Mitochondrial DNA Sequence Analysis — Validation and Use for Forensic Casework

M. M. Holland and **T. J. Parsons**
The Armed Forces DNA Identification Laboratory
Office of the Armed Forces Medical Examiner
The Armed Forces Institute of Pathology
Rockville, Maryland
United States of America

TABLE OF CONTENTS

1042-7201/11-01-99/21–50/$15.00 •

Mitochondrial DNA Sequence Analysis — Validation and Use for Forensic Casework [a,b]

REFERENCE: Holland MM, Parsons TJ: Mitochondrial DNA sequence analysis — Validation and use for forensic casework; *Forensic Sci Rev* 11:21; 1999.

ABSTRACT: With the discovery of the polymerase chain reaction (PCR) in the mid-1980's, the last in a series of critical molecular biology techniques (to include the isolation of DNA from human and non-human biological material, and primary sequence analysis of DNA) had been developed to rapidly analyze minute quantities of mitochondrial DNA (mtDNA). This was especially true for mtDNA isolated from challenged sources, such as ancient or aged skeletal material and hair shafts. One of the beneficiaries of this work has been the forensic community. Over the last decade, a significant amount of research has been conducted to develop PCR-based sequencing assays for the mtDNA control region (CR), which have subsequently been used to further characterize the CR. As a result, the reliability of these assays has been investigated, the limitations of the procedures have been determined, and critical aspects of the analysis process have been identified, so that careful control and monitoring will provide the basis for reliable testing. With the application of these assays to forensic identification casework, mtDNA sequence analysis has been properly validated, and is a reliable procedure for the examination of biological evidence encountered in forensic criminalistic cases.

KEY WORDS: DNA sequencing, forensic science, mtDNA, PCR.

INTRODUCTION

The role of DNA profiling in forensic investigations has become increasingly important as the analysis techniques have evolved. Cases such as the Nicole Brown Simpson murder trial, the identification of Nicolas Romanov II, the last Russian Tsar, and the identification of the Tomb of the Vietnam Unknown Soldier have brought DNA profiling to the forefront of public awareness. The expanded use of DNA profiling has significantly assisted the investigation of crimes, and has allowed for the identification of human remains where other methods have failed. As a result, the DNA recovered from a minute bloodstain found at the scene of a crime can be associated with a suspect through DNA profiling. Biological specimens (e.g., stains or hair) found at a suspect's residence can be associated with a victim. In many cases, DNA testing can play a critical role in establishing the identity of the victim. Although alternative biological-based identification methods exist, such as blood group markers and polymorphic protein variants [46,55,234], DNA profiling remains the most powerful and robust method of identification since the discovery of the human fingerprint [222].

The majority of DNA profiling methods currently performed in forensic laboratories analyze chromosomal DNA found in the cell nucleus (i.e., nucDNA). Variations in the length of specific nucDNA fragments, generated using either restriction digestion (RFLP: restriction fragment length polymorphism) [15,22,29,108,162,203] or polymerase chain reaction (PCR) amplification (AmpFLP or AFLP: amplified fragment length polymorphism) [27,30,54,72,190], are the most commonly used methods to differentiate between individuals. The most commonly used AmpFLP's are the short tandem repeats (STRs). The length of a DNA fragment is dependent on the composition of tandemly repeated sequences (i.e., a variable number of tandem repeats or VNTRs). The more "repeat units", the larger the DNA fragment. Individuals will have two fragments or "alleles", representing one inherited allele from the mother and one from the father. The inherited alleles can vary in length, resulting in a heterozygous banding pattern (i.e., "profile"), or the alleles can be of the same length, resulting in a homozygous profile. A marker, or forensic DNA "locus", is considered "discriminating" if there are a number of different alleles observed in the population (e.g., 5–10 or more), and if a large percentage of individuals in the population have heterozygous profiles (e.g., 70–95%). In turn, the more loci analyzed to generate a DNA profile, the more likely the biological specimen originated from a single source. Thus, the interpretation of a DNA profile "match" between the evidence and an individual has great significance, as most multi-locus RFLP or AmpFLP DNA profiles will conclusively identify the origin of the specimen, making nucDNA markers the gold standard or method of choice in the forensic community [42,50,72,84].

[a]The opinions and assertions contained herein are solely those of the authors and are not to be construed as official or as views of the U.S. Department of Defense or the U.S. Department of the Army.

[b]Editor's note: Comments on this article will appear in a forthcoming issue of this journal.

RFLP and PCR-based DNA profiling systems have survived the scrutiny of both the academic and legal communities, and although the reliability and acceptability of the procedures are still attacked in courts of law on a routine basis, there is now a general acceptance of both the methodologies used to generate a nucDNA profile, as well as the statistical weight placed on a profile [123,164]. In particular, AmpFLP analysis has routinely been admitted as evidence, given that PCR-based testing and length-based fragment analysis have already been generally accepted in the courts; in addition to length-based differences, sequence specific polymorphisms can be detected with PCR in conjunction with the reverse dot blot formats (i.e., AmpliType™ PM and HLA DQA1), a PCR-based method widely accepted in the U.S. legal system [39,83,176,189]. Given these efforts, crime laboratories currently have the ability to evaluate biological evidence with great accuracy, precision, and reliability.

Recent years have seen the advent of mitochondrial DNA (mtDNA) profiling, with a greatly increasing use and emphasis. In addition to nucDNA, mitochondria found in the cytoplasm of a cell contain a second human genome, the mtDNA genome. Although nucDNA profiling is highly informative and the method of choice if available, the mtDNA genome contains useful information which can be used to help establish identity or the source of a biological specimen. mtDNA has two primary advantages over nucDNA. First, on average there are thousands of copies of mtDNA in each cell compared to two copies of nucDNA, making mtDNA analysis a more sensitive assay, and thus, more successful on highly degraded specimens (e.g., old skeletal material and hair shafts). Second, mtDNA is maternally inherited, increasing the range of references available for the identification of human remains, e.g., distant maternal relatives become potential sources for comparison. As a result, a great body of mtDNA forensic profiling has already been performed, with a growing number of laboratories performing the analysis, and with the presentation of mtDNA in the courtroom both in the U.S. and around the world.

Although much is known about mtDNA and the techniques used to analyze the locus, forensic mtDNA profiling is still relatively new, and has some unique features when compared to nucDNA, so it is still being challenged in admissibility hearings. Thus, it is the purpose of this paper to review the characteristics and biology of mtDNA, and based on these characteristics, to describe how mtDNA data is being interpreted forensically. We will demonstrate that the component methods used to perform mtDNA analysis (i.e., DNA extraction, PCR amplification, and sequencing of DNA) are well established and validated, and that combined, these methods have been used to study mtDNA recovered from a wide range of forensically relevant biological sources, including ancient specimens (e.g., mummified soft tissue, bone, and hair). In turn, mtDNA analysis has been used successfully in a variety of forensic identification cases. As a result, mtDNA analysis has become a validated, robust, reliable, and well established forensic DNA profiling system.

I. MITOCHONDRIAL DNA (mtDNA) BIOLOGY AND BACKGROUND

A. Characterization of the Locus

In addition to the nucDNA genome, human cells contain additional genetic elements within mitochondria. Mitochondria are double-membraned organelles present in the cytoplasm, and are the site of many crucial metabolic processes such as oxidative phosphorylation. For this reason, the mitochondrion is often referred to as the energy powerhouse of the cell. It is now established that present-day mitochondria are derived evolutionarily from an ancient bacterial ancestor that at one time formed an intracellular symbiosis with early eukaryotic (or pre-eukaryotic) cells [61]. In the ensuing hundreds of millions of years, this ancestor lost the ability to function as an independent organism. The ancestral genome is now greatly attenuated, so that most of the functional proteins of mitochondria are coded for by genes present in the nucleus [124]. What remains within the human mitochondrion is a ~16,569 bp circular genome that encodes 37 densely packed genes (**Figure 1**). Twenty-two of the genes encode transfer RNAs (tRNA), 2 encode ribosomal RNAs (12S and 16S rRNAs), and 13 encode protein enzymes involved in the electron transport chain of oxidative phosphorylation and ATP production [228]. The mitochondrial genome has been extensively characterized from the standpoint of function, population variation, and genetic disease. Deletions, duplications and numerous point mutations have been identified within the coding region that are cause to various pathological syndromes, and the progressive accumulation of some of these in normal adults is thought to contribute significantly to degeneration with aging (extensive reference lists are in [9,227,228]).

The primary sequence of the mitochondrial genome was determined in 1981 [6]. The two strands of mtDNA have significantly different base compositions, with a pyrimidine-rich "light strand," and a purine-rich "heavy strand." In addition to the coding regions of mtDNA, there is only one significant non-coding section, the control region (CR). The CR is also sometimes called the D-loop, so named for structures visible by electron microscopy that are formed during mtDNA replication. The CR is

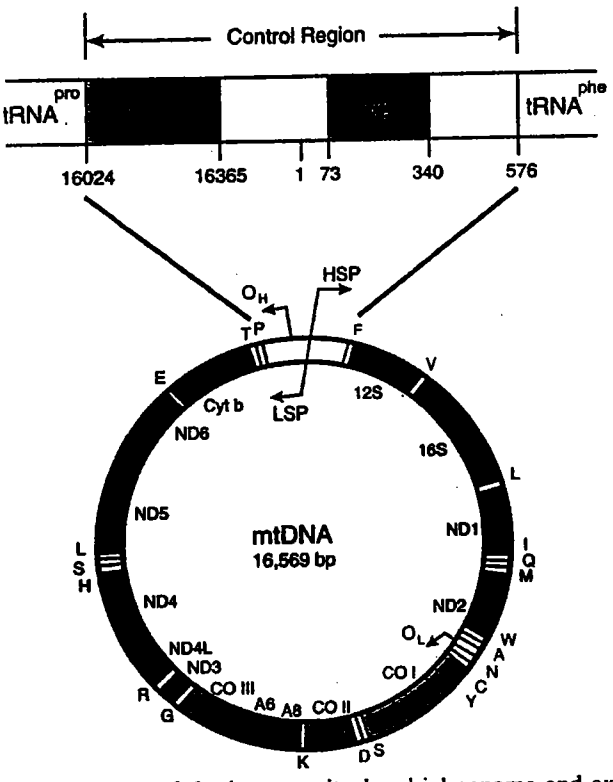

Figure 1. Map of the human mitochondrial genome and expanded diagram of non-coding control region. Listed are the genes for 12S and 16S ribosomal RNAs, subunits of the NADH-coenzyme Q oxidoreductase complex (ND), cytochrome *c* oxidase complex (CO), cytochrome *b* (cty b), ATP synthase (A), and 22 tRNAs (labeled with single letter amino acid code). The light strand (O_L) and heavy strand (O_H) origins of replication, and the light strand (LSP) and heavy strand (HSP) transcriptional promoters are shown by arrows. Control region diagram shows flanking tRNAs and location of hypervariable region 1 (HV1) and 2 (HV2); numbering system follows that of the standard reference sequence [6].

approximately 1125 base pairs in length, and contains promoters for polycistronic RNA transcription of genes on both the light and heavy strands, as well as the origin of DNA replication for the heavy strand. The base numbering system of the standard reference sequence (or Cambridge Reference Sequence, CRS [6]) begins arbitrarily near the middle of the control region, so that the control region spans positions 16,024 to 16,569, then continues from base position 1 to 576 (Figure 1). Relative to the coding regions, portions of the CR are highly variable among individuals, presumably due to decreased selective constraint of the non-coding DNA. For this reason, forensic identity testing has so far focused on sequence variation within two hypervariable regions of the CR (e.g., [89,172,238]). Hypervariable region 1 (HV1) extends from position 16024 to ~16365, while hypervariable region 2 (HV2) extends from ~73 to ~340 (the boundaries are not rigidly defined, and vary among particular studies or laboratories).

B. Copy Number and Inheritance

A characteristic of mtDNA that is of great advantage in forensic testing is the high copy number per cell. Whereas nuclear DNA is present in two copies per diploid cell, a mature oocyte is estimated to have thousands of mitochondria and more than 100,000 copies of mtDNA [152,178]. Somatic cells range from having ~200-1700 mtDNA copies depending on tissue type [19,188]. This relative abundance of mtDNA imparts a correspondingly higher likelihood of its recovery from samples where DNA is generally highly degraded, and is one of the principle reasons why mtDNA testing is used.

A second characteristic that lends great power to mtDNA analysis as a forensic tool is that mtDNA is maternally inherited. A primary reason for this may be simply numerical: sperm heads contain only a few copies of mitochondrial DNA compared to the many thousands of copies in the ovum [35]. However, there also appears to be a specific recognition mechanism that can eliminate even the few paternal mitochondria that may be introduced into the ovum. For example, when mitochondria from human sperm cells were introduced into somatic culture cells devoid of mtDNA, 10–20% of the cells contained functioning sperm mitochondria immediately following introduction, while only a very small fraction of the cells ($1/10^5$) survived more than 48 hours [139]. However, when mitochondria from somatic cells were introduced into cultured cells, there was rapid replacement of endogenous mtDNA [115]. This points to the existence of mechanisms that specifically eliminate sperm-derived mitochondria, but not mitochondria derived from somatic cells.

Additional information regarding mtDNA inheritance comes from studies of other mammals. Further supporting a mechanism for the elimination of sperm-derived mitochondria, Sutovsky et al. [211] found that sperm mitochondria were undetectable by the late four-cell stage of *in vitro*-fertilized bovine embryos. However, Gyllensten et al. [65] studied mouse species whose mtDNA are readily distinguishable, and performed repeated hybridizations between *M. spretus* females to C57BL males. Using highly sensitive PCR techniques designed to specifically amplify paternal DNA, low levels of paternal sequences (0.01–0.1% relative to maternal contributions) were detected. Similar results were obtained by Kaneda et al. [113], also using interspecific mouse hybrids, with paternal mtDNA detectable throughout development and birth. However, when Kaneda et al. performed crosses using mice of the same species, paternal mtDNA was detected only through the early pronucleus stage. These latter results suggest that the mechanism for elimination of paternally derived mitochondria (or mtDNA) is based on

a self:non-self recognition system that does not function with mitochondria derived from a different species. This would account for the persistence of paternal mtDNA that Gyllensten et al. [65] observed with inter-specific mouse hybridization.

While the mechanism for elimination of paternal mtDNA is not fully elucidated, nor known to be absolute in terms of extremely low level persistence, it is clear that from the practical standpoint of mtDNA forensic testing, mtDNA behaves as maternally inherited. Parsons et al. [173] report comparison of mtDNA sequences of 69 father:child pairs, and in no case was any trace of the paternal sequence detected by direct sequencing of PCR-amplified mtDNA (this being the same methodology in use for forensic testing). While many instances have now been observed of mixtures of more than a single mtDNA type within an individual (a condition known as heteroplasmy, reviewed in more detail below), in no case have the mixtures involved more than a small number of base positions (i.e., usually 1 or 2). If such heteroplasmic mixtures were the result of paternal inheritance, the expectation would be for mixtures at many more positions, as there are on average eight differences in control region sequences between two randomly selected Caucasian individuals.

C. Heteroplasmy

1. Overview

During development, mtDNA molecules are replicated independently of one another, are not strictly tied to mitotic or meiotic cell division, and are thought to be essentially non-recombining [reviewed in 100]. Further, mitochondrial DNA replication is associated with a much higher error rate than is nuclear DNA (e.g., [23,122]). These factors create the possibility that the population of mtDNA molecules found within an individual could be diverse, with many variants replicating and segregating independently. Indeed, if all the mtDNA variants present in a mother were passed to her offspring, one would predict that over population genetic history the accumulation of variants would be so extensive as to invalidate the concept of a mitochondrial DNA "type" that would be useful for identity testing. However, we know from vast experience in human population genetic studies and forensic testing that this is not the case: i.e., when PCR-amplified mtDNA is sequenced, individuals typically harbor a single mitochondrial DNA type that is distinguishable from that of other maternal lineages.

While mechanisms clearly exist that restrict the level of mtDNA variation that is passed between generations (genetic bottlenecks in mtDNA transmission, discussed below), it is now known that mixtures of two or more subpopulations of mitochondrial DNA — a condition known as heteroplasmy — can occur within individuals. Heteroplasmy has the potential to both complicate and strengthen forensic identity testing, and must be taken into account. Fortunately, the frequency of heteroplasmy in the population (when analyzed by DNA sequence analysis) is relatively low (i.e., 2–8% of the population), so it is not an issue in a majority of cases, and our current knowledge is sufficient to appropriately deal with it in a great majority of cases where it does occur.

The expectation of readily detectable sequence variation within the mtDNA population of single individuals motivated early studies that compared the sequences of multiple fragments of cloned mtDNA from single individuals, as well as from different tissues within an individual [156,157], and from retinal tissue suspected of being highly susceptible to DNA damage [17]. The then-surprising result of these studies was that remarkably little variation was detected. As a result, an expectation was established that variation within individuals was minimal, and that individuals could be considered to be essentially homoplasmic. This assumption of homoplasmy was not violated by a vast body of work performed in the field of human population genetics and molecular evolution. Many population studies of human CR sequence variation were performed using PCR amplification and direct sequencing (reviewed in a subsequent section), using both manual and automated fluorescence sequencing techniques. These studies were conducted under the assumption of homoplasmy, and none reported other than a single mtDNA type within thousands of individuals. The explanation for this is that heteroplasmy can be difficult to distinguish from "background" in sequencing data, where apparent signal from alternative nucleotides can be present at variable levels due to artifacts of the sequencing chemistry and/or detection methods. However, newer and "cleaner" sequencing technologies are now being used, and as a result, it is not particularly difficult to distinguish heteroplasmy from background. However, the data from population genetic studies were not analyzed with heteroplasmy in mind, and the positions where heteroplasmy occurred in those samples were undoubtedly either called as the predominant nucleotide, or denoted as isolated "ambiguities" of unknown cause.

The first documented instance of point mutation heteroplasmy within the human CR occurred in a forensic case involving the skeletal remains of Tsar Nicholas II [58,104]. Since then, heightened scrutiny has resulted in multiple reports of point mutation heteroplasmy in the CR, indicating that it is not an extremely rare occurrence [13,14,38,98,101,161,172,208,219,220,235]. In addition, it has been known for some time that length heteroplasmy is common in two polycytosine stretches (one in HV1 and

one in HV2 [12,140]). Within HV1, most people have a T at position 16,189 that interrupts a run of C's on each side. When a C is substituted for the T at 16,189 (a condition present in ~20% of the general population), this creates an unbroken run of C's that is apparently replicated with low fidelity by the mitochondrial DNA polymerase system. The result is a population of molecules within an individual that differ in the length of the C-stretch, producing an abrupt crash in the quality of direct sequencing data due to template molecules that are out-of-register with one another. When the HV1 C-stretch is present, it is virtually impossible to characterize the population of length variants by direct sequencing, so forensic labs often do not call a "type" for the HV1 C-stretch. The correct sequence outside of the C-stretch itself, however, can be determined by sequencing both strands and using alternative sequencing primers. Length heteroplasmy also occurs in the HV2 C-stretch (positions 303–315), but in that case there is often an identifiable predominant length variant, even when heteroplasmy is present [172].

Although there have been numerous reports of CR heteroplasmy, the sample size is too small to know with precision the frequency with which heteroplasmy occurs at most nucleotide positions. In addition to DNA sequence analysis, there are other, non-sequencing methods that can unambiguously indicate the presence of heteroplasmy: e.g., denaturing high performance liquid chromatography (DHPLC) [221], denaturing gradient gel electrophoresis (DGGE) [77,205], and single-strand conformation polymorphism (SSCP) [219] analysis. In work performed in the authors' laboratory [219], DGGE was used in a population study to assess the frequency and distribution of heteroplasmy within HV1. Blood samples from 253 individuals of various races (predominantly Caucasians) were assayed for heteroplasmy, with the result that 35 (~14%) showed some level of detectable heteroplasmy; two individuals were heteroplasmic at two positions. This high percentage is due in part to the high sensitivity of the DGGE method in detecting mixed sequences. Mixing experiments showed that DGGE is often capable of detecting a mixture even when the minority component is 1% of the total, and in this study was deemed capable of detecting mixtures at all positions in the region that had a minority component of 5% or greater. This is substantially more sensitive that direct sequencing, where the sensitivity threshold is ~10–15% at best [236]. In fact, only three of the individuals in the study had heteroplasmy that was detectable by direct sequencing. This underscores that when asking the question, "Is this individual heteroplasmic?" one also has to ask, "At what level of sensitivity?" Certainly, given the untold trillions of mtDNA molecules in an individual, we are all heteroplasmic at some trace level, but only occasionally at levels that are of functional significance to forensic identity testing.

A significant point that comes out of the DGGE heteroplasmy study cited above, as well as the published [13,14,38,98,101,161,172,208,219,220,235] and unpublished data (AFDIL) on heteroplasmy detected by sequencing, is that not all positions in the CR have an equal propensity for heteroplasmy. The DGGE study indicated two predominant hot spots for heteroplasmy, with 12 independent instances of heteroplasmy (34% of the total) occurring at position 16,093, and 6 instances occurring at 16,129. Four other sites had two instances of heteroplasmy, while heteroplasmy at all other sites occurred only once. A general correlation can be seen between the sites where heteroplasmy occurs and the sites where differences between individuals accumulate over evolutionary time in the human population [79,106,226]. This is to be expected as both heteroplasmy and population polymorphism have the same root cause, i.e., the substitution of nucleotides at particular base positions on mtDNA molecules. It has been well established that the rate of base substitutions between individuals in the population is quite heterogeneous among sites [79,226]. The sites where multiple instances of heteroplasmy have been observed, particularly the two HV1 hot spots 16,093 and 16,129, are all sites that are polymorphic for population variation as well. The same is true of the HV2 C-stretch, where individuals commonly show length heteroplasmy [140], and where the population is roughly evenly divided between having 7 Cs or 8 or more C's [26]. This general information on the distribution of sites where heteroplasmy occurs can guide interpretation when heteroplasmy becomes an issue in forensic casework (discussed below).

2. Intergenerational Substitution

Given that heteroplasmy occurs at quite tractable levels at only a site or two within individuals, it is not a factor that *a priori* confuses forensic identity testing. Indeed, it represents an additional level of variation that can increase the power of mtDNA testing [104]. If a reference sample and an unknown sample match each other by sharing a constellation of unusual polymorphisms, the co-occurrence of heteroplasmy at a particular site would provide additional evidence restricting the pool of potential donors, and increasing the odds that the reference is, in fact, the source of the unknown sample. However, this concept must be framed with reference to the manner with which heteroplasmic mtDNA variants segregate within an individual during development, and how they are passed between generations. As it turns out, these processes involve one or more genetic "bottlenecks." This creates the potential for different hetero-

plasmic ratios among different tissues of the same individual, as well as differences between maternal relatives. In extreme cases, the difference could be large enough to appear as a homoplasmic nucleotide substitution. While this feature of heteroplasmy can cause complications in forensic testing our current understanding of mtDNA biology is such that the complications can be dealt with appropriately in a great majority of situations that will be encountered.

The bottleneck theory was proposed in early studies to account for the rapid segregation of variant mtDNA sequences observed within pedigrees of Holstein cows [80,81,117]. Frequent transitions to heteroplasmy were observed, as well as shifts between generations from one apparent homoplasmic state to another, sometimes without a heteroplasmic intermediate. In the bottleneck theory, it is proposed that the number of mtDNA genomes is reduced to a relatively small number during some stage of oogenesis, or germ cell development. A subset of the mtDNA population is transmitted through the bottleneck and this founder population (or a portion thereof) is subsequently replicated to produce ~100,000 mtDNA copies in the mature ovum. The founder population, as a small subsample, may differ significantly in the proportions of variants from the original mtDNA pool, giving rise to differential segregation between different meioses.

It is quite apparent that an mtDNA bottleneck occurs during human oogenesis. Bendall et al. [13] studied 180 twin pairs and discovered CR heteroplasmy (detectable by direct sequencing) in four instances. Analysis of additional family members in these heteroplasmic lineages showed differential segregation of varying magnitude between generations. Applying a population drift model to the data, Bendall et al. estimated the size of the meiotic (intergenerational) bottleneck to range from 2 to >100 mtDNA molecules for particular mother:child transmissions, with an overall best estimate of 3–20 molecules. This is similar to the value Marchington et al. [141] calculate from studying individual ova from a single heteroplasmic woman. However, Howell et al. [97] present data from a particular heteroplasmy segregating in a lineage where the bottleneck does not always appear to be greatly restricted — as also appeared to be the case with certain transmissions in Bendall et al. Therefore, it may be that the size of the effective bottleneck can vary stochastically from transmission to transmission, or in response to other genetic factors [45].

The data of Parsons et al. [173] are also consistent with a narrow bottleneck. In that study, 327 mother:child events were surveyed (directly or indirectly) within 134 independent mtDNA lineages. Within these, there were ten instances of a substitution of one mtDNA type for another; i.e., generally one nucleotide substitution for another. In one lineage, two such apparent "substitutions" were observed. Additional such "intergenerational substitutions" were reported by Parsons et al. [174], indicating a higher rate than is predicted by evolutionary estimates of the mtDNA substitution rate. This higher rate was also reported by Howell et al. [98], but others studying single or restricted numbers of lineages failed to observe such a high rate of substitutions [107,202].

The explanation of why near-term empirical studies, when taken together, indicate a higher intergenerational substitution rate than predicted by evolutionary/phylogenetic studies has not been resolved. It has been suggested that it is because near-term substitutions occur at hot spots [107,137,171] providing an inflated estimate of the rate when applied to all CR sites. However, this would only be true if phylogenetic analyses had substantially underestimated the rate of reversion mutations at these hot spots [79,99,173]. Whatever the explanation, it is the empirical studies that are directly relevant to forensic identity testing, and it is apparent that maternal relatives can differ from one another in CR sequence with an appreciable frequency. Interpretation in mtDNA identity testing must take this into account when maternal relatives are compared, to avoid the potential for false exclusion. Further, when apparent differences are noted between individuals suspected of being maternal references, interpretation can be greatly abetted by careful analysis of both samples for low levels of heteroplasmy that would indicate the segregation of mtDNA variants within a lineage.

3. Heteroplasmic Segregation: Disease Studies and Development

A large body of work has investigated the segregation of disease-causing mtDNA mutations within individuals and between generations. While a detailed review is beyond the scope of this paper, we will provide a brief overview, focusing on how this information relates to issues of forensic significance. Common syndromes associated with tRNA or protein-coding point mutations are myoclonic epilepsy and ragged-red fiber (MERRF), mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms (MELAS), maternally inherited myopathy and cardiomyopathy (MCM), Leber's hereditary optic neuropathy (LHON), and neurogenic muscle weakness, ataxia, and retinitis pigmentosa (NARP). Common diseases associated with large-scale deletions of mtDNA are Kearns-Sayre syndrome and Pearson's syndrome. These diseases are sometimes due to homoplasmy for the mutant condition, but frequently are associated with varying levels of heteroplasmy. Pathological conditions become evident when the energy-producing capacity of mitochon-

dria falls below threshold levels in particular tissues with critical energy requirements, and the severity of symptomology is often correlated with the proportion of mutant mtDNAs (extensive reference lists relevant to this paragraph are in [9,227,228]). While there are abundant reports of heteroplasmy associated with mtDNA disease, these contribute little to our knowledge of the frequency of heteroplasmy in the CR because the pathogenic nature of these heteroplasmies ensures their detection even if the causal events are rare.

Many studies have investigated the segregation of disease-related mutations within pedigrees (e.g., [21,91,125,126,135,154,181,201,213,225]). Taken together, the information from intergenerational transmissions strongly reinforces the existence of a genetic bottleneck between generations, with different offspring of the same mother demonstrating widely different heteroplasmic proportions. While this appears to be the general rule, some disease studies have suggested that the bottleneck is not always tightly restricted [132,177,225]; as was the case in some instances with CR heteroplasmies [13], and neutral mutations in the coding region [97]. It is possible that these studies were influenced by the pedigrees evaluated and/or the limited number of pedigrees evaluated.

It is important to note that mutations involved in mitochondrial disease have functional consequences to cells and tissues, and the segregation and accumulation of these pathogenic mutations may be different than for non-coding CR variants. This is suggested by a general (but not absolute) trend for the progressive increase of deleterious mutants across generations. While this trend is clearly superimposed with a high degree of sporadic segregation consistent with a bottleneck (discussed above), it appears that in some cases there is a general replicative advantage for the mutant condition. Large-scale deletion mutations may have a replicative advantage due to their smaller size (e.g., [227]), but a progressive accumulation across generations occurs with some deleterious point mutations as well [201,246]. The cause of this is not understood, but may be related to the functional effects of the mutations. Therefore, one should use caution in using disease studies as a guide for how segregation may operate on forensically relevant CR variants.

Many studies of disease-associated heteroplasmies indicate an uneven distribution of mutant mtDNAs in various tissues of afflicted individuals [25,71,96,103,116, 125,130,135,144,167,213,245,248]. This phenomenon might reflect differential segregation of mtDNA variants during early development, but in general the studies do not discriminate this possibility from that of variable post-mitotic selection on, or accumulation of, mutant mtDNAs. The widespread occurrence of heteroplasmic mutations throughout multiple tissues strongly suggests that the individuals originally inherited heteroplasmy from their mother. In addition, a number of studies indicate that a relatively uniform distribution of heteroplasmic mixtures occurs throughout the establishment of the primary germ layers, as evidenced by a uniform distribution among fetal tissues, but subsequent differential accumulation or segregation can occur to result in different heteroplasmic mixtures among tissues in adults [138,144,219].

A common finding in inter-tissue disease studies is that levels of the mutant type are highest in tissues with high-energy demands such as brain, skeletal muscle, liver, and heart. The high-energy demands of particular cells may produce a more oxidative environment that leads to higher mutagenesis [71]. However, others have suggested the possibility that the functional nature of the disease mutations play a role in their differential accumulation in a tissue-specific fashion [144]. Once again, the dynamics of segregation and accumulation of deleterious mutations may be quite different than those of CR sequence variants.

There are a number of studies involving CR heteroplasmies in humans indicating that the primary bottleneck in mtDNA segregation occurs prior to maturation of the oocyte, and hence, prior to fertilization and cell division during embryogenesis [16,140,141]. This is also the picture that emerges from a detailed study of the segregation of neutral, non-functional variants in artificially heteroplasmic mice, where the bottleneck was demonstrated to occur during the expansion of the population of oogonia from primordial germs cells during oogenesis [110]. The implications of this are that the state of heteroplasmy or homoplasmy present within an oocyte determines to a large extent that a similar condition will be found throughout an individual after development [141,182]. However, there is evidence that post-zygotic differential segregation of mtDNA variants does occur to some extent in the CR. This was seen in a DGGE survey of CR heteroplasmy in 21 adults [220]. Examination of various tissues (blood, bone, brain, hair, liver, muscle) from these individuals indicated that the levels of heteroplasmy varied in some instances among the tested tissues. Some tissues were detectably heteroplasmic, while others were not. However, in almost all cases in this study, the heteroplasmy was quite low level (undetectable by sequencing) and the variability between tissues reflected rather minor fluctuations, some near the limit of detectability.

The issue of between tissue variation is relevant to forensic testing because it creates the potential for reference and questioned samples from the same individual to differ in heteroplasmic ratios, or in a more extreme case in an apparently homoplasmic manner. The data reviewed

above indicate that in general this is a rather restricted problem, by virtue of the primary bottleneck in mtDNA segregation occurring prior to embryogenesis. However, the data indicate that drift of heteroplasmic ratios can occur during development, and forensic scientists must be prepared to deal with this in both procedure and interpretation. Particularly, recent studies (primarily in the forensic context) have indicated the existence of another developmental bottleneck that exists in the generation of individual hairs from a single individual [14,208,235]. For example, in a particular woman, multiple extracts and amplifications from blood and buccal swabs gave consistent levels of CR heteroplasmy (the levels were consistent within each, but differed slightly between the blood and buccal swabs) [235]. However, extracts from different hairs of the same individual showed greatly skewed ratios; in five hairs thymine predominated at position 16,355, in three hairs cytosine predominated, and in two there was a roughly even mix.

Testing of shed hair recovered at crime scenes is a major application for mtDNA analysis, because hairs lacking roots generally possess insufficient amounts of nuclear DNA for testing. Therefore, the scenario of hair sequences differing from those of blood, buccal, or other hair samples of the same individual will be encountered. In such cases it is important to carefully scrutinize the reference samples (blood, buccal swab) for low-level heteroplasmy that would be expected to segregate differentially among individual hairs. Also of importance is to sample multiple hairs from the suspected individual, to test whether differential segregation of heteroplasmic variants is occurring in a manner consistent with the type observed in the questioned sample. Interpretation of mtDNA evidence in such a case will be discussed in a subsequent section.

D. Population Genetic Variation

Studies of the variation in human mitochondrial DNA have been pivotal in our understanding of the evolutionary origin and population genetic history of modern *Homo sapiens* (e.g., [31,43,79,150,175,186,224]). The maternal inheritance, lack of recombination, and high evolutionary mutation rate of mtDNA are properties that aid tremendously in the ability to draw inferences relating to genetic history. The control region is the most rapidly evolving segment of mtDNA, and sequence data from the CR has been characterized from a great many populations worldwide. A recent review lists 4079 HV1 and 969 HV2 sequences compiled from 38 separate primary publications [73]. It was this wealth of sequence information that revealed the high variability of control region sequences, and prompted the development of forensic analysis of the CR.

The large number of published sequences for the most part do not correspond exactly to the boundaries of the regions that are analyzed in forensic testing [153]. Additionally, it is difficult to evaluate the accuracy of the sequence data when compiled from such a wide variety of sources. For these reasons, forensic laboratories have assembled databases of their own, comprised of complete HV1 and HV2 sequences that have been determined in accordance with specified quality assurance criteria (e.g., [172,238]). The current database in use by the Armed Forces DNA Identification Laboratory and the FBI (and shared by other forensic laboratories worldwide) contains HV1 and HV2 sequences from 1657 individuals. The size of this database at the time of this writing, broken into component ethnic groups, is as follows: African American, 149; Afro-Caribbean 116; African, 115; U. S. Caucasian, 604; English Caucasian, 102; Austrian Caucasian, 101; French Caucasian, 109; American Hispanic, 99; Asian American, 58; Japanese, 162; Korean, 42 [26].

Sequence databases are our best source of information regarding the power of mtDNA for identity testing. Given that it is expected that maternal relatives will match each other (barring mutation or "substitution"), it is clear that mtDNA is not a unique identifier. Moreover, mtDNA lineages extend back into time to various depths of antiquity, with the result that one's mtDNA lineage generally encompasses many more individuals than would be included in one's known relatives. In addition to the potential for matching other individuals of the same lineage, matches will also occur to other mtDNA lineages that have mutated independently to consist of the same mtDNA sequence. The population genetic processes that govern the longevity and spread of mtDNA lineages, and the rate of change to and from a particular sequence, are complex. Therefore, we presently must turn to an empirical evaluation of the net result of these processes: the frequency with which particular sequences are detected within various populations.

Within the database of 604 U. S. Caucasian HV1 and HV2 sequences, there are 451 distinct mtDNA types. Among these, there are a small number of relatively common types, and a larger number of rare types. The most common mtDNA type occurs in 26 individuals, or 4.3% of database. There are 188 other mtDNA types that occur in more than one individual (ranging from one type that occurs 15 times to 31 types that occur twice), but 390 occur in only a single individual. Using Mitosearch, a program developed by the Federal Bureau of Investigation [26] one can perform a pairwise comparison of all sequences within the database. For U. S. Caucasians, this

results in 669 instances of a match out of 182,106 separate comparisons, for an empirically determined probability of 0.0037 for a random match. Thus, we can predict that two randomly chosen individuals from this population will match once out of ~270 times. However, given a known sequence of a particular type, the chance that another random individual will match depends greatly on the relative rarity of that particular type.

All other population databases show a pattern similar to that of the U. S. Caucasians in having many rarer types and fewer common types. However, it is known that different populations can have different levels of mtDNA diversity, with pools of mtDNA types that differ significantly [147–149,207]. For example, the African American database ($n = 149$) has a most common type occurring in 2.7% of individuals in the database, 27 additional types that occur more than once, and 118 types that occur only once. In the African American database, the average number of nucleotide differences between individuals is 14, higher than in Caucasians; the average number of nucleotide differences between individuals in the U.S. Caucasian database is eight. This is consistent with a higher mtDNA diversity known to occur in African populations, reflecting a greater evolutionary age of African versus non-African (and particularly, Caucasian) sequences. Moreover, if we compare all the sequences in the U.S. Caucasian database with those in the African American database, there is only a single instance of a sequence match, compared to pairwise match frequencies more than 200 times higher in each of the databases considered separately. This indicates the extent of the sequence differences between the databases and illustrates the necessity of maintaining separate databases, at least for the major population/racial groups.

It is apparent that, in general, the strength of mtDNA evidence is limited by the size of the current databases. Since there are many rare haplotypes, the "apparent frequency" of the unique haplotypes (1/n, where n is database size) is, in a great majority of cases, an overestimate of the true frequency. This can be illustrated by simulating the growth of a database. For the purposes of this example, the forensic databases of European-derived sequences (U.S. Caucasian, French Caucasian, Austrian Caucasian, English Caucasian) were pooled [there is evidence suggesting that there is little or no substructure within mtDNA of various native European populations, 149]. From this larger pool, 100 sequences were sampled at random to produce a smaller database, then sequences were added to this by further random addition in increments of 100. This resulted in a growing set of databases of size 100, 200, 300, 400, 500, 600, 700, and 800.

For these incremental databases, **Figure 2** plots the total number of mtDNA types and the total number of unique mtDNA types. Both of these values are still increasing steadily at the maximum database size of 800, even within Caucasians where diversity is lower than in some other major groups. We can see that when new sequences are added, a great majority of sequences that were unique in a smaller database remain unique, with a concomitant decrease in their "apparent frequency." For example, when 100 new sequences were added to the database of size 700, 66 new apparently unique sequences were added, while only six sequences were added that matched a sequence that was previously unique. Thus, in increasing the database size, the apparent frequency of six previously unique sequences increased, while the apparent frequency of 445 previously (and still) unique sequences decreased. It is not known when most of the mtDNA types in the population will have been sampled, but the limit has not yet been closely approached. Large-scale efforts that are underway to increase database size can be expected to increase the power of mtDNA testing for the many rare mtDNA types.

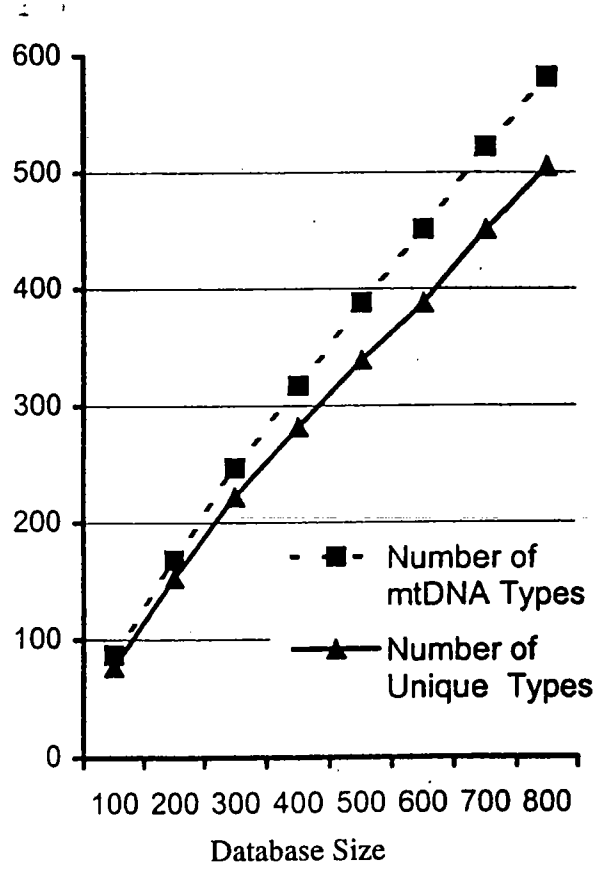

Figure 2. Graph showing effect of database size on the total number of mtDNA types within the database and the number of mtDNA types that are unique in the database. Databases increasing in size by increments of 100 were obtained by random subsampling of a large database of Caucasian sequences, as described in text.

II. INTERPRETATION OF MtDNA DATA IN THE FORENSIC CONTEXT

A. Match Significance

When one applies mtDNA sequence typing as a forensic test, it is clear that the purpose is to answer two related questions: (a) through sequence comparison, could an individual be excluded as a possible source of a questioned sample?, and (b) if not, what does this inclusion suggest regarding the chance that the individual is, in fact, the source of the questioned sample? If the answer to the first question is yes, the exclusion is an outcome of obvious significance (for which mtDNA testing has a clear and proven utility). However, if an individual is included as a potential source, one is immediately faced with the second question. Unless the discriminatory potential of a test can be objectively evaluated, an inclusion could mean anything. It is therefore incumbent on the forensic scientist to determine a means to evaluate and communicate the significance of an mtDNA inclusion or "match." In this section we will explore various alternative and/or complementary methods for communicating this significance, without intending to suggest that these definitively or exclusively constitute the manner in which mtDNA evidence should be presented in a court of law. In fact, mtDNA evidence has been introduced in only a small number of cases in the United States, and it seems likely that the presentation of mtDNA statistics in court will continue to evolve. What remains clear nonetheless is that there are many scientifically accurate statements that can be made concerning mtDNA match significance. As long as appropriately conservative approaches are taken, with clear statements of the premises and potential limitations of various statistics, forensic scientists have at their disposal a number of ways to aid investigators, lawyers, judges, and juries in a proper understanding of what the mtDNA evidence means.

In general, the significance of an mtDNA sequence match is dependent on the case in question, and the mtDNA type involved. In only a restricted set of circumstances can mtDNA matching be considered definitive evidence of identification. This can occur when the population of potential sources is "closed," that is, narrowed down to an identified set of individuals. An example would be an aircraft incident where recovered remains simply need to be associated with any of the crewmembers known to be aboard. If a reference sample (often a maternal relative in such a case) for a crewmember can be distinguished from those of all other crewmembers, a match to a set of recovered remains can be considered a positive identification. Here, the identification is not based on DNA statistics, but on a process of elimination that is, in turn, dependent on the surety that the sample could only have come from a known set of individuals. The identification of Michael J. Blassie, whose skeletal remains were interred for 14 years in Arlington Cemetery as the Vietnam Unknown Soldier, was an instance where mtDNA was used to distinguish among a restricted set of reference lineages.

More commonly, mtDNA testing involves cases where the potential source population is not closed or completely characterized. In these cases, mtDNA evidence alone cannot be considered a definitive identification, and we have to address the relative significance, or exclusion potential, of a match. Such assessments are based on some indication of the relative frequency of the mtDNA type in the population, or, relatedly, the chance that random individuals in the population will also match that mtDNA type. As a general indication of the power of mtDNA testing, one can present the frequency with which sequences in a database match one another when all pairwise comparison are made. For example, in a database of 604 U.S. Caucasian sequences, there are 669 instances of matching in 182,106 comparisons: on average, the chance that two randomly chosen individuals in this population would match is one in 272. While such average values can provide the framework for thinking about mtDNA match probabilities, this is not the best way to evaluate match significance in any particular case. A more accurate and useful guide for interpretation is based on the relative rarity of the particular sequence in question. The relative rarity (or relatedly, the frequency) of a sequence determines how likely it would be for a randomly selected individual to match that particular mtDNA sequence by chance.

The simplest factual statement that can be made concerning the relative rarity of an mtDNA sequence type is to report the number of times that the sequence has been observed in various databases. While this has been referred to as the "counting method," the manner in which it has been presented in mtDNA court cases differs from the "counting method" that the original NRC report defined as a means for estimating the frequency of a profile from the number of observations in the database [163]. In mtDNA cases, the number of observations in the database has been reported without any further interpretation precisely because the counting method is a poor method for estimating the frequency of rare mtDNA types. To date, the counting method (or "pseudo-counting method" as it has also been called) is the only method that has been reported in the United States. Despite some arguments, this presentation has been deemed acceptable by the courts in most instances. However, the counting method presentation of mtDNA evidence was in one instance found inadmissible under the Frye standard. The grounds