the counting method, or the number of times a particular sequence has been observed in a population, divided by the number of sequences contained within the appropriate database.”

And Weedn (17, p 191) then of AFDIL notes:

“The discriminatory power of mtDNA sequencing is not as high as that of nuclear DNA testing, as it is only a single genetic system.”

The final conclusion of the FBI guidelines asserts (8, p. 76):

“In routine forensic cases, where nuclear DNA is extracted from a sample, polymorphic nuclear markers such as DQ α , short tandem repeat loci, etc., can and should be utilized. These markers should take precedence over mtDNA analysis. Not only are such markers more statistically informative than mtDNA, they also are not as susceptible to contamination as the analysis of mtDNA. However, mtDNA can and should be the method routinely used with certain classes of biological evidence which by nature have very little DNA, such as telogen hairs, hair shafts, bones, and teeth.”

THE VALIDATION STUDIES

By consensus, then, the major areas of concern for the forensic use of mitochondrial DNA sequencing appear to be heteroplasmy, contamination, and suitable databases to weigh evidence of matching sequences and to a lesser extent evidence that environmental insults do not affect sequencing. As Wilson et al. (8, p 76) noted in the Guidelines: "Validation studies must be conducted on any new forensic technique". Based on their two reported validation studies (15-16), they feel they have addressed all of the critical issues and were ready for casework and the courtroom.

Validation of homoplasmy:

They note:

"We have observed identical mtDNA sequence from DNA extracted from hair shaft and blood in approximately 30 individuals." (15, p 667).

"Body fluid samples from six different donors were used as described previously" . . . (16, p 71)

"In all cases, the sequence obtained from hair shaft matched that of the donor as determined by separate sequencing of DNA extracted from bloodstains." (16, p 71).

"Head, chest, pubic, limb, and axial hairs were collected from one donor in order to test whether the specific body area had any effect on the amplified mitochondrial DNA sequence. Additionally, hairs were collected from two individuals from the front, crown, right, left, and back of the scalp. . . . All typings from the same individual matched each other and also matched the known mtDNA sequence obtained from DNA extracted from bloodstains." (16, p 72).

“Hairs were collected from two individuals who were deceased for a period of approximately 2 years. Blood and tissue samples from these individuals had been collected post mortem. Sequences from head hairs matched the known sequence derived separately from the blood and tissue.” (16, p 72).

Validation of Contamination Issues

“Because of the sensitivity of this method. The issue of detection and quantitation of external contamination is important. We have observed low levels of amplified product in negative controls and reagent blanks when amplifying human mtDNA.” (15, p. 667).

“However, in cases where low level contamination is found, the experiments discussed herein suggest that with careful quantitation, correct typing results can be achieved when the ratio of sample to contaminant is greater than or equal to 10:1.” (15, p. 667).

“Previously sequenced amplicons were quantified by CE and two samples with different mtDNA control region sequences were chosen for this study. Amplified DNA was mixed together from the two samples in ratios of 4:1, 8:1, 12:1 and 16:1. Sequencing reactions were performed on the mixtures as previously described. Editing of the analyzed data from the sequence of the mixed samples was conducted by two individuals. Ambiguous bases (bases which could not be designated) and errors (bases which were called incorrectly) were noted. In the 4:1 mixture, both ambiguities and errors were observed. The 8:1 mixture yielded only two ambiguous calls from both editors at positions that differed in sequence between the mixed templates. No errors or

ambiguities were noted at the 8:1 mixture. All base calls were completely consistent with the more abundant sequence at and above 8:1.” (15, p 667).

“We have shown empirically that under the conditions described in this paper, base-calling errors do not occur at mixtures greater than 8:1, and hence we have adopted a 10:1 working ratio. In order to confirm that correct typing results can be achieved using this ratio, additional tests were conducted. Amplified DNA samples from a variety of donors were mixed in 10:1 ratios with other amplified DNAs. The dominant and minor samples were blindly reversed to the sequence editors. At the 10:1 ratio in a total of five such tests, all of the base calls from both editors were consistent with the known sequence of the more abundant sample with no errors or ambiguous calls.” (15, p 668).

“It should be stressed, however, that with evidentiary material such as hair, contamination by adherence should also be thoroughly addressed with properly designed validation studies. Hairs contaminated with semen, blood, etc., should be cleaned and typed in order to confirm that the cleaning procedure effectively removes adhering contaminants prior to DNA extraction.” (15, p 668).

“In a more relevant and important set of experiments from a forensic standpoint, hairs were deliberately contaminated with body fluids from individuals who were known to have a different mitochondrial DNA sequence from the hair donors. (16, p 72).

“ In one initial experiment, the mitochondrial DNA type from the extracted hair matched that of the blood donor rather than that of the hair donor. Thereafter, more rigorous cleaning methods were investigated.” (16, p. 72).

A total of 48 reactions were performed on 12 contaminated hairs: 4 hairs were contaminated with blood, 3 with saliva, and 5 with semen. We found that 29 typed on the first attempt, for a success rate of 60%. Additional testing on these samples raised the success rate to 88%, all matching the known sequence from the hair donor” (16, p. 72).

“It should be noted that in the present study, the vast majority of instances in which successful typing on the first attempt was not achieved was not due to amplification failure, but to the fact that contamination levels exceeded the 10% limit, and hence analysis was not attempted.” (16, p. 72).

THE TESTIMONY

All of the courtroom testimony on these issues reviewed here is taken from transcripts by two of the FBI scientists (Wilson and DiZinno) responsible for the validation studies (Wilson was the first author on all of the papers, DiZinno a co-author on many) and the day-to-day operations of the mitochondrial DNA lab (now DNA2) at the FBI lab in Washington DC, as well as one independent scientist (Stoneking) who was also a co-author with Wilson and DiZinno of the FBI’s Guidelines paper. The transcripts record testimony at the capital murder trial of Tennessee v Paul Ware in Chattanooga, the first trial in North America using mtDNA (Mark Richard Wilson, August 28, 1996), or in the preliminary hearing (Joseph DiZinno, August 1, 1997) and admissibility hearing (DiZinno and Stoneking, November 21, 1997) associated with the capital murder case of Commonwealth of PA v Patricia Rorrer in Allentown, Pennsylvania.

HOMOPLASMY

Cross of Wilson, August 28, 1996.

Q. So there is no inclusion in this; it merely cannot be excluded?

A. Well, if, a biological specimen comes from a particular person, you would expect that the DNA sequence obtained from the known sample of that individual with the question specimen would have the same sequence, so it is consistent with originating from Mr. Ware.

Direct of DiZinno, August 1, 1997

A (In part) "So all maternal relatives will have the same mitochondrial DNA."

Q. Let me ask you this, Mr. DiZinno. Does it matter that you compared the defendant's blood to hair rather than hair to hair? Does that matter in mitochondrial DNA analysis?

A. Generally no, it does not.

Q. And how do you know that?

A. We've done a number of validation studies that show that. We've taken blood, tooth, saliva, from the same individual and always obtained the same sample, the same type. ... and we have seen cases where a blood sample, when looking at 600 of these bases, may be one base different than a hair sample, but we see that in both blood and generally in the saliva and the hair and we're able to address that issue. But generally we don't see a difference in sequences between the same tissues from the same individual."

Cross of DiZinno, August 1, 1997.

Q Now, you stated that there was – there are differences at various bases when you test blood versus hair. Is that correct?

A At times there can be. There's something called heteroplasma, where you can take a tissue, let's say a blood sample, and then a hair sample, and over 600 bases, it may differ by one base. But what we see when we see that, let's say we're looking at that one base, and let's say the blood is a G and the hair is a C, but what we'll see in the blood is often times, we can actually see a G and C mixture at that base. So if we see that difference at one base, we'll go back and get known – more known samples from that individual, saliva, we'll even go to other maternal relatives and see if we see that difference in other related individuals. So it actually makes the test even more probative to have 600 bases the same and a mixture at that – at a certain specific base.”

Q Well have you conducted any research into heteroplasma?

A Our laboratory has not. We're about to do that.

Q You're about to do that. You haven't done it yet?

A We have not, no, that's correct. Well, that's not correct. We had one individual who was heteroplasmic and we did a number of analyses with that individual's blood, hair, we had a couple of maternal relatives, and we've written a paper and it should be published in Human Genetics this month.

Cross of Stoneking, November 21, 1997.

Q And that principal is called a heteroplasmy, is that correct?

A That's correct.

Q Now that is a readily studied phenomenon in genetics and mitochondrial DNA, is it not?

A That's correct.

Q Prior to that phenomenon, didn't most scientists ascribe to what was known as homoplasmy?

A No, I wouldn't say that is correct. We've always known that humans – that we are a population of mitochondrial DNAs and that – and everyone has a certain amount of underlying variation within that population. The question is what is the amount of that variation and what is the ability of our techniques to detect that variation.

The contradiction between the early testimony and statements in the major FBI papers and the admission that heteroplasmy can occur emerges in testimony later in 1997. Stoneking's assertions are contradicted by all of the forensic papers published prior to 1996 (e.g., the quotes above from 5), including some that he co-authored (8) and also Poulton's (9) more recent and clearly contradictory statement. In 1997, the same FBI scientists (6) reporting a single case of familial heteroplasmy, repeat the statement that, "in humans has been demonstrated to exhibit a high degree of homoplasmy." This implies that our understanding of the degree, pattern, and potential consequences of mutation and heteroplasmy for forensic mtDNA typing is, at best, still in flux.

Recently, published data from the forensic labs has challenged everyone's comprehension of the rate and pattern of mutation and heteroplasmy in the mitochondrial genome, including the hypervariable regions (6-7, and reviewed in 18). I believe that the data contradict co-author Di Zinno's testimony above implying that interpretations of heteroplasmy are easy.

For example, Wilson et al. (6) report that a female's blood was typed for inclusion in the mtDNA database using a particular sequencing kit and typed as a C at position 16355 "and little evidence of heteroplasmy was observed" (Fig. 1, in their paper). Three hairs from the same individual also typed as C at this position at the same time. Using a new and improved sequencing kit, the individual was sequenced again using a hair, yielding an identical sequence except for a T at position 16355. (Of course this implies that different methods can produce different results when sequencing mtDNA, a problem with shared databases). As they noted: "The present case is the first report of human control region heteroplasmy within a known living family that is studied in multiple tissues." To explore the phenomenon, they took samples from numerous tissues from the daughter, her mother, and her maternal brother for mtDNA sequencing. They performed new analyses of two blood stains of the daughter which yielded a mixture, with a predominance of C over T at this position (6, Fig. 2). Four separate buccal swabs from the daughter showed a balanced condition with equal amounts of T and C (6, Fig. 3). Telogenic hair roots showed a mixture- with C predominating in five hairs, T in three, and equal concentrations in two (6, Fig. 4). "Similar results from blood stains, buccal swabs, and hairs were obtained from the mother (data not shown)." Unlike his sister and mother, the son showed a slightly different pattern. For example in one buccal swab and two of three hair roots, T predominated rather than C (6, Figs. 5, 6). Finally they typed one hair from the daughter over its entire length, concluding, "Different portions of same hair exhibited approximately constant amounts of cytosine and thymine (6, Fig. 7). This experiment was repeated with a single pubic hair from mother with similar results (not shown)."

In my opinion, the daughter's single hair showed considerable variation in the relative concentrations of T and C, which could have resulted in different consensus sequences (compare the

tip to the three other segments, see 6 or Fig. 7). What is indisputable is that had the daughter left a hair at a crime scene, she could have been typed as a T or a C or as heteroplasmic for T and C. Depending on which tissue was used for the known sample, and which sequencing kit was used, on the basis of that position, she could have been included, excluded, or the comparison declared inconclusive. If excluded, it would have been a false negative. Similarly, if her hair was found at a crime scene, it could be declared to match an individual with the same sequence with a homoplasmic T, a homoplasmic C or an individual like her that was heteroplasmic for both C and T. In two out three such cases the error or incomplete information would result in declaring a false positive match.

Of course the FBI would argue (see DiZinno above), that this can all be sorted out by taking many samples, including some from maternal relatives (as was done with the Tsar). The problem is that in forensic cases, the evidence is often in short supply and may consist of a single hair which is not amenable to confirmatory re-testing. Since the evidence at a crime scene is by definition from an unknown individual, it will usually be impossible to obtain samples from the unknown relatives of whoever left the sample. The documentation of heteroplasmy resulted in a change in the way sequence mismatches are interpreted by the FBI fully one year after they first brought the technique to the courtroom. They now conclude (6, p 170):

“Depending on the situation, if there is an apparent difference of one or two nucleotides between two samples, one should consider the possibility of heteroplasmy.

Due to the high variability of mtDNA control region sequences, instances where samples differ by a single nucleotide will be rare. In such instances, however, the present study indicates that current sequencing methods will demonstrate the heteroplasmic condition. As in the case of some of the hairs in our study, if the hair

demonstrates mtDNA homoplasmy, and the blood demonstrates heteroplasmy for the particular nucleotide site, then it is not possible to exclude the two samples as potentially originating from the same source.”

Once it became clear that maternal relatives would not always possess identical mtDNA sequences, the forensic community realized that additional analyses might be necessary. The most recent published work on the subject was reported in a study with numerous cooperating forensic scientists from AFDIL, the FBI (Wilson), and the Forensic Science Service in Great Britain, and one academic specialist from Gettysburg College. This superb paper makes a number of important points with more than adequate sample sizes to test and illustrate them, yet these points contradict many of the assumptions and conclusions reported in earlier papers and in testimony by the same groups and scientists. Of course this is an exemplar of how science actually operates qua science, rather than how it is often presented or mis-presented in courtrooms.

Parsons et al. (7, p 363) note in their introduction that:

“Thus, the rate and pattern of mtDNA substitutions remains an unsettled issue of central importance. ... However, a high substitution rate raises the possibility that maternal relatives will sometimes differ at one or more base positions. This has been seen in the case of Tsar Nicholas II and was encountered with *surprising* frequency in the course of forensic casework involving sequence comparisons of family reference samples (see below). The number of forensic cases involving mtDNA is increasing rapidly (for example, in the ongoing program to aid in ‘full accounting’ for > 10,000 U.S. personnel missing from conflicts since World War II and in identity testing of shed

hairs found at crime scenes). Thus, mtDNA substitution will have practical implications that must be addressed explicitly.” (emphasis added)

In this study they used 357 individuals from numerous families in a number of databases to get sufficiently explicit about the issue to satisfy everyone.

“As an internal control for sequence quality and to further establish the reproducibility of DNA sequencing in a forensic context, sequence determination was replicated in independent laboratories for 63 of the 357 individuals sequenced in our study.”

“Additionally, sequences were compared for 69 father:child pairs and as expected, no evidence for paternal transmission of mtDNA was observed.”

“Overall 327 generational events were screened and within these, ten instances of substitution were detected (Table 1)”.

As they note this yields an average estimate of 1 mutation per 33 generational events with the sequences derived from blood showing 8/153 (1/19) and cell lines 2 in 174 (1/87) mutations per generational event, with the rates for the two tissue types significantly different from each other. On the basis of their data, they infer that the mutation rate is much higher than anyone had expected but it is consistent with another estimate using a longitudinal analysis of a single family with mitochondrial disease (19). They conclude: “Most individuals are heteroplasmic at low levels for length variants in this HV-2 C-stretch (AFDIL, unpublished data).”

Commenting on heteroplasmy, in general, they note that one Amish lineage showed extensive levels, a second sibship showed good evidence for it, and a third also showed low levels of heteroplasmy at the 309.1 insertion. They suggest that: “In several other lineages, the

electropherograms were suggestive that heteroplasmy might be present at low levels, but firm conclusions could not be drawn from the direct sequencing data.” They conclude that:

“In our study heteroplasmy was clearly detected in five individuals from three lineages. While this adds to a growing body of evidence that point mutation heteroplasmy in the human CR may not be particularly rare (30, 31, 37, 46, 48, 49), a suitable estimate of the frequency of heteroplasmy is still lacking.”

and

“Our results have implications for the use of CR sequences in forensic identity testing. mtDNA is often employed to compare questioned samples to presumed maternal references. It is now clear that the mtDNA substitution rate is sufficiently high that differences between true maternal relatives will be encountered not infrequently, providing the grounds for a false exclusion. Presently, it is our policy not to report an exclusion based on a single CR sequence difference. In such a case, one can evaluate significance of the evidence through likelihood ratio calculations that incorporate the probability that a mutation has occurred within the lineage (30, 31).”

Stoneking (Testimony in admissibility hearing on November 21, 1997) made the same claim when asked:

Q So in order to exclude a particular person’s sequence, all you have to do is find one bases diversion, isn’t that correct?

A Actually according to the guidelines we prefer to have two or more bases just to allow for the possibility of mutation.

I would add that the substitution rate is high enough and the direction random enough that it would also provide the grounds for false inclusions, especially if sequences differing by one (7, and Stoneking testimony) or two (6) base pairs, as well as the original zero base differences are all counted as inclusions. The overall rate of heteroplasmy in the control region sequenced by forensic labs is now thought to range from a minimum of 10% to over 20% of all humans. The base mutation rate is thought to range from 1 in 19 to 1 in 84 events per generation, both considerably higher than anyone would have guessed. Now geneticists, both forensic and evolutionary, are arguing about whether this results from mutational hot spots, natural selection, the population bottlenecks associated with mtDNA transmission, or some combination of known and unknown phenomena (7, 18, 20-21).

Whatever the causes, the potential consequences are clear. We must get a feel for the common patterns of heteroplasmy and its expected frequency. We must discover how frequently mutations result in somatic mosaicism in different tissues of a single individual and how often different mitotypes segregate within a group of maternal relatives. Until we understand the frequency and pattern of mutation and the heteroplasmy it can cause, the amplification and sequencing of mtDNA to provide forensic evidence about identity will remain problematic at best and at times, may be completely unreliable.

CONTAMINATION

When asked in a courtroom about contamination, Wilson and DiZinno reiterated their writings that contamination is a potential source of error. For example in his cross-examination on August 28, 1996, Wilson was asked:

Q And if you'll follow me as I read and tell me if this is correct, "The most critical potential source of error in mitochondrial DNA sequencing is contamination," is that correct?

A Right. In my opinion, that's correct.

To understand the rest of the testimony, I would note that in the Ware case, as well as every other case where I have seen sequencing data (5 to date), the negative controls and reagent blanks produce DNA sequences. They usually are not high quality sequences and when quantified with capillary electrophoreses in those samples that are reported, they are reported as being quantified at less than 10% of the DNA in the evidence samples. This is not unexpected as the protocol calls for more PCR cycles on the evidence and negative controls than is customary in other kinds of forensic samples (36-38 for hair versus the 32 used to generate mtDNA sequences in blood for known samples). This low level contamination is discussed extensively in the FBI's published papers and protocol, but in the courtroom something else occurs. To continue with Wilson's August 28, 1996 cross-examination:

Q Is there any evidence of contamination in this case?

A No.

Q None?

A None.

Q Zero?

A No contamination, no.

Q Okay, what's a reagent blank?

A Reagent blank is a sample that you run as a control to show that a particular reaction is, is free from contamination, or lower than a certain threshold, so that you can call a sample the correct type.

Q Does a reagent blank have any DNA?

A Sometimes they do, yes?

Q Okay. Wrong question. Is it supposed to have any DNA?

A Well it's not an issue of supposing to or not.

Q Is it supposed to?

A It's not an issue of supposed to or not.

Q You can't answer that yes or no?

A Can't answer that yes or no.

Q Do you put DNA into a reagent blank?

A No.

Q Then how do you get DNA out of it?

A Well, in certain instances, there will be a low level of DNA in a reagent blank, and what we do is take it all the way through the process, try to determine its source. We have found that if the level stays low, beneath a certain threshold that we've experimentally determined, that even though there might be a little bit of DNA in a blank, in a reagent blank, we can go forward with the type of the sample and expect a reliable result, and that's what our validation study showed.

Q Isn't that by definition contamination?

A I think – I don't think of it in that way, no, as contamination. Contamination to me means potentially affecting your result.

Mr. DiZinno was asked a similar question but gave an entirely different answer during his cross-examination in a hearing on November 21, 1997:

Q Now, it's a fact, is it not, that mitochondrial DNA is more prone to contamination problems than nuclear DNA?

A It is, and let me explain that if I can.

Q Certainly.

A The reason we look at mitochondrial DNA is because there's more copies of it per cell than nuclear DNA. ... Also if there is contamination, let's say there is blood on the hair, well, there's always going to be a lot more mitochondrial DNA than nuclear DNA. In that respect, it is more susceptible to contamination because there is more there, per se.

A We run a number of controls with our samples to determine if the DNA that we're getting is the DNA from the hair and not from the contaminant. Also our validation studies were done to prove that. So if we see a DNA in one of those controls, we stop, we repeat, we do whatever we have to do to make sure that the DNA that we're sequencing is the DNA from the hair and not from a contaminant. I don't know if that answers your question?

Q Is it true, sir, that your lab allows an acceptable level of contamination?

A Yes.

Q And that would be where the contaminant is greater than or equal to ten to one or ten percent of the overall sample?

A That's correct, and can I explain that?

(I would note here that the context of this discussion makes it clear that Dr. Dizinno understood that the contaminant needed to be less than 10% rather than greater than 10% of the target DNA even though the attorney asked the question the other way round).

Q You certainly can.

A Again, as I said before, we run these controls with the hair, a number of controls, where we expect not to see contamination. Being that there's so much DNA present, there may be contamination on – what was on as the hair possibly got in the extraction process that we may see DNA now sequenced in one of the controls.

Well the bottom line is really how does that sequence from the control affect the DNA from the sample? We want to be sure that we're getting the DNA from the sample, not from anything else.

So as part of our validation studies, we did studies where we controlled known amounts of contaminant to the hair. And what we found is actually anywhere from six to one or eight to one, a ratio, meaning that eight or six parts per hair, one part contaminant, we would start to see the actual sample from the hair, not from the contaminants.

If it was two to one, something like that or lower, we would get a mix or we wouldn't get the true type from the hair.

So in an effort to be conservative, we're going to ten to one. It has to be ten to one or greater in the controls and the sample has to be different from the sequence and the controls has to be different than the sample for us to allow for that.

So we saw it. We know that sometimes on occasion we have it, and we developed a method through our validation study to deal with it. So again, we're sure that the sample, that the sequence we are obtaining is the sequence from the forensic sample, not from a contaminant."

Q You probably anticipated I was going to ask you the that question.

A Probably.

One would presume if Di Zinno had anticipated that question he would have gotten the parameters of the experiment reported in Wilson et al. (15), a paper he co-authored, correct, rather than reporting them incorrectly, but in a fashion that makes it appear that more contaminant is necessary to disturb accurate sequencing than was actually observed in his studies (see above). I realize that this might be a simple mistake in memory, but the fact that the numbers presented are wrong in 2 directions, and both times favor a conclusion of increased reliability for the FBI's protocol is problematic.

In addition, one would presume that the two individuals then in charge of the mtDNA lab at the FBI would have gotten together on what is and is not contamination. DiZinno readily admits that DNA in negative controls is contamination, but that the lab accepts a specific low level of contaminant (> 10:1, target:contaminant) proven to his satisfaction by the validation studies to be unimportant. In contrast, Wilson, who says the same thing about contamination in the papers he has written, has redefined contamination to mean only that level of "unexpected" DNA in negative controls that he believes would affect the result for courtroom purposes. In this way he avoids the threatening word entirely when

speaking to judge or jury, except when confronted during cross-examination when he asserts that it does not exist in a case, when it demonstrably does, and his colleague DiZinno would admit that it does while stating it is below the threshold where he would worry about it.

Finally, one would presume that validation studies that “proved” the FBI protocol works the way that the FBI asserts it works, would be based on sufficient samples and done with sufficient rigor to warrant some faith in that conclusion. The sample sizes of experiments exploring contamination, however, are inadequate to say much of anything. For example, the absence of ambiguous base calls or errors in sequencing at 10:1 ratios of target to contaminant DNA in a sample of 5 tests of this hypothesis is consistent with a true “error” rate of anywhere from 0% (their claim) to 35% errors (someone’s nightmare, Table 1). To suggest this is a conclusive test is the logical equivalent of watching Michael Jordan hit 5 free throws in a row and concluding that it is safe to assume that he *never* misses. If the true error rate was as low as 5%, which would demonstrably impact the reliability of the process, they would need a considerably larger sample size than 5 to adequately test the hypothesis (Table 1). The same sample size problem afflicts their more “forensically relevant” studies of external contaminants of hair. Their claim that their tests of the hypothesis that their new cleaning method “always” works with each type of contaminant is based on inadequate samples (blood (N=4), saliva (N=3), semen (N=5), see above).

The bottom line is that contamination issues are more important when doing mtDNA sequencing than with any other kind of forensic DNA typing. The very small amounts of material in the usual evidence (e.g., single hairs) means that equally small amounts of contamination can invalidate the sequencing process. In part this stems from the high copy number of mtDNAs, and in greater part from the high number of PCR cycles used to generate sequences for typing. That this is not an academic

problem is indicated by the fact that negative controls and reagent blanks regularly amplify human DNA in the FBI's mtDNA lab (Yes, Virginia, DNA does fly- with apologies to Rock Harmon) and have been shown to reduce the reliability of mtDNA sequencing. The FBI reports that at ratios of 4:1, both errors and ambiguities (which now might be attributed to heteroplasmy rather than contamination) occurred, and ambiguities even occurred at 8:1 ratios (see above). Since no sample sizes are given for this experiment I presume the sample size was one. It remains possible that errors would have turned up at 8:1 or even 12:1 if enough had been tested. Before they changed their washing method they obtained mtDNA sequence from the blood contaminating a hair, rather than from the target hair itself. Luckily, since they knew the sequences of both, they could identify the problem and the problem emerged with a sample size of one. In casework, such an error might be undetectable and thus could result in a false positive or negative. Finally are 12 hairs (4 with blood, 3 with saliva, and 5 with semen) enough to safely conclude that the new cleaning methods always work, even if you test each hair 4 times? The repeated testing of the same contaminated hair does not increase the valid sample size, but is itself a form of pseudoreplication which gives more weight to an experiment than it deserves (22).

While the FBI protocol is designed to deal with laboratory contamination, detected in reagent blanks and negative controls, and some forms of external contamination of hair, even they admit that undetected contamination is still possible. For example, contamination can occur at a crime scene before the evidence is brought into the lab. Wilson et al. (16, p 71) reported, "one sample from cat blood initially amplified and sequenced, but the result appeared to be of human origin. This suggests that the initial stain was contaminated with human DNA." They also noted: "only one instance (cat blood) of undetected human DNA contamination was found." In my opinion, the flaws in their validation studies, especially the inadequate sample sizes in tests of critical hypotheses that are the foundation of their

protocol, make the entire enterprise of mitochondrial DNA amplification and sequencing as done by the FBI potentially unreliable, and in fact, at present invalid. This is not to say that I believe that additional rigorous testing, with adequate samples is impossible. It might be the case that a comprehensive and well designed validation study would unequivocally demonstrate that the current protocol, or one like it, would manage the problem of contamination in exactly the fashion claimed by the FBI. On the basis of the data in hand, however, such a conclusion is presently premature and unwarranted.

SEQUENCE MATCH FREQUENCIES

Assuming for the moment that the potential problems associated with heteroplasmy and contamination do not apply in a specific case, then we are still left with the question of what it means when the mtDNA sequence from a crime scene hair matches the sequence from a known individual (usually the defendant) or a maternal relative of a presumed victim. As with any genetic analysis, the fact of a match needs to be weighted for reliable interpretation. The audience, whether judge, jury, or student, needs to know how rare or common such matches are expected to be. What we can say, *a priori*, is that since the mtDNA acts as a single locus, and the variation in the control region is not distributed at random in the about 600 base pairs usually sequenced, mtDNA typing is “not as statistically informative” as nuclear DNA typing(8). It is expected to show discriminatory power that is reduced by orders of magnitude relative to other forms of forensic DNA analysis (Table 2, in 17).

What do the forensic expert witnesses say?

Direct of Wilson, August 28, 1996

A (In Part) “Last step of the process is if we determine that the sequences are the same, and we look at all these 600 bases, we try to give the, the Courts an idea of how rare it is. The way we do this is we take the sequence and we look at a computerized

database of the sequences that have been generated from our lab and other labs, and we ask the question, how rare is it, how common is it. And for example, here, if there's a hundred people in this picture and it's only been seen once, we just say it's only been seen once and we've sequenced a hundred individuals, to keep it simple."

Q. How rare is the sequence that, that you ran with regard to Paul Ware?

A. We have a database that consists of a total of 742 individuals. 319 of these individuals are from African descent and 423 are from Caucasian descent, European or Caucasian descent. This particular sequence that we obtained from K6, the saliva sample and the two hairs have not been observed before in either the Caucasians or the African sequences, so it simply had not been observed before.

Continuing with the Cross-examination of Wilson on August 28, 1996, he discussed the number of base pair differences between individuals:

A The average number of differences between any two Caucasian individuals is approximately six.

Q Okay. Well, so you're telling – and I think I am understanding this, but it's been a while since I took a biology course – what you're telling us is that one out of every 424 people, at least in your database, have this?

A No, no, I'm not saying that at all.

Q What are you saying then?

A All I'm saying is that we have a database of a certain size, and this particular sequence has not been observed in this database before.

I am not saying that it's a particular frequency, one over this or that, because it cannot be expressed that way because the database is not large enough at the present, in its present form, present size to be able to assign a frequency, you know, like one percent or whatever.

This event would have to be observed many more times in order to assign it a frequency, so what we do is we just state a fact. Instead of a frequency, we give a fact, and the fact is here's the size of the database and here's the number of times it's been observed before.

Q You can't say that's his hair, can you, from your stuff?

A I'm not saying that this is his hair –

Q Okay.

A -- to the exclusion of all others, no.

Q Okay.

A. I'm saying that we have a technique here which is very discriminating which points to the conclusion that this hair originated from the defendant.

Shortly after this, in a report dated 11/20/96 in a different case (PA v Dillon), the FBI states:

“Using the mtDNA database currently available to the FBI laboratory, the mtDNA sequence obtained from specimens Q2, Q7, and K4 has not previously been observed in 319 individuals of African descent, 334 individuals of Caucasian descent, nor in 90 individuals of Hispanic descent.”

In the preliminary hearing associated with PA v Rorrer, DiZinno was asked (August 1, 1997) during his direct:

Q. Now, do you have an opinion, Mr. DiZinno, to a reasonable degree of scientific certainty based on your microscopic examination and your mitochondrial DNA examination, as to the likelihood that it was not Patricia Rorrer whose hair was found in the victim's vehicle at the crime scene?

After some legal objections and discussion he answered:

A. Yes. The mitochondrial DNA sequence is unrelated or independent of the microscopic appearance of the hair, so there are two unrelated issues that you're looking at, the microscopic appearance of the hair and the mitochondrial DNA sequence. The sequence does not affect how the hair looks. ... and since these are independent events, when I put them together, I would say it's highly unlikely that the Q1, Q7 and Q8 hairs came from anyone else but Patricia Rorrer. Is it possible that someone else has the same mitochondrial DNA sequence and the same hair characteristics as Patricia Rorrer, as the Q1 hairs and Patricia Rorrer? It's possible. But I would say it would be highly unlikely.

After being asked about the database, he responded in part:

A At the time this case was done, our data base consisted of 742 individuals and it compared the sequence from the questioned known hairs, which is now one sequence since it's the same, and compared it to the 742 individuals in our data base. And they had not seen that sequence before in our data base of 742 individuals.

Q What is the size of the database now?

A Right now its 1,043.

Q And how many of those – can you break that down into ethnic backgrounds at all?

A Yes.

Q Is it done on the report, on your Commonwealth exhibit A?

A Yes.

Q 6a?

A In the report that contains the 742 individuals, the 742 individuals consist of 319 individuals of African descent, 333 individuals of Caucasian descent, and 90 individuals of Hispanic descent.

Q Now just one last question. Has this sequence that you found on defendant's blood and on the hair found in the victim's vehicle and at the scene where the victim's body was found, has that sequence turned up in any other of your – in your data bank other than defendant hairs?

A Not at this- we checked it – it originally was checked against 742 and I had not seen it. I recently checked it against 1,043 and had seen it once, but when I checked what sequence it was, it was the sequence from this case, which was put into our data base as a data base sample. So if you say the data base is 1,042, eliminating that one, I still have not seen that sequence before.

During the cross-examination, he testified:

Q Are you familiar with the British Home Office Laboratory's study that out of that 100 that were in the database, four had the exact same sequence of mitochondrial DNA?

A Um-hum. Yes.

Q So four unrelated, random individuals out of 100 had the exact same sequencing of DNA?

A I can give you a better statistic. In our database of 1,043, there are 89 that have been seen more than once.

Q So people sharing the same DNA sequencing?

A Yes.

Q And they're unrelated, random individuals?

A Yes. It can happen.

Later in the admissibility hearing for the same case (November 21, 1997), DiZinno testified again.

Q. ... you're not allowed to predict the frequency at which you observe different sequences within a data bank as it relates to the general population, is that correct?

A. That's correct, let me explain that a little bit if I can. Other DNA techniques, nuclear DNA techniques that have been used for years have very, very large databases. Tens of thousands, possibly even more. And because they have such a large database, if a sequence from a question source matches a known source or a type in sequence, then they compare that type to a much, much larger database. And then the population geneticist can calculate an estimate of what is the estimate of that type being seen in the

population. So it's a frequency, X percent of the population would be expected to have that sort of sequence. But our database is small. Originally, 742 now 1043. And we don't feel that it's enough for us to be able to say if it's been seen once before, or many times, zero before."

Q. So in other words, multiple people sharing the same sequence of bases in hypervariable region one and hypervariable region two is not necessarily uncommon?

A: I would say it's uncommon, 89 out of 1043, that's not very common.

On the same issue, during the same hearing (November 21, 1997), Stoneking testified on direct:

Q Maybe I'm a little confused on that. Is there actually a two-fold step? First determine the sequence and then weigh it against known sequences that are contained in the database?

A Right. Once we determine the match, then we compare it to the database that we have to get an idea of how frequent that particular sequence is.

A (In Part) ... If it doesn't match, you're done. If it does match, then you search the database to ask, what is the frequency, what is the likelihood that another individual would have the same sequence.

Later during cross-examination, he testified:

Q So I would assume that, in your opinion, you can – you consider mitochondrial DNA sequencing to be a property of ethnicity, is that correct?

A Well, I would say I consider that sometimes one can get information about ethnicity from an mitochondrial DNA sequence but not in every case.

Q So I guess what I'm asking you would be, would you believe it to be a greater chance of finding certain sequences in certain subpopulations of people?

A Yes.

Q At this time you believe there's enough research that's been conducted as to the potential that certain sequences may be more common than other sequences in order to render opinions as to frequencies of a sequence appearing in a population?

A I don't think the database is adequate to render an opinion as to an exact frequency, but I do think it is adequate to give us an idea whether a sequence is common or rare.

A The number 1,043 simply relates as being the current size of the database with respect to being a random sample of the existing sample population.

 Ignoring for the moment the simple accounting errors evident in the above testimony, with Wilson apparently classifying 90 Hispanics as Caucasians in 1996, and one 1997 report finding an unexplained extra Caucasian (334 versus 333) in the database, what can we conclude? I would summarize the testimony as: We have these databases, (e.g., Wilson and the reports) or one database (DiZinno and Stoneking) of sequences and we compare a forensic sequence to the database(s) and see if any match the evidence. We report the size of those databases and the number of times the sequence has been seen, eliminating the known and evidence sequences from the case at hand. We will, however, add those same known sequences from the case to the database(s) to increase our sample size. We absolutely refuse to say anything more about the likelihood of matching because our databases are too small to estimate frequencies accurately and directly and there is no simple way to estimate the likelihood of matching because our databases are too small to do so.

Despite the paucity of frequency data, they are willing to make statements about rarity and commonness of sequence matching and somehow in Karnak like fashion always conclude that such a match is unexpected due to its rarity. Wilson asserts that mtDNA analyses are “very discriminating”. I would ask, compared to what and how does he know? The danger, of course is that such a statement leaves the audience with the possible belief that mtDNA typing is similar to other forms of DNA typing, where matches are also described as rare, in its ability to discriminate among individuals. It is not that discriminating.

DiZinno exacerbates this problem because he is actually willing to testify that such matches are so rare that he can conclude that microscopically matched known and evidence hairs when combined with the “independently” matched mtDNA sequences in that evidence hair and that known person’s blood allows one to conclude: “with a reasonable degree of scientific certainty”, that it is highly unlikely that the hair came from someone other than the person on trial. How does he know that? Is he multiplying some *unknown* but positive frequency of hair matching with some equally *unknown* but positive frequency of mtDNA sequence matching to come up with a magically *known* joint probability that strongly implies though not proves identity?

Contra his repeated statements, the hair and mtDNA comparisons are not actually independent. Hair examiners regularly tell the world that one thing they generally can do is determine the ethnic background of someone based on hair. In other words hair characters vary between ethnic groups (e.g., Africans versus Caucasians versus Asians). Mitochondrial DNA sequence frequencies also vary among the same ethnic groups, such that mtDNA matching is more likely within one’s ethnic group than across such groups (see Stoneking, above, and Tables 2 and 3 below). You cannot multiply

the probabilities of these two events, whether they have been estimated, or exist only in the examiner's mind, as appears to be true here.

Just as important, the implication that one cannot estimate objectively the likelihood of matching based on these small databases is wrong. The British forensic scientists do it based on pair-wise comparisons using the same databases (e.g., 9 and 24), and the entire forensic community appears to recommend it when dealing with the probability of mutation (7). DiZinno is simply wrong when he states that the databases associated with frequency estimates in nuclear DNA testing are much larger than the mtDNA databases. In fact, many are considerably smaller on a per locus basis with sample sizes of 90-120 individuals common in databases used to estimate frequencies for PCR loci including many of the short tandem repeats (STR's).

The counting method was first suggested for looking at match frequencies when databases were not large enough to estimate multilocus genotype frequencies for nuclear DNA typing. The disingenuous method used when reporting mtDNA sequence matches in court (It hasn't been seen before in 1072 individuals) is misleading, if not downright dishonest. When a database contains 90 individuals it is not likely to contain every possible sequence if the average frequency of matching is as high as 0.1 or 0.2, and cannot contain every sequence if the match frequency is .01 or less. To state that a sequence has not been seen under such circumstances, without an explanation is wrong. To state that a sequence has not been seen in 1072 samples is actively misleading. The common sense interpretation of such evidence is that the probability of matching is less than 1 in 1072 and if the listener wrongly associates this test with other DNA tests could easily be mistaken for 1 in millions or billions. Shouldn't the proponent of such evidence make sure that this error is not possible?

The method of presentation also fails to account for ethnically dependent match probabilities and for the fact that the particular sequence has always been seen once before- in the known individual. A less biased way to present such information, until the databases are large enough to have sampled all sequences at their true frequencies is to state that it has been seen in 1 out of 334 Caucasians (the 333 unmatched in the database plus the new sequence added to it) and in 0/319 individuals of African descent with the appropriate confidence limits on those empirical frequencies (e.g., $=(1/N+1) \pm CI$, where N is the size of the ethnic database containing the new sequence, and $0/N \pm CI$, where N is the size of the ethnic databases that do not contain the new sequence). I believe that the FBI scientists know this, how else can one explain putting such a sample in the database and using it in some cases, but then removing it to provide a “better” number and stating that it has not been seen before except in the defendant in this case.

That mtDNA is not as probative as other forms of DNA analysis is not disputed by anyone on a theoretical level. So why isn't this reported routinely in the courtroom by the proponents of the evidence? The way it's currently presented in courtrooms in the U.S. instead is actively misleading. One can get a feel for this by doing what some of the same scientists who have written and testified on the issue of mtDNA recommended when examining nuclear DNA databases and what the British have done with their mtDNA databases. By performing pair-wise comparisons on all of the members of a database, we might not be able to answer the question of exactly how likely is it that someone other than maternal kin would match *this* evidence sequence, but we can say how often we see any two individuals in the database with matching sequences. When this is done, matching is not a rare event, and mtDNA typing is not very discriminating relative to a multilocus match with nuclear DNA.

To examine match frequencies, I used the concordance of mtDNA sequences from the Internet in January, 1998 (World Wide Web, <http://sscf.ucsb.edu/~kmiller>, 23). The concordance lists sequences based on HV1 sites from 16,024-16,365 and HV2 sites 000073- 000340, the same as the FBI, and that includes most of the 742 person database used by the FBI in casework. I counted the number of individuals and sets of individuals with matching sequences and noted to which subgroups within the database they belonged. I discovered that the database was not quite as advertised in testimony- with the African descent portion corresponding to 122 Africans from Africa, 115 individuals of African-Caribbean descent from the U.K., and 90 African Americans from the U.S., (I cannot account for the two “missing” individuals). It also included 90 Hispanics, 100 Caucasians from the U.K. and 233 Caucasians from the United States (Tables 2 and 3). When looked at as individuals, between 15% and 37% of the people in each database matched at least one other person in the same database (Table 2). Matching was much more likely within, than between subpopulations, when major groups were compared and among the African subgroups as well (Table 2). This was not true of the Caucasian subgroups where the combined database had the greatest frequency of matching individuals owing to reasonably high levels of matching between the two subgroups (U.K. and U.S. databases, Table 2).

When pair-wise comparisons were done, the probability of anyone in the database matching someone else ranged from a high of 1 in 114 in the Hispanic database to a low of 1 in 468 in the African-Caribbean database (Table 3). In line with the assumption that matching is more likely within races and subpopulations, the probability of matching decreased in the combined databases (with the exception of the combination of African-Caribbean and Hispanic, where it increased, relative to the

African-Caribbean alone) reaching a minimum when all databases were combined (1 in 646 with N=750, Table 3).

The bottom line is that the words used to express the discriminatory power of mtDNA sequences are ambiguous at best. When someone claims that matches are “very rare” or that the method is “very discriminating” and points to the defendant as the source of the evidence, one can only ask: What does rare or discriminating in this context really mean? My problem is that because the word DNA typing is used, the judge, jury, or both might be lulled into concluding that the words rare and discriminating, which are used in both contexts, mean the same thing. In this sense the naive observer might bootstrap the mtDNA evidence with the greater discriminatory power of other forms of DNA evidence.

Even more important, now that the FBI and other forensic practitioners, are unwilling to exclude someone on the basis of a single base sequence difference, the true comparison should be how many sequences match another at every base or at every base but one. This number is considerably larger than total perfect matches, but to the best of my knowledge has not been computed or presented in court. Since the word match now encompasses perfect matches and one base mismatches, the probability of a random match must include both events as well.

It is especially problematic when listeners are actively misled about match probabilities. For example, DiZinno notes that there are 89 shared sequences in their combined database of 1073 individuals. His claim that matching is rare because 89 out of 1073 is not common (see above) is either an error or an active misrepresentation of the facts. Eighty-nine shared sequences means that at least 178 individuals share sequences not 89 as a reading of his testimony would allow some to conclude. *Caveat emptor* should not be the rule when offering scientific evidence. In fact, the true number of

individuals matching at least one other individual is about 300 (Wilson, testimony, January 29, 1998).

One cannot determine the exact number unless the evidence is provided the other way (i.e., how many unique sequences exist in the database). Since at least 226 individuals matched someone in the earlier database of 742 (30%), using the number of perfectly shared sequences and the total number of individuals in the combined database, the number 89 is simply wrong.

It is also misleading because match frequencies vary among and within races. To state how frequently a sequence from an African American defendant is seen in Caucasians, Hispanics, or even Africans from Africa is irrelevant if not downright biased against any defendant. The appropriate way to present such evidence is to note how many matching sequences (including those differing by one base with the new heteroplasmy driven match rules) have been found in African Americans (1 or 3 of $91 \pm$ Confidence Limits) who have been typed and how often in other groups. If the pair-wise match frequencies (adding single base mismatches and matches together) and the proportion of each group matching at least one other individual is presented, the judge or jury will be hearing an objective and unbiased representation of the weight of the evidence based on the data in hand. Anything else is either a guess or wishful thinking.

CONCLUSION

Regardless of my opinion, or anyone else's for that matter, the FBI validation studies were published in peer-reviewed journals. Similarly, scientists with good reputations have put their imprimatur on the FBI's resulting protocol. Since such peer review is considered an important precursor to admissibility under Daubert, such publication history assumes great legal import. For example, Stoneking has testified (November 21, 1997, Admissibility Hearing):

Q. All right. Now I believe it was brought out that you actually helped the FBI set up their lab?

A. I did discuss with them what would be involved in setting up the laboratory at the FBI when they initially became interested in this, yes.

Q. Are you aware of the protocol they have issued for the conduction of mitochondrial DNA tests in their lab?

A. Yes, I have seen their protocols.

Q. And is what the FBI does, is that generally accepted based on underlying theory and technique?

A. My opinion, yes, the protocols are based on sound underlying theory and techniques.

I cannot believe that Stoneking would have reached this conclusion if he had read the “validation” papers, where he was not a co-author, or heard some of the testimony of Wilson and DiZinno as they interpreted these critical issues. With respect to the peer review of those papers in those journals, I would note that the International Journal of Legal Medicine had a very short turnaround for acceptance of the validation paper, much shorter than the other papers on these issues published in other journals (e.g., Human Biology or Nature Genetics). Wilson testified that he never saw hard copy reviews of the “Validation” paper but only spoke to the editor on the phone about suggested revisions. I wish I could publish an important foundation paper in such perfect condition that only minor revisions were required. Similarly, the Biotechniques paper was submitted as a techniques paper rather than as a validation paper and might have been reviewed as such. I cannot believe that any statistically competent reviewer would have accepted the small samples ($N's < 5$) presented as reliably validating the 10:1 rule

for contamination. Nonetheless, it is now cited in court by FBI forensic scientists as a paper reporting on one aspect of the “extensive” validation studies that they have done. The original assumption of homoplasmy was based on similar experiments and small sample sizes (15-16), and was, in short order, disproved as the relevant sample size increased (6-7). Why should anyone assume that it will be different with respect to the potentially disturbing influence of contamination?

In my opinion, the mitochondrial DNA typing protocol used by the FBI to generate evidence about identity in criminal trials and especially the testimony itself are flawed. The so-called validation studies used as a foundation for the protocol are themselves unreliable, often suffering from sample sizes that are too small to rigorously test the hypotheses they have been reported to address. Courtroom testimony of the proponents of the technique, as well as their early published outlines of appropriate validation procedures underscore the importance of rigor in such studies. The current statistical presentation of the likelihood of random matching in the courtroom is biased and invalid as well. Finally, the emergence of significant levels of heteroplasmy at frequencies and with patterns that have not yet been characterized “suitably” much less definitively, is a nearly insurmountable problem undermining the reliability of this form of identity testing with this kind of evidence.

I would not argue here that mtDNA evidence should be inadmissible, for that is a legal rather than a scientific question. I do believe that the technique has not yet been validated well enough to insure its reliability as evidence of identity. I would not dispute the current usefulness of the technique as an investigatory tool. With proper care it can offer strong evidence of exclusion (with 2 or more base mismatches as long as care is taken to exclude heteroplasmy and contamination as the reason for a non-match). I do feel that until we are able to provide definitive evaluations of the effects of mutation,

contamination and appropriate statistical tools to provide a reliable estimate of the frequency of random sequence matching, it is dangerous to use mtDNA sequence matching as positive evidence of identity.

A review of what was proposed by the FBI as necessary validation considerations, “before mtDNA sequencing is used in forensic casework.” (8, p 72, and see above for direct quotes), shows that, in fact, human hairs from the same individual can differ in their sequence, sequences can differ among different tissues from a single individual, and sequence information is not necessarily identical among maternal relatives. In addition, a review of the manner in which they present match likelihood and fail to caution about the reduced discriminatory power of mtDNA relative to the other forms of forensic DNA analysis, shows that when presented in court such evidence can be positively misleading. If the trier of fact assumes or is led to believe that mtDNA has been validated to the same degree and has the same discriminatory power as nuclear DNA typing, they will have reached a demonstrably erroneous conclusion. In the United Kingdom, the use of mtDNA is currently used primarily for investigation and exclusion and if it is brought to court the major differences between mtDNA and other forms of forensic DNA analysis are presented “up front”. In conclusion, then, if the FBI scientists agree with their own published recommendations in the “Guidelines” (8), they would conclude that mtDNA sequencing should not be used yet in forensic casework.

ACKNOWLEDGMENTS

I would like to thank numerous colleagues including, Jonathan Kohler, David Werret, Barry Scheck, Mitchell Holland, Barbara Hager, and especially William Thompson for general or specific comments on the manuscript or on the presentation at the 1998 Meeting of the American Association for the Advancement of Science meeting, on which the manuscript is partially based. Remaining errors or stubbornness are entirely my own.

REFERENCES:

1. National Research Council Committee, "DNA Technology in Forensic Science." *National Academy Press*. Washington D.C., 1992.
2. Kelly, J. F., and Wearne, P. K., "Tainting Evidence: Inside the Scandals at the FBI Crime Lab." *Free Press*, 1998. New York, New York.
3. Higuchi R., von Beroldingen, C. H., Sensebaugh, G. F., and Erlich, H. A. "DNA Typing from Single Hairs," *Nature*, Vol. 332, 1988, pp. 543-546.
4. Paabo S., Gifford, J. A., and Wilson, A. C. "Mitochondrial DNA Sequences from a 7000-Year Old Brain." *Nucleic Acids Research*, Vol. 16, 1988. pp. 9775-9787.
5. Budowle, B., Adams, D. E., Comey, C. C., and Merrill, C. R. "Mitochondrial DNA: A Possible Genetic Material Suitable for Forensic Analysis." *DNA and Other Polymorphisms in Forensic Science*. H. C. Lee and R. E. Grenslen, Eds. Year Book Medical Publishers, Chicago, IL., 1990. pp. 76-97.
6. Wilson, M. R., Polanskey, D., Repogle, J., DiZinno, J. A., and Budowle, B. "A Family Exhibiting Heteroplasmy in the Human Mitochondrial DNA Control Region Reveals Both Somatic Mosaicism and Pronounced Segregation of Mitotypes." *Human Genetics*, Vol. 100. 1997. pp. 167-171.
7. Parsons, T. J., Muniec, D. S., Sullivan, K., Woodyat, N., Alliston-Greiner, R., Wilson, M. R., Berry, D. L., Holland, K. A., Weedn, V. W., Gill, P., and Holland, M. M. "A High Observed Substitution Rate in the Human Mitochondrial DNA Control Region." *Nature Genetics*. Vol. 15. 1997. pp. 363-367.

8. Wilson, M. R., Holland, M. M., Stoneking, M., DiZinno, J. A., and Budowle, B., "Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science." *Crime Laboratory Digest*. Vol. 20. 1993. pp. 68-77.
9. Poulton, J., "Transmission of mtDNA: Cracks in the Bottleneck." *American Journal of Human Genetics.*, Vol. 57, 1995. pp. 224-226.
10. Holland, M. M., Fisher, D. L., Mitchell, L. G., Rodriguez, W. C., Canik, J. J., Merrill, C. R., and Weedn, V. W. "Mitochondrial DNA Sequence Analysis of Human Skeletal Remains: Identification of Remains from the Vietnam War." *Journal of Forensic Sciences*. Vol 38. 1993. pp. 542-553.
11. Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagleberg, E., Sullivan, K., "Identification of the Remains of the Romanov Family by DNA Analysis." *Nature Genetics*. Vol. 6, 1994. pp. 130-135.
12. Ivanov, P.L., Wadhams, M. J., Roby, R. K., Holland, M. M., Weedn, V. W., and Parsons, T. J. "Mitochondrial DNA Sequence Heteroplasmy in the Grand Duke of Russia Georgij Romanov Establishes the Authenticity of the Remains of Tsar Nicolas II." *Nature Genetics*. Vol. 12. 1996. pp. 417-420.
13. Sullivan, K. M., Hopgood, R., Lang, B., and Gill, P., "Automated Amplification and Sequencing of Human Mitochondrial DNA." *Electrophoresis*, Vol. 12, 1991. pp. 17-21.
14. Sullivan, K. M., Hopgood, R., and Gill, P., "Identification of Human Remains by Amplification and Automated Sequencing of Mitochondrial DNA. *International Journal of Legal Medicine*, Vol. 105, 1992. pp. 83-86.

15. Wilson, M. R., Polanskey, D. P., Butler, J., DiZinno, J. A., Repogle, J., “Extraction, PCR Amplification, and Sequencing of Mitochondrial DNA from Human Hair Shafts.” *Biotechniques*. Vol. 14, 1995. pp. 662-669.
16. Wilson M. R., DiZinno, J. A., Polanskey, D., Repogle, J., Budowle, B., “Validation of Mitochondrial DNA Sequencing for Forensic Casework Analysis.” *International Journal of Legal Medicine*, Vol. 108. 1995. pp. 68-74.
17. Weedn, V. W., “Forensic DNA Tests.” *Clinics in Laboratory Medicine*. Vol. 16, 1996. pp. 187-196.
18. Gibbons, A., “Calibrating the Mitochondrial Clock.” *Science*. Vol. 279, 1998. pp. 28-29.
19. Howell, N., Kubacka, I., and Mackay, D. A., “How Rapidly Does the Human Mitochondrial Genome Evolve.” *American Journal of Human Genetics*., 1996. pp. 501-509.
20. Jazin, E., Soodyall, H., Jalonen, P., Lindholm, E., Stoneking, M., and Gyllensten, U., “Mitochondrial Mutation Rate Revisited: Hot Spots and Polymorphism.” *Nature Genetics*. Vol. 18. pp. 109-110.
21. Parsons, T. J., and Holland, M. M., “Response to Jazin et al.”, *Nature Genetics*. Vol. 18. p. 110.
22. Hurlburt, S. H., “Pseudoreplication and the Design of Ecological Field Experiments.” *Ecological Monographs*. Vol. 54, 1984. pp.187-211.
23. Miller, K. W. P., Dawson, J. L., and Hagelberg, E.. “A Concordance of Nucleotide Substitutions in the First and Second Hypervariable Segments of the Human mtDNA Control Region.” *International Journal of Legal Medicine*, Vol. 109, 1996. pp. 107-113.

24. Piercy, R., Sullivan, K. M., Benson, N., and Gill, P., "The Application of Mitochondrial DNA Typing to the Study of White Caucasian Genetic Identification." *International Journal of Legal Medicine*, Vol. 106, 1993. pp. 85-90.

Table 1. Sample Sizes Needed to Estimate the Expected Frequency of Critical Error Events.¹

True Frequency of Errors Consistent with a Sample of Size (N) with No Errors	N, Sample Size of Experiment with Zero Errors Consistent with True Frequencies > 0
.01	268
.05	51
.10	24
.35	5

¹ Assume that the sample gives a frequency of zero occurrences of error or ambiguity (i.e. no errors or ambiguity is observed). Given an $a = 0.05$ and a $b = 0.10$, what sample size would allow one to conclude that a *sample* of zero errors (a frequency of zero) could have occurred by chance given the specified true frequencies of error in the total population

Table 3. Pairwise Comparisons of mtDNA Sequences Within and Between FBI Databases Included in the Worldwide Concordance. ¹

DATABASE	P/N	R	1/R
African Database (N=122)	51/7,381	.0069	1/145
African American (N=90)	13/4,005	.0032	1/308
African Caribbean (N=115)	14/6,555	.0021	1/468
Hispanic (N=90)	36/4,005	.0089	1/114
US Caucasian (N=233)	131/27,028	.0048	1/206
UK Caucasian (N=100)	20/4,950	.0040	1/248
Total Caucasian (N=333)	257/55,278	.0046	1/215
AFR+AFAM (N=212)	54/22,366	.0024	1/414
AFR+AFCR (N=237)	53/27,966	.0019	1/527
AFAM+AFCR (N=205)	34/20,910	.0016	1/615
AFCA+HISP (N=205)	51/20,910	.0024	1/414
Combined (N=750)	435/280,875	.0015	1/646

¹ **Random Match Probability, R = P/N, where P = frequency of pairs of identical sequences, and**

N = Total Number of Pairwise Comparisons, which equals, $n*(n-1)/2$, where n is the number of individuals in an individual or combined database.

Table 2. Number of Individuals in Different Populations with mtDNA Sequences Identical With at Least One Other Individual in the Designated Population.

	US CAUC	UK CAUC	TOTAL CAUC	HISP	AFAM	AFCR	AFRIC
N	233	100	333	90	90	115	122
USC	74(0.32)	61	82	3	0	22	0
UKC	30	29(0.29)	41	7	0	5	0
TOTC	NA	NA	123(0.37)	10	0	27	0
HISP	1	5	5	21(0.23)	1	2	0
AFAM	0	0	0	1	18(0.20)	6	2
AFCR	6	4	6	2	9	17(0.15)	2
AFRI	0	0	0	0	1	2	37(0.30)

The numbers in parentheses along the diagonal is the proportion of individuals in each database that match at least one other member of that database.