Received 20 January 2015
Received in revised form 4 March 2015
Accepted 6 March 2015
Available online 11 April 2015
Peer Reviewed
2157118X.5.1.R1

Sequence, Haplotype, and Ancestry: Using the Mitochondrial DNA Hypervariable Region to Predict Forensic "Race"

Alexander F Christensen, Ph.D. [*, a]

[a] Joint POW/MIA Accounting Command—Central Identification Laboratory, 310 Worchester Avenue, Hickam AFB, HI 96853.

[*] Corresponding Author: Alexander F. Christensen, PhD, D-ABFA, is a forensic anthropologist at the Joint POW/MIA Accounting Command Central Identification Laboratory (JPAC/CIL). He earned his BA in archaeological studies and classics at Yale College in 1991 and PhD in anthropology at Vanderbilt University in 1998. Before joining the JPAC/CIL, Dr. Christensen taught at Rutgers University–Camden and Augusta State University. His primary research interest is the structure of prehistoric and historical Mesoamerican populations, and he has used osteological, archaeological, linguistic, and ethnohistoric data to investigate this question. He is a member of the American Academy of Forensic Sciences, the American Anthropological Association, and the American Association of Physical Anthropologists, and has published articles in the Journal of Forensic Sciences, Forensic Science International, Human Biology, World Archaeology, Journal of Archaeological Method and Theory, Journal of Human Evolution, International Journal of Osteoarchaeology, and other professional journals. Dr. Christensen is a diplomate of the American Board of Forensic Anthropology (#79).

Introduction

In forensic casework, DNA is primarily used for the individuation of evidence, or the identification of remains. This is particularly true for autosomal DNA (auDNA), which can be used in a positive identification, but is generally the case for Y-chromosome (yDNA) or mitochondrial DNA (mtDNA) as well (Edson and Christensen this volume). While variation in yDNA and mtDNA actually only places an individual within one or more particular groups of paternally or maternally related individuals, it is used in conjunction with other circumstantial evidence (such as a passenger manifest) to determine individual identity. But what about when individuation is not possible, either because references are not available for comparison or because no victim names are known? How well can DNA be used to assign an individual to a population group?

Several recent studies have evaluated the degree to which genetic data correlates with ancestry and therefore its predictive value for "race" and/or "ethnicity" (e.g., Lao et al. 2010; Royal et al. 2010). This chapter focuses on mtDNA, as it is the type most frequently sequenced from skeletal remains, due to its high copy number and good preservation. Because mtDNA is a nonrecombining, uniparentally inherited genome, correlating it with individual ancestry raises a different, if related, set of issues from those posed by autosomal DNA.

A given mtDNA haplotype can be assigned to a particular ancestral group using either a phylogenetic approach (e.g., Lao et al. 2010) or a computational approach (e.g., Egeland et al. 2004). The first classifies each sequence into a broad haplogroup and follows a simple chart or map to determine with which continental population that haplogroup is associated. This approach is often used in studies of population history and admixture (e.g., Stefflova at al. 2009). The second is a statistical model, independent of phylogeny: How similar is this sequence to each different sequence in the reference populations and to which population does it show the greatest affinity? One such study achieved close to 90% accuracy in prediction of what was referred to as "coarse ethnic group"—that is, Caucasian, African, Asian, and Hispanic (Lee, Măndoiu, and Nelson 2011).

This chapter takes a step back from these approaches and breaks the problem down into three independent questions that must be addressed in any forensic case:

- How accurately can a given sequence be assigned to a specific haplogroup?
- How tightly defined is the ancestral population with which that haplogroup is associated?
- What is the correlation between that ancestral population and the circumstances of the particular forensic case?

Sometimes, these questions will allow a given case to be placed within a forensically useful ancestral category with a fair degree of certainty; sometimes they will not. This chapter will consider the three questions in order, and then specifically examine

mtDNA haplogroups A, B, C, and D, which are of particular interest for the casework of the Joint POW/MIA Accounting Command—Central Identification Laboratory (JPAC-CIL). Each of these haplogroups can be found in both Native Americans and East Asians, and the distinction between these two groups is often of forensic relevance in CIL cases. Finally, two case studies will be considered.

Illustrative examples are drawn from CIL casework, which consists primarily of skeletal remains believed to be associated with US casualties from World War II, the Korean War, and the Vietnam War. For CIL military casework, mtDNA analyses are conducted by the Armed Forces DNA Identification Laboratory (AFDIL), and the frequency of each mtDNA sequence is reported against its casework population database (CPD). The CPD supplements the original Scientific Working Group for DNA Methods reference database of 4,839 individuals (Monson et al. 2002) with additional sequences from AFDIL's own analyses for a total of 10,428, sorted into forensically relevant subsamples. For identification purposes, AFDIL compares evidence sequences to conflict-specific databases of family reference samples (FRSs)—that is, samples taken from maternal relatives of casualties that provide the mtDNA sequences that those casualties can be expected to have had.

Throughout these examples, matches to casualty references will be listed under the racial classifications used in those service members' original files. Individuals who might now be classed as Hispanic, for instance, might be listed as Mexican, Puerto Rican, or White; in the latter case, Hispanic surnames are noted. The terms used in the records vary over time and between individuals, and include some that are inappropriate at best to modern ears, including several casualties from the Korean War classified as "Mongolian," whose actual ancestors came from China, Japan, and Korea, and others classified as "Malayan" who appear to be of Filipino, Chamorro, or Native Hawaiian ancestry.

Hypervariable Region Sequence and Phylogeny

Because mtDNA mutates without recombination, haplotypes can all be placed within a single phylogeny based upon their evolutionary relationships (Figure 15.1). Clades within this phylogeny are called haplogroups. Early studies of mtDNA variation at the population level (e.g., Torroni et al. 1993) defined haplogroups based on restriction fragment length polymorphisms (RFLPs) and assigned them letter designations. As work progressed, it was discovered that some of these lettered haplogroups nested within others, and an increasing use was made of sequence data rather than RFLPs (Macaulay et al. 1999; Ingman and Gyllensten 2001). In the current system, the letter designations are preserved for historical continuity, but new clades are designated with a series of numbers and letters within the original haplogroups, such as B4a1a1a or D4b1a2a1b (van Oven and Kayser 2009). The current phylogeny is based upon extensive sequencing of entire mtDNA genomes, and its broad outlines are unlikely to change (Behar et al. 2012).

While haplogroups have been the focus of population genetics studies, forensic analyses have individuation as their focus. As a result, forensic laboratories generally sequence the hypervariable region (HVR) and, occasionally, the broader control region (CR). The high mutation rate across the HVR makes it ideal for individuation, but also means that a particular mutation may well have occurred multiple times and therefore be of limited phylogenetic significance (Behar et al. 2007). Individuals belonging to different haplogroups and exhibiting different polymorphisms elsewhere in their mitochondrial genome may bear the same sequence in the HVR. Some polymorphisms in the HVR are more stable and are associated with particular haplogroups, but most haplogroups are defined on the basis of coding region polymorphisms that are not sequenced in standard forensic casework. In some cases, forensic individuation may require analysis of such single-nucleotide polymorphisms (SNPs) in the CR to differentiate between otherwise identical HVR sequences (e.g., Asari et al. 2007; Just et al. 2009).

Table 1 lists the basic HVR motifs associated with each lettered haplogroup. To determine in more detail what haplogroup a particular HVR sequence might belong to, several references are available. The current master phylogeny is available at http://www.phylotree.org (van Oven and Kayser 2009; references in this chapter are to build 14, dated April 5, 2012) and http://www.mtdnacommunity.org (Behar et al. 2012). Individual polymorphisms may be searched against these charts to locate where a given sequence might fall. mtDNAmanager (http://mtmanager.yonsei.ac.kr; Lee et al. 2008) provides a web-based application to compare sequence data to published sequences and predict haplogroup assignment, as well as a reference chart of HVR mutation motifs for each haplogroup (http://mtmanager.yonsei.ac.kr/help/MutationMotifs.pdf; references in this chapter are to the version dated October 2–3, 2011). Haplogroup assignment can be double-checked by comparing the sequence data to the set of complete mtDNA haplotypes made available by

mtDNA Community (12,813 as of the April 6, 2012, release).

As a general rule, sequences that exhibit few polymorphisms in comparison to the Cambridge reference sequence (CRS) will be from European twigs of the phylogeny, in H, V, and U. This is because the CRS belongs to haplogroup H2a2a1. In particular, any sequence that does not exhibit a 73G almost certainly belongs either to R0 (including HV, H, and V) or L0, as those are the primary positions on the phylogeny where the original 73A has mutated to G. (The same G–A transition also appears in minor branches L3h1a1 and C4c2, but they are very rare.) Further, the absence of 73G combined with the presence of 72C places a sequence securely in HV0 or its daughter, V. Sequences with 263A are more common, occurring in multiple twigs of the tree, but without much phylogenetic significance. As a result, throughout this chapter, all sequences are listed with an assumed 73G-263G unless stated otherwise.

Macrohaplogroup R is marked by the loss of 16223T. While this mutation does occur in other branches, the presence of 16223T is a good indicator that the sequence falls within L, M, or N (although Behar et al. 2007 report that 2.5% of all pre-R genomes exhibit 16223C, while 1.1% of R genomes have mutated back to 16223T). Within these macrohaplogroups, some of the branches, particularly those found in Europeans, East Asians, and Native Americans, are quite well defined in the phylogeny. Others are still very poorly delineated, as they are found in parts of the world where the number of samples to date is very low in comparison to the amount of phylogenetic variation. Recent work in South and Southeast Asia, for instance, has defined a large number of new haplogroups within M, some (but not all) of which have distinctive HVR profiles (see Chandrasekar et al. 2009; Peng et al. 2010). A large number of sequences from African populations are available (Behar et al. 2008), but it is likely that much diversity remains unsampled there as well.

Throughout the process of haplogroup assignment, care should be taken to avoid overspecificity, especially if the evidence sequence lacks rare polymorphisms. Sometimes, a given HVR sequence might be indicative only of a general area of the tree, or even of multiple, distantly separated branches. As an example, consider the sequence 16223T-16278T-

Table 1. Mitochondrial DNA Haplogroups

Haplogroup	Defining HVR Polymorphisms[a]	Simplified Macroclade Structure
L0	16129A, 16187T, 16189C, 16223T, 16230G, 16311C, 146C, 152C, 195C, 247A, 263A	
L1	16187T, 16189C, 16223T, 16278A, 16311C, 152C, 182T, 185T, 195C, 247A	>L1′2′3′4′5′6
L5	16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16278A, 16311C, 182T, 195C, 247A	>L2′3′4′5′6 > L1′2′3′4′5′6
L2	16223T, 16278A, 16390A, 146C, 150T, 152C, 182T	>L2′3′4′6 > L2′3′4′5′6 > L′2′3′4′5′6
L6	16048A, 16223T, 16224C, 16278A, 16311C, 146C, 152C, 182T, 185C	>L3′4′6 > L2′3′4′6 > L2′3′4′5′6 > L1′2′3′4′5′6
L4	16223T, 16311C, 16362C, 195C	>L3′4 > L3′4′6 > L2′3′4′6 > L2′3′4′5′6 > L1′2′3′4′5′6
L3	16223T	>L3′4>L3′4′6 > L2′3′4′6 > L2′3′4′5′6 > L1′2′3′4′5′6
M	16223T, 489C (control region)	>L3
C	16223T, 16298C, 16327T, 249del, 489C	>CZ > M8 > M
Z	16185T, 16223T, 16260T, 16298C, 249del, 489C	>CZ > M8 > M
E	16223T, 16362C, 16390A, 489C	>M9 > M
G	16223T, 16362C, 489C	>M12′G > M
Q	16129A, 16223T, 16241G, 16311C, 16362C, 489C	>M29′Q > M
D	16223T, 16362C, 489C	>M80′D > M
N	16223T	>L3
I	16129A, 16223T, 16391A, 199C, 204C, 250C	>N1a′c′d′e′I > N1 > N1′5 > N
W	16223T, 16292T, 189G, 195C, 204C, 207A	>N2 > N
Y	16126C, 16231C	>N9 > N
A	16223T, 16290T, 16319A, 235G	>N
O	16213A, 16223T	>N
S	16223T	>N
X	16189C, 16223T, 16278T	>N
R	None	>N
HV	73A	>R0 > R
V	16298C, 72C, 73A	>HV0a > HV0 > HV > R0 > R
H	73A	>HV > R0 > R
J	16069T, 16126C, 295T	>JT > R2′JT > R
T	16126C, 16294T	>JT > R2′JT > R
F	16304C, 249del	>R9c > R9 > R
B	16183C, 16189C	>R
P	None	>R
U	None	>R
K	16224C, 16311C	>U8b > U8 > U2′3′4′7′8′9 > U > R

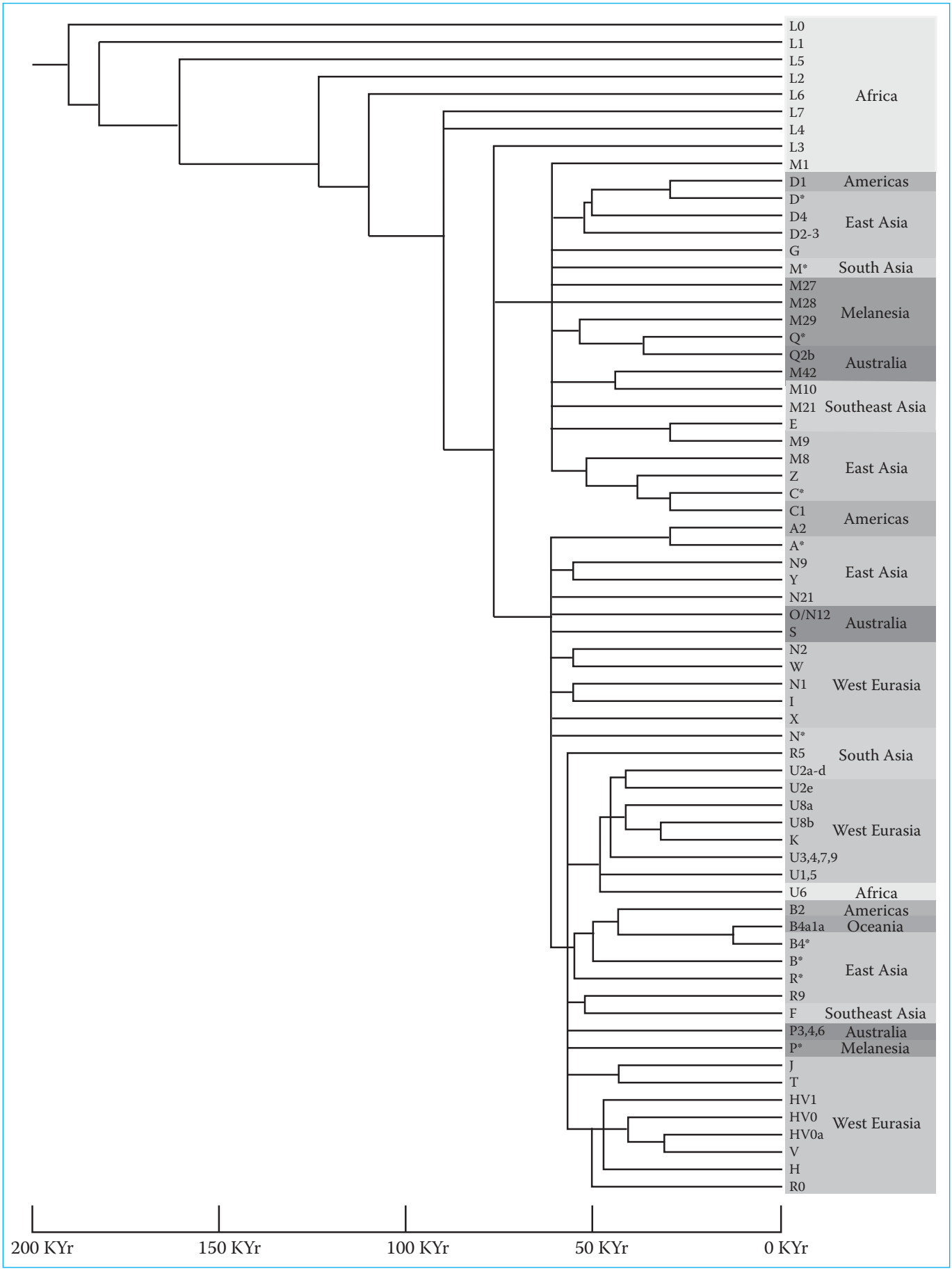

Fig.1 Simplified phylogeography of human mtDNA. Dates are approximate. An asterisk after haplogroup designations indicates "members of this haplogroup not specified on other branches."

16362C (here and elsewhere in this chapter, C-stretch insertions are ignored). In CIL casework, this sequence has been found in multiple sets of remains from both North and South Korea, as well as World War II contexts on Pacific islands. The mtDNAmanager reference list shows this as the profile for haplogroup D4g1, which is consistent with Northeast Asian ancestry, such as Korean or Japanese. In fact, this sequence is found in over 2% of modern Japanese and almost 1% of Koreans (Sekiguchi et al. 2008; Jin et al. 2006). Does this mean that all of those individuals actually belong to haplogroup D4g1? Not necessarily. With the addition of 16189C, this sequence is also consistent with G2a; since 16189 has a fairly high mutation rate, it is possible that some haplotypes within this haplogroup would show the same HVR sequence. Given that D and G are related clades within macrohaplogroup M and show similar distributions, it might be safer to state that the sequence probably falls within D or G, and might specifically belong to D4g1.

However, this same sequence is found in 11/10,428 individuals in AFDIL's CPD: two Koreans, four Turkmen, one Asian American, three African Americans, and one Hispanic. It also matches family references for two casualties from the Korean War: one African American and one Korean American. Are the African American and Hispanic matches due to Asian gene flow? Probably not. Instead, it turns out that the same HVR profile can be found in some haplotypes within L3b1a (Soares et al. 2012). So without additional sequencing of other regions of the mtDNA genome, all we can say about this particular HVR sequence is that it belongs either to D/G or to L3b1a. Because one of the mutations that define macrohaplogroup M (including D and G) is 489C, sequencing of the full CR would differentiate D/G and L3b1 sequences.

Phylogeny and Ancestry

Once a haplotype is located within the phylogeny, what can we say about its ancestry? On a local level, population genetic studies have shown a weak correlation between mtDNA and geography because of the movement of wives between communities (e.g., Besaggio et al. 2007). On a broader, continental level, it does have a high correlation with geography, as haplogroups branched off from each other with the spread of modern humans across the world (Fig.1; Maca-Meyer et al. 2001). The very diverse macrohaplogroup L, aside from its descendant clades M and N, is restricted in distribution to African populations, which exhibit great diversity and time depth (Behar et al. 2008). Macrohaplogroups M and N branched off from L3 around the time of our ancestors' departure from Africa and subsequently diversified across Eurasia. A few branches, such as M1 and U6, indicate back-migrations into Africa after this date (Olivieri et al. 2006). Within M, several clades have received their own letter designations (C, D, E, G, Q, and Z), while dozens of others have simply been numbered. Some of these clades are both very distinctive and very restricted in distribution (e.g., Q in Papua and Australia), while others spread across much of Asia and even (in the case of C and D) the Americas. Within N, there are two broad clades: R and N*, with the latter also including the haplogroups A, I, O, S, W, X, and Y. R is in turn divided into R* (including B, F, J, P, and T), R0 (including HV, H, and V), and U (including K). The R0 and U clades are restricted to western Eurasia, while R* and N* include some branches restricted to Asia (and in the case of A and B, the Americas) and others found in Europe (such as T).

Beyond these broad patterns, however, there is a fair amount of variation. Within any one designated haplogroup, smaller branches may have much more restricted distributions. Some of these distributions may be widely disjunctive, as well, as a result of historical migrations (Table 2). Thus, while haplogroups L1, L2, and L3 are restricted to sub-Saharan Africans and their descendant populations, specific branches within them have been documented from Slavic groups and likely date to prehistoric small-scale population movements (Malyarchuk, Derenko, et al. 2008).

Ancestry and Forensic Relevance

Sequence 1 exhibits an extensive series of polymorphisms that allow it to be securely placed in the phylogeny: 16223T-204C-199C, in that order; place it in the N1a′d′e′I branch, with 152C-16248T-16355T narrowing it to N1a. N1a1a is marked by 16147A and then 16320T. Finally, 207A and the reversal of a previous 16172C mutation mark N1a1a2. This subclade has only been reported from Europe and has actually been sequenced from Neolithic farmers from Germany (Palanichamy et al. 2010). Sequence 2 exhibits the same 16223T-204C-199C, followed by the 250C-16391A that marks haplogroup I (a subclade of N1, and similarly distributed in Europe). Its placement within I1a1 is determined by the successive mutations 16311C-16172C-203A. In the Pacific islands, European maternal ancestry indicates that the remains are likely to come from a US or Australian casualty who can eventually be identified by the CIL.

Sequences 3 and 4 both exhibit the 16223T-16362C that characterizes haplogroups D, G, and some other, minor branches of M. G3 is marked

by 16274A, and G3a2 by 143A-152C-16189C-16265C, placing sequence 3 securely. Sequence 4 is more difficult, however, as it also exhibits the 16390A that characterizes haplogroup E. However, no sequence within E in the reference phylogeny exhibits 16260T, which characterizes G2b2b. By contrast, one of the reference sequences for G2b2b exhibits 16093C-16189C-16390A as private mutations (that is, differences present in just one sequence and therefore not used in the current phylogeny) and thus matches sequence 4 perfectly. Haplogroup G is generally restricted to Northeast Asia, and both sequences have been reported from Japanese individuals (Sekiguchi et al. 2008; Nohira, Maruyama, and Minaguchi 2010). On a Pacific island, such a sequence is most consistent with the remains belonging to an Imperial Japanese service member.

Sequence 5 is phylogenetically the most distinctive of all. Haplogroup Q is characterized by 16129A-16223T-16241G-16311C, and 16144C-16148T-16265C-16343G-89C-92A-146C marks Q1. This is an ancient branch (perhaps as much as 50,000 years old) only found in the indigenous population of Papua New Guinea and Island Melanesia (Friedlaender et al. 2005). Such a sequence indicates that the remains are almost certainly of indigenous Melanesian origin.

Table 2. Selection of mtDNA Subhaplogroups with Disjunctive Distributions

Haplogroup	Subhaplogroup	N/12813[a]	n[b]	Primary Region
L1		200		Africa
	L1b1a8[c]		0	Europe
L2		232		Africa
	L2a1k		2	Eastern Europe
M		1386		Eurasia
	M1		80	North Africa
C		423		Northeast Asia
	C1b-d		211	Americas
	C4c		4	Americas
	C5c1		11	Eastern Europe
D				East Asia
	D1		50	Americas
	D2a		48	Americas
	D3		7	Americas
	D4b1a2a1b		4	Eastern Europe
	D4e1c		2	Americas
	D4e4b		2	Eastern Europe
	D4h3a		45	Americas
	D5a3a		4	Northeast Europe
A		426		Northeast Asia
	A2		277	Americas
X		161		Circumpolar
	X2a		24	Americas
	X2g		1	Americas
B		452		East Asia
	B2		122	Americas
	B4a1a1a		67	Polynesia
U		1208		Europe
	U2a,b,c,d		28	South Asia
	U6		119	North Africa, Mediterranean
Z		62		Northeast Asia, South Asia
	Z1a1a		22	Scandinavia

How Far Back Is That Asian Ancestry? Distinguishing Asian and Native American Sequences

Traditional triracial categorizations in physical anthropology lumped Native American and Asian populations together as "Mongoloid" (e.g., Brues 1990). In fact, the founding gene pool of Native Americans is a reduced subset of northeast Asian variation, as the Americas were colonized by small groups crossing the Bering Strait who were largely isolated over subsequent millennia. This is particularly clear in the mitochondrial genome. Initial studies of mtDNA phylogeography defined haplogroups A, B, C, and D as the four founding Native American types and recognized that they also occurred in Asia (Torroni et al. 1993). As the phylogenetic details were refined, it became clear that the American branches A2, B2, C1, and D1 were distinct from the Asian branches of the same haplogroups, and that haplogroup X should also be added as a founder. More recently, full genome sequencing has revealed that extant Native American mitochondrial lineages trace back to at least 15

Table 3. Cases from Pacific Islands

Individual	1	2	3	4	5
Recovery location	Papua New Guinea	Philippines	Philippines	Papua New Guinea	Papua New Guinea
Sequence	16086C, 16147A, 16223T, 16248T, 16320T, 16355T, 152C, 199C, 204C, 207A	16172C, 16223T, 16311C, 16362C, 16391A, 189G, 199C, 203A, 204C, 250C	16189C, 16223T, 16265C, 16274A, 16362C, 143A, 152C	16093C, 16189C, 16223T, 16260T, 16362C, 16390A, 195C	16144C, 16148T, 16188T, 16223T, 16241G, 16265C, 16311C, 16343G, 89C, 92A, 146C, 208C
Haplogroup	N1a1a2	I1a1	G3a2	G2b2b	Q1
Ancestry	European	European	East Asian	East Asian	Melanesian
N/10428[a]	0	0	0	0	0
Probable origin	United States	United States	Japanese	Japanese	Papuan

distinct branches, each with Asian cousins: A2*, A2a, A2b, B2, C1b, C1c, C1d*, C1d1, C4c, D1, D2a, D3, D4h3a, X2a, and X2g (Perego et al. 2010). Furthermore, at least one additional haplogroup, an otherwise undocumented branch of M, has been found in a 5,000-year-old burial from British Columbia (Malhi et al. 2007), indicating that it is possible or even likely that other as of yet unknown lineages are, or were, present in the New World, if at very low frequencies. Nonetheless, the vast majority of Native American sequences belong to one of the four originally identified branches.

In CIL casework, recognizing whether a given sequence is of Asian or Native American origin is frequently probative, as the majority of the remains received by the CIL come from Asian or Pacific contexts. While there are US casualties of Asian ancestry missing in Korea and Southeast Asia, AFDIL can compare unknown sequences from those areas to their family references. A sequence of Asian origin that does not match a known Asian American casualty almost certainly represents a non-US individual, and the remains are therefore not likely to be identified by the CIL. Any sequence of Native American origin found in the region, by contrast, has a very high likelihood of representing a US casualty who is (hopefully) identifiable.

Haplogroup A

Haplogroup A is a subclade of macrohaplogroup M and is distinguished by the HVR profile 16223T-16290T-16319A-235G. It has a restricted distribution in northern Asia and is found throughout the Americas (Achilli et al. 2008; Derenko et al. 2007). Within haplogroup A, there are three successive splits marked by HVR polymorphisms: 152C differentiates A3, A4, A7, A9, and A11 from A5, A8, and A10; 16362C differentiates A4; and within A4, 16111T-146C-152T-153G differentiates A2. Within A2, A2a and A2b are found in both northeast Asia and the Americas, while all the other clades, most of which share also 64T, are restricted to the New World. For forensic purposes, it is safe to consider any haplogroup A sequence exhibiting 16111T and/or 64T to be almost certainly of Native American ancestry, however distant.

Five cases provide examples from haplogroup A (Table 4). Individual 1 was recovered from the island of Tarawa, where thousands of US and Japanese service members were buried in 1943. He exhibits the A4 profile without any of the distinctive A2 polymorphisms and also has a 200G, which marks the A4c′d branch. Since A4d also has 151T, while A4c has no further HVR mutations, he most likely falls within the latter clade. While no exact match to this sequence could be located in the literature, A4c is a North Asian, and particularly Siberian, clade (Derenko et al. 2007), and a similar sequence has been reported from Korea (Jin et al. 2006). It thus appears more consistent with Japanese ancestry than American, although the recently discovered New World branches of haplogroups C and D (Table 2) show that other related but long isolated branches may still turn up.

Individuals 2, 3, and 4 were recovered from North Korea. The first bears a 16187T that is characteristic of A5a; this sequence is common in both Korean and Japanese populations. The second and third both exhibit the markers of A2 (except for 152T, which has mutated back to 152C in multiple branches within A2). They cannot be placed more precisely than that, which is not surprising, as A2 contains a broad range of variation, consistent with a small founding population spreading across a wide area and growing exponentially (Kumar et al. 2011). This haplogroup is indicative of Native American ancestry, which is consistent with their FRS. Individual 5, recovered from Vietnam, does not match any sequences in the CPD or FRS database. However, despite the lack of database matches, this sequence can also be securely placed in haplogroup A2. All three sequences almost certainly belong to US service members.

Haplogroup B

Haplogroup B is a subclade of macrohaplogroup R and is characterized

Table 4. Haplogroup A Examples

Individual	1	2	3	4	5
Recovery location	Tarawa	North Korea	North Korea	North Korea	Vietnam
Sequence	16093C, 16223T, 16290T, 16319A, 16362C, 152C, 200G, 235G	16187T, 16223T, 16290T, 16319A, 235G	16111T, 16223T, 16290T, 16319A, 16362C, 16378T, 64T, 146C, 153G	16051G, 16111T, 16223T, 16249C, 16290T, 16319A, 16362C, 146C, 153G, 235G	16111T, 16223T, 16290T, 16319A, 16362C, 146C, 153G, 159C, 235G
Haplogroup	A4c	A5a	A2	A2	A2
Ancestry	Northeast Asian	Northeast Asian	Native American	Native American	Native American
N/10428[a]	0	1 Kazakh	0	1 Caucasian	0
FRS matches[a]	None	None	1 White (Hispanic surname)	1 White, 1 American Indian	None

in the HVR by (a) the change at 16223 to the CRS that typifies R, and (b) 16183C-16189C. Given that both 16183 and 16189 exhibit high mutation rates (the former site is not even considered in the formal phylogeny), it can be difficult to distinguish B if no other mutations are present. Branches are found throughout East Asia, the Pacific, and the Americas (Achilli et al. 2008; Derenko et al. 2012; Li et al. 2007; Melton et al. 1995). Within B, the primary division is between B4, marked by 16217C, and B5, marked by 16140C and four other mutations outside the HVR. Within B4, 16261T divides B4a, g, h, and i from B4b, d, and e (the last a minor branch marked by a slew of other polymorphisms as well). A simple sequence of 16189C-16217C could belong to either B4b or d. The difficulty here is that B4b includes haplogroup B2, the haplogroup present in the New World, which is differentiated from B4b by five mutations outside the HVR. Therefore, HVR data alone cannot always discriminate between Asian and American Bs. Fortunately, B4b1, the largest clade within B4b, is marked by 16136C, usually with multiple other polymorphisms as well, which does allow its separation from B2.

Table 5 provides some examples of the difficulties raised by haplogroup B in evidence cases. Individual 1 was recovered from North Korea and exhibits the basic B4 profile plus 16390A. Unfortunately, this polymorphism is not significant within the current B4 phylogeny. The sequence does match four Hispanics in the CPD, as well as two casualties who both appear likely to fall within that same modern category. It therefore appears likely that these remains will be identifiable as those of a US casualty. Similarly, individual 2 has the basic B4 profile plus 150T. This sequence matches 10 Navajo (out of the 146 in the CPD), but it is certainly possible that it might also match indigenous Southeast Asians, given how poorly that region is represented in databases and the fact that B4c is present in the region (Peng et al. 2010). Without additional information, either from elsewhere in the genome or from the archaeological context, these remains cannot be securely assigned to either American or Asian ancestry.

Individual 3, by contrast, presents a set of polymorphisms distinctive to B2g1. The transversion of C to G at 114 is particularly significant, as most mutations are transitions (between either C and T or A and G). This is a haplotype that appears to be of native Mesoamerican origin. The remains were recovered from a World War II aircraft crash site in Germany, and in that context, US ancestry is the most plausible explanation.

Individual 4 exhibits the HVR motif of B5a, including another distinctive transversion from C to A at 16266. This is a haplogroup that is quite common in mainland Southeast Asia (Peng et al. 2010). Given that these remains were recovered in Vietnam, they almost certainly represent an indigenous individual.

The last three sequences present a different conundrum. They exhibit a set of polymorphisms distinctive to B4a1a: 16189C-16217C-16261T-146C. This branch appears to have arisen in early Austronesian populations spreading out across islands of Southeast Asia (Melton et al. 1995). Individual 5 was recovered from North Korea, but the same sequence has also been found in remains recovered from a World War II context in Papua New Guinea. Individuals 6 and 7 come from Saipan and exhibit the 16247G that marks B4a1a1a. The combination of 1627C-16247G-16261T has been labeled the "Polynesian motif" (Redd et al. 1995), and it is one of a handful of lineages carried by the Austronesian populations that expanded into Polynesia (Kayser 2010).

At first glance, the population database matches of these three sequences do not match what might

Table 5. Haplogroup B Examples

Individual	1	2	3	4	5	6	7
Recovery location	North Korea	Cambodia	Germany	Vietnam	North Korea	Saipan	Saipan
Sequence	16182C, 16183C, 16189C, 16217C, 16390A	16183C, 16189C, 16217C, 150T	16183C, 16189C, 16217C, 16298C, 114G, 146C	16140C, 16183C, 16189C, 16266A, 210G	16182C, 16183C, 16189C, 16217C, 16261T, 146C	16182C, 16183C, 16189C, 16217C, 16247G, 16261T, 146C	16182C, 16183C, 16189C, 16217C, 16247G, 16261T, 146C, 151T
Haplogroup	B4/B2	B4/B2	B2g1	B5a	B4a1a	B4a1a1a	B4a1a1a
Ancestry	East Asian/Native American	East Asian/Native American	Native American	Southeast Asian	Island Southeast Asia	Polynesian	Polynesian
N/10428	4 Hispanic	10 Navajo	5 Hispanic	0	2 Chinese, 1 African American	4 Asian, 4 African American, 1 Hispanic, 1 other	2 Asians, 1 Hispanic, 1 other
FRS matches	1 White (Hispanic surname), 1 Mexican	None	1 White (Hispanic surname)	None	2 Filipino	1 Black, 2 Hawaiian, 1 Chinese Hawaiian, 1 Asian American	1 White, 3 Hawaiian, 1 Chinese, 1 "Mongolian"

be expected from their phylogenetic position. The "Asian" samples are all either Taiwan Chinese, who could be of indigenous Austronesian ancestry, or Asian Americans, a category that often includes Pacific Islanders as well. But how are distinctly Pacific sequences appearing within the African American and Hispanic populations? Some insight is provided by the list of matching casualties from within the FRS database, for whom some individual biographical data are available. Individual 5 matches two Filipino American casualties, as expected given the distribution of B4a1a. The group of casualties matching individual 7, despite the variety of racial categories, are all of Native Hawaiian maternal ancestry. Given the population history of Hawaii, it is perfectly possible that the CPD Hispanic who matches this sequence is in fact also of Native Hawaiian maternal descent. Individual 6 also matches three casualties of Native Hawaiian maternal ancestry and one of uncertain Asian ancestry. However, the last match is to an African American whose maternal lineage can be traced to the late nineteenth century in Tennessee and is very unlikely to have any Hawaiian ancestry.

So what accounts for the "Polynesian motif" appearing in an African American? In fact, several African American data sets include sequences belonging to B4a1a, as well as the Southeast Asian haplogroup F3b (Allard et al. 2005; Diegoli et al. 2009; Stefflova et al. 2009). These sequences are usually attributed to recent non-African admixtures. However, it is more likely that they trace back to Madagascar, one source of the Atlantic slave trade (Lee et al. 2009; see Razafindrazaka et al. 2010 for a recent analysis of Malagasy mtDNA).

So what is the significance of a B4a1a or B4a1a1a sequence in CIL casework? It depends upon the recovery area. For remains recovered in Saipan or elsewhere in the Pacific, it is possible that they belong to a Filipino American or, less likely, African American service member. It is much more likely, however, that they belong to an indigenous individual. By contrast, remains recovered on the mainland of Asia or in another combat theater probably belong to a US service member of uncertain ancestry.

Haplogroup C

Haplogroup C is a branch of haplogroup M distinguished by 16223T-16298C-249del (which it shares with haplogroup Z), and 16327T. C1, the primary Native American branch, is further distinguished by 16325C-290del-291del. C and its sister Z are largely restricted to northeast Asia, although some branches reach further afield (Table 2; Derenko et al. 2010; Ebenesersdóttir et al. 2011). Five cases from the sister haplogroups serve as examples (Table 6).

Individuals 1 and 2 exhibit sequences that can be securely placed within the phylogeny. The first has the standard C1 motif plus 16086C-16189C-16278T-143A, which together mark C1b4. The second exhibits the C1 motif minus 16223T (an example of the parallel occurrence of the T–C transition at this locus). The 215G matches the profile expected for C1c1. Both of these sequences are present in CPD Hispanics, indicating likely Native American ancestry, which in turn, when recovered from North Korea and Papua New Guinea, indicates that the remains most likely belong to US service members.

However, we can actually draw even more specific conclusions from a careful examination of the literature. In the FRS database, sequence 1 matches three Puerto Ricans, and Martínez-Cruzado (2010) has identified it as one of the founding Native American lineages in Puerto Rico. As many US troops of Puerto Rican ancestry did serve in Korea, it is likely that these remains represent one of them. Sequence 2 is consistent with Mexican American examples (Kumar et al. 2011), suggesting that that origin is more likely.

Individual 3 presents a very unusual sequence. It bears the C1 motif, but the other five polymorphisms are absent from all published examples. Given the number of polymorphisms,

Table 6. Haplogroup C Examples

Individual	1	2	3	4	5	6
Recovery location	North Korea	Papua New Guinea	Saipan	North Korea	North Korea	Hawaii
Sequence	16086C, 16183C, 16189C, 16223T, 16278T, 16298C, 16325C, 16327T, 143A, 249del, 290-291del	16298C, 16325C, 16327T, 215G, 249del, 290-291del	16131C, 16164G, 16207C, 16223T, 16224A, 16234T, 16298C, 16325C, 16327T, 249del, 290-291del	16093C, 16223T, 16234T, 16288C, 16298C, 16327T, 249del	16223T, 16298C, 16327T, 16357C, 204Y, 207A, 249del	16185T, 16223T, 16260T, 16298C, 16301Y, 16362C, 151T, 249del
Haplogroup	C1b4	C1c1	C1	C5c1	C4c1?	Z
Ancestry	Native American	Native American	Northeast Asian/ Native American?	East European	Asian/Native American?	East Asian
N/10428	7 Hispanic	2 Hispanic	0	1 Caucasian	0	0
FRS matches	3 Puerto Rican	1 White (Hispanic surname)	None	1 White	1 American Indian	None

this sequence appears to represent an old lineage within C1 that has not been sampled in published studies. It is certainly possible that it is indeed Native American; it is also possible that it is Northeast Asian, belonging either to C1a with a subsequent loss of the 16356C that marks that haplogroup or to a yet undescribed clade. As these remains are from a Pacific island, the former option would be consistent with US ancestry, the latter with Japanese ancestry.

Individual 4 is placed within C5 by the 16288C and then C5c1 by 16093C-16234T. Given C5's distribution in Northeast Asia and the recovery location of these remains in North Korea, it would be logical to assume a Korean or Chinese ancestry. However, the sequence also happens to match one White US service member. Further investigation reveals that C5c1 is indeed an Eastern European branch, first defined from Polish individuals, and owes its Asian maternal ancestry to a settler from the steppes millennia ago (Derenko et al. 2010). Without the FRS match, it would have been easy to set these remains to the side as not from the United States and therefore not identifiable by the CIL.

Individual 5 provides a similar case. The sequence lacks the diagnostic markers of C1 and C5, which leaves C4 and C7 as phylogenetic possibilities, since neither of these clades is defined by HVR mutations. The 16357C indicates a possible placement within C4a, best documented from South Asia (Chandrasekhar et al. 2009). However, the sequence also matches a casualty of Native American ancestry. Given this match, it is possible that the sequence actually falls within the recently defined C4c, which has been documented from multiple Native Americans and represents an additional founding lineage in the hemisphere (Hooshiar Kashani et al. 2012). Several of the sequences that fall within this haplogroup lack any distinctive HVR mutations.

Individual 6 was recovered on the island of Oahu and is included here to show the difference between haplogroups C and Z. Although recovered from a context where a prehistoric Native Hawaiian interment was a possibility, the mtDNA sequence obtained is inconsistent with that origin, since haplogroup B4a1a1a would be expected in that case. Haplogroup Z is quite rare and therefore poorly documented, so no identical sequences were encountered in the literature. However, very similar examples have been reported from Guangdong, China (Chen et al. 2008; Wang et al. 2010). The remains likely represent a nineteenth- or twentieth-century descendant of East Asian immigrants; depending upon the recovery context, they could be from either a historical cemetery or a more recent missing person case.

Haplogroup D

Haplogroup D is one of the largest clades within haplogroup M and is found across Asia and the Americas (Achilli et al. 2008; Derenko et al. 2010). Its basic HVR profile is 16223T-16362C, which unfortunately is also the root profile for multiple other haplogroups, such as M6, M9 (which includes E), and G. As illustrated previously, some HVR profiles within D also match even more distantly related phylogenetic twigs. As a result, it is often not possible to locate a related sequence precisely within the phylogeny. In some cases the sequence may be useful even without a more precise cladistic position. For instance, the basic D profile is found in approximately 4% of both the Japanese and Korean populations (Jin et al. 2006; Sekiguchi et al. 2008). In a forensic situation where there is a high prior probability of encountering remains of either of those ancestries, such as a recovery on a Pacific island, this profile may be sufficient to determine the disposition of remains without a more exact phylogeny.

Table 7 provides six casework examples of haplogroup D sequences. The first three were all recovered in North Korea. Individuals 1 and 2 each exhibit the 16325C that marks D1. Like A2, D1 has a fairly broad and shallow phylogeny, and it can be difficult to place sequences more precisely. Individual 1's 16274A-16368C

Table 7. Haplogroup D Examples

Individual	1	2	3	4	5	6
Recovery location	North Korea	North Korea	North Korea	Vietnam	Papua New Guinea	Papua New Guinea
Sequence	16223T, 16274A, 16325C, 16362C, 16368C	16147T, 16223T, 16325C, 150T, 152C, 185A, 489C	16093C, 16209C, 16223T, 16274A, 16295T, 16325C, 16362C, 189G, 195C, 203A, 204C, 228A, 298T, 325T	16129A, 16223T, 16362C, 152C	16129A, 16223T, 16243C, 16362C, 151T, 152C	16164G, 16172C, 16182C, 16183C, 16189C, 16223T, 16266T, 16362C, 150T
Haplogroup	D1i	D1?	D4g2a	D4a	D4a	D5a2a1
Ancestry	Native American	Native American?	East Asian	East Asian	East Asian	East Asian
N/10428	4 Hispanic	0	0	10 Asian	0	8 Asian
FRS matches	2 White (1 also called Mexican)	1 White (with Spanish surname)	None	None	None	None

does not match any specific branch; however, one sequence that is placed in D1i exhibits the same motif. Individual 2 lacks the 16362C of D, but because his full CR was sequenced, we know that he has the 489C that marks macrohaplogroup M, and within this, the presence of 16325C and match to a casualty of apparent Hispanic ancestry suggest D1. The other polymorphisms, 16147T-150T-152C-185A, do not match any documented sequence. Individual 3 has a 16274A-298T that places him within D4g2a; the other polymorphisms match a published Japanese sequence from this clade (Tanaka et al. 2004). In North Korea, the first two individuals are most likely US service members, while the third is either Korean or Chinese.

Individuals 4 and 5 both show the 16129A-152C that marks D4a, in the latter case with an additional 16243C-151T not shown in the phylogeny. This haplogroup has a different significance in the two different recovery locations. In Southeast Asia, D is present, albeit at low frequencies (Peng et al. 2010). In Melanesia, it is absent in the indigenous populations (Friedlaender et al. 2007). Therefore, the former case is most likely an indigenous Vietnamese individual, while the latter is most likely an Imperial Japanese service member. Finally, individual 6 bears a profile distinct to haplogroup D5a2a1. This haplotype was first described from Siberia, but is also present in China (Starikovskaya et al. 2005; Yu et al. 2010). In Vietnam, it most likely indicates Sino-Vietnamese ancestry.

Skeletal Ancestry versus Mitochondrial Lineage: A Korean War Case

From 1996 through 2005, recovery teams led by CIL anthropologists excavated numerous sites in North Korea. From one of these sites, a foxhole on a battlefield from November 1950, a team recovered the skeletons of three US service members with fragments of their uniforms but without any identification media. The remains were well preserved, allowing a detailed anthropological analysis as well as mtDNA testing (Table 8). Determinations of ancestry were based upon cranial morphology and discriminant function analyses conducted in FORDISC 3.0 (Jantz and Ousley 2005), which yielded consistent results. Subsequently, two of the three individuals have been identified, allowing verification of the ancestry determinations.

For individual 1, skeletal ancestry appeared consistent with the mtDNA haplogroup and database matches, indicating that this service member was most likely classified as Negro in 1950, although the individual has not yet been identified to confirm this. For individuals 2 and 3, the results are less clear. In the former case, the skeleton appears to be of African ancestry, while the mtDNA haplotype belongs to the European haplogroup U, and the database matches are predominantly Hispanic. In the latter case, the skeleton appears to be of European ancestry, while the mtDNA haplotype is very strongly African.

Sequence 2 presents an example of how the details of the mtDNA phylogeny do not necessarily fit our "racial" categories particularly well. Malyarchuk et al. (2010) have argued that U5 (defined by 16192T-16270T) is the oldest European haplogroup, and that it evolved over the last 30,000 years in southwestern Europe and the adjoining southern rim of the Mediterranean. U5b1b1b in particular (defined by 16320T) is documented from both US Hispanics and Senegalese. Martínez-Cruzado et al. (2005) found this haplotype (which they labeled U5b2) in 9/800 Puerto Ricans and considered it to be of West African origin; however, it also appears plausible that it reached US Hispanics from a North African or Iberian source. In this particular case, it is likely of West African origin, as the remains were identified as those of the Black casualty whose family reference matched.

Sequence 3 belongs to one of the deeper branches on the African mitochondrial phylogeny, L1, which is found at low frequencies in both West and East Africa (Gonder et al. 2007). While this exact sequence is not found in the literature, the L1b1a clade is documented in African Americans and West Africans and also in individuals from elsewhere in Africa, Portugal, Cyprus, Jordan, and Israel (Behar et al. 2008). This lineage almost certainly entered the US population through

Table 8. Skeletal Group from North Korea

Individual	1	2	3
Sequence	16223T, 16278T, 16294T, 16309G, 146C, 152C, 195C	16189 C, 16192T, 16270T, 16320T, 150T	16114T, 16126C, 16187T, 16189C, 16215C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 152C, 182T, 185T, 195C, 247A
Haplogroup	L2a1a	U5b1b1b	L1b1a1
Ancestry	African	African?	African
N/10428	33 African-American, 1 Hispanic	8 Hispanics	0
FRS matches	9 Black, 1 White	2 Puerto Rican Whites, 1 Black	1 White
Skeletal ancestry	Negroid	Negroid	Caucasoid
Casualty race	Unknown	Black	White

the slave trade, but the one casualty whose reference matched was White. In fact, the remains were subsequently identified as that individual, who may or may not have been aware that he had a distant African ancestor.

In these latter two cases, if ancestry had been assigned to the remains based purely upon the mitochondrial haplogroup, it would have been incorrect. The skeletal determination, however, matched the bureaucratic one.

Ours or Theirs? A South Asian Case Study

During World War II, 594 US and Allied aircraft carrying 1,659 personnel were lost flying "over the hump," carrying supplies from British India to China. These crash sites have proven to be some of the most difficult in the world to locate and recover, given their remoteness. In several cases, hikers and other private citizens have picked up remains at remote sites and subsequently turned them over to JPAC. One such individual provided a selection of bones that he said he had obtained from a US crash site in the mountains along the India–Burma border. However, without archaeological provenience, the CIL was unable to verify whether the remains actually did represent US casualties. Samples were taken from all of the bones for mtDNA analysis, and AFDIL reported five sequences (Table 9). The first three sequences matched multiple individuals of European ancestry in the CPD and therefore appear fully consistent with the alleged provenience. Sequence 1 exhibits the 16356C-195C that defines U4, with a 146C that marks U4b1a3, sequence 2 the basic K profile of 16224C-16311C, and sequence 3 the 16126C that marks JT with the additional 16069T-295T of J. The absence of 150T-152C, which defines J2, indicates that sequence 3 belongs to J1.

Looking at the two sequences that do not match any references in the CPD, it is easy to hypothesize that sequence 4, with only three polymorphisms, is likely to be European, especially as 16304C is a fairly common variant. However, the combination of 199C and 16304C, along with the 16223T that excludes the sequence from macrohaplogroup R, is in fact very distinctive and places this sequence within haplogroup M35b (Chandrasekhar 2009). M35 itself has no HVR polymorphisms, but 199C marks the M35a′b branch, and 16304C in turn marks a clade of M35b1 and M35b2. Like many South Asian haplogroups, it is clear that M35 is very diverse, particularly in comparison to the limited data so far available on it, and is present in individuals from throughout the subcontinent. M35b2 is in fact best documented from Eastern European populations, where it appears to have been carried by the ancestors of the Vlax Roma (Salihović et al. 2011; Mendizabal et al. 2011). It is now present in a high percentage of Vlax, but also in the non-Romany population of Slovakia (Malyarchuk, Perkova, et al. 2008).

However, all European M35b sequences known to date exhibit the string of polymorphisms 16129A-16230G-16233G-16344T in addition to the 199C-16304C, suggesting that sequence 4 is more likely to be from an indigenous South Asian than an Eastern European. Furthermore, the two sequences in the whole genome database that match this HVR profile exactly are from a Thai "sea gypsy" and a dental patient from Andhra Pradesh, India (Behar et al. 2012), while other related examples come from the Tharu of Nepal (Fornarino et al. 2009). Given this distribution, it appears most likely that the bone in question does not belong to a US casualty and was instead improperly associated with the other remains, either by the individual who turned the remains over to US authorities or by a third party who possessed them previously.

Conclusions

Because race is a cultural category, albeit one based upon perceived biological differentiation, it will never perfectly correspond with determinations of ancestry based upon any form of biological variation. Uniparental lineage markers such as YDNA and mtDNA need to be used with particular care. Nonetheless, mtDNA variation can help us to hypothesize the ancestry of a given set of remains, and what race their owner might have been classified in, if we are clear about its limitations. The three questions outlined in this chapter provide a guide to these limitations in any given case. At one

Table 9. Skeletal Group from South Asia

Individual	1	2	3	4	5
Sequence	16356C, 146C, 195C	16224C, 16311C	16069T, 16126C, 295T	16223T, 16304C, 199C	16069T, 16126C, 16145A, 16231C, 16261T, 16355T, 150T, 152C, 195C, 215G, 295T, 310.1T
Haplogroup	U4b1a3	K	J1	M35b	J2a1a1a
Ancestry	European	European	European	South Asian	European
N/10428	4 Caucasian	17 Caucasian, 1 Hispanic, 13 "other"	7 Caucasian	0	0
Probable origin	United Stat	United States	United States	Indigenous	United States

end are those cases where a distinctive HVR sequence can be firmly placed in the phylogeny, on a branch with a restricted geographic distribution, and that maternal geographic origin is of high forensic significance; at the other end are those where a nondescript HVR sequence can only be placed in one or more broad areas of the phylogeny with little geographic restriction, and where minimal forensic significance can be placed on that maternal origin. An example of the former is provided by haplogroup Q1 remains recovered from Papua New Guinea; an example of the latter by remains that could be placed in D, G, or L3b1a recovered from Southeast Asia.

Acknowledgments

I am grateful to all of the personnel at the Armed Forces DNA Identification Laboratory (AFDIL) for everything that I have learned from them over the years. In particular, Suni Edson and I had lengthy discussions of many of the issues in this chapter, and Rebecca Just helped me get a handle on phylogeny. I would also like to thank all of my colleagues at the Joint POW-MIA Accounting Command—Central Identification Laboratory (JPAC-CIL), particularly those whose forensic anthropological reports I refer to in the preceding case studies.

Disclaimer

The opinions expressed herein are those of the author and are not those of the US government, Department of the Navy, or the United States Pacific Command.

References

1. Achilli A, Perego U A, Bravi C M, et al. The phylogeny of the four pan-American mtDNA haplogroups: Implications for evolutionary and disease studies. PLoS ONE, 2008, 3: e1764.
2. Allard M W, Polanskey D, Miller K, et al. Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. Forensic Science International, 2005, 148: 169–179.
3. Asari M, Umetsu K, Adachi N, et al. Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. Legal Medicine, 2007, 9: 237–240.
4. Behar D M, Rosset S, Blue-Smith J, et al. The genographic project public participation mitochondrial DNA database. PLoS Genetics, 2007, 3: e104.
5. Behar D M, van Oven M, Rosset S, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. American Journal of Human Genetics, 2012, 90: 675–684.
6. Behar D M, Villems R, Soodyall H, et al. The dawn of human matrilineal diversity. American Journal of Human Genetics, 2008, 82: 1130–1140.
7. Besaggio D, Fuselli S, Srikummool M, et al. Genetic variation in northern Thailand hill tribes: Origins and relationships with social structure and linguistic differences. BMC Evolutionary Biology, 2007, 7: S12. http://www.biomedcentral.com/1471-2148/7/S2/S12
8. Brues A M. The once and future diagnosis of race. In Skeletal attribution of race: Methods for forensic anthropology, ed. G. W. Gill and S. Rhine, 1–7. Albuquerque, NM: Maxwell Museum of Anthropology Anthropological Papers no. 4. 1990.
9. Cerezo M, Achilli A, Olivieri A, et al. Reconstructing ancient mitochondrial DNA links between Africa and Europe. Genome Research, 2012, 22: 821–826.
10. Chandrasekar A, Kumar S, Sreenath J, et al. Updating phylogeny of mitochondrial DNA macrohaplogroup M in India: Dispersal of modern human in South Asian corridor. PLoS ONE, 2009, 4: e7447.
11. Chen F, Wang S-Y, Zhange R-Z, et al. Analysis of mitochondrial polymorphisms in Guangdong Han Chinese. Forensic Science International: Genetics, 2008, 2: 150–153.
12. Derenko M, Malyarchuk B, Denisova G, et al. Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and Eastern Europe. PLoS ONE, 2012, 7: e32179.
13. Derenko M, Malyarchuk B, Grzybowski T, et al. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. American Journal of Human Genetics, 2007, 81: 1025–1041.
14. M Derenko, B Malyarchuk, T Grzybowski, et al. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. PLoS ONE, 2010, 5: e15214.
15. Diegoli T M, Irwin J A, Just R S, et al. Mitochondrial control region sequences from an African American population. Forensic Science International: Genetics, 2009, 4: e45–e52.
16. Ebenesersdóttir S S, Sigurðsson Á, Sánchez-Quinto F, et al. A new subclade of mtDNA haplogroup C1 found in Icelanders: Evidence of pre-Columbian contact? American Journal of Physical Anthropology, 2011, 144: 92–99.
17. Egeland T, Bøvelstad H M, Storvik G O, et al. Inferring the most likely geographical origin of mtDNA sequence profiles. Annals of Human Genetics, 2004, 68: 461–471.
18. Fornarino S, Pala M, Battaglia V, et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): A reservoir of genetic variation. BMC Evolutionary Biology, 2009, 9: 154. http://www.biomedcentral.com/1471-2148/9/154
19. Friedlaender J S, Friedlaender F R, Hodgson J A, et al. Melanesian mtDNA complexity. PLoS ONE, 2007, 2: e248.
20. Friedlaender J S, Schurr T, Gentz F, et al. Expanding southwest Pacific mitochondrial haplogroups P and Q. Molecular Biology and Evolution, 2005, 22: 1506–1517.
21. Gonder M K, Mortensen H M, Reed F A, et al. Whole-mtDNA genome sequence analysis of ancient African lineages. Molecular Biology and Evolution, 2007, 24: 757–768.
22. Hooshiar Kashani B, Perego U, Olivieri A, et al. Mitochondrial haplogroup C4c: A rare lineage entering America through the ice-free corridor? American Journal of Physical Anthropology, 2012, 147: 35–39.
23. Ingman M, Gyllensten U. Analysis of the complete human mtDNA genome: Methodology and inferences for human evolution. Journal of Heredity, 2001, 92: 454–461.
24. Jantz R, Ousley S. FORDISC 3.0: Computerized forensic discriminant functions. Knoxville: University of Tennessee. 2005.
25. Jin H J, Kwak K D, Hong S B, et al. Forensic genetic analysis of mitochondrial DNA hypervariable region I/II sequences: An expanded Korean population database. Forensic Science International, 2006, 158: 125–130.
26. Just R S, Leney M D, Barritt S M, et al. The use of mitochondrial DNA single nucleotide polymorphisms to assist in the resolution of three challenging forensic cases. Journal of Forensic Sciences, 2009, 54: 887–891.
27. Kayser M. The human genetic history of Oceania: Near and remote views of dispersal. Current Biology, 2010, 20:

R194–R201.
28. Kumar S, Bellis C, Zlojutro M, et al. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evolutionary Biology, 2011, 11: 293. http://www.biomedcentral.com/1471-2148/11/293
29. Lao O, Vallone P M, Coble M D, et al. Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal, and mitochondrial DNA. Human Mutation, 2010, 31: E1875–E1893.
30. Lee C, Măndoiu I I, Nelson, C E. Inferring ethnicity from mitochondrial DNA sequence. BMC Proceedings, 2011, 5(Suppl. 2): S11. http://www.biomedcentral.com/1753-6561/5/S2/S11
31. Lee E J, Anderson L M, Dale V, et al. MtDNA origins of an enslaved labor force from the 18th century Schuyler Flatts Burial Ground in colonial Albany, NY: Africans, Native Americans, and Malagasy? Journal of Archaeological Science, 2009, 36: 2805–2810.
32. Lee, H. Y., Song, I., Ha, E., Cho, S.-B., Yang, W. I., and Shin, K.-J. 2008. mtDNAmanager: A web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics 9:483. http://www.biomedcentral.com/1471-2105/9/483
33. Li H, Cai X, Winograd-Cort E R, et al. Mitochondrial DNA diversity and population differentiation in southern East Asia. American Journal of Physical Anthropology, 2007, 134: 481–488.
34. Maca-Meyer N, González A M, Larruga J M, et al. Major genomic mitochondrial lineages delineate early human expansions. BMC Genetics, 2001, 2: 13. http://www.biomedcentral.com/1471-2156/2/13
35. Macaulay V, Richards M, Hickey E, et al. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPS. American Journal of Human Genetics, 1999, 64: 232–249.
36. Malhi R S, Kemp B M, Eshleman J A, et al. Mitochondrial haplogroup M discovered in prehistoric North Americans. Journal of Archaeological Science, 2007, 34: 642–648.
37. Malyarchuk B A, Derenko M, Grzybowski T, et al. The peopling of Europe from the mitochondrial haplogroup U5 perspective. PLoS ONE, 2010, 5: e10285.
38. Malyarchuk B A, Derenko M, Perkova M, et al. Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs. European Journal of Human Genetics, 2008, 16: 1091–1096.
39. Malyarchuk B A, Perkova M, Derenko M, et al. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. Annals of Human Genetics, 2008, 72: 228–240.
40. Martínez-Cruzado J C. The history of Amerindian mitochondrial DNA lineages in Puerto Rico. In Island shores, distant pasts: Archaeological and biological approaches to the pre-Columbian settlement of the Caribbean, ed. S M Fitzpatrick, A H Ross. Gainesville: University Press of Florida, 2010: 54–80.
41. Martínez-Cruzado J C, Toro-Labrador G, Viera-Vera J, et al. Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. American Journal of Physical Anthropology, 2005, 128: 131–155.
42. Melton T, Peterson R, Redd A J, et al. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. American Journal of Human Genetics, 1995, 57: 403–414.
43. Mendizabal I, Valente C, Gusmão A, et al. Reconstructing the Indian origin and dispersal of the European Roma: A maternal genetic perspective. PLoS ONE, 2011, 6: e15988.
44. Monson K, Miller K, Wilson M, et al. The mtDNA population database: An integrated software and database resource for forensic comparison. Forensic Science Communications, 2002, 4(2).
45. Nohira C, Maruyama S, Minaguchi K. Phylogenetic classification of Japanese mtDNA assisted by complete mitochondrial DNA sequences. International Journal of Legal Medicine, 2010, 124: 7–12.
46. Olivieri A, Achilli A, Pala M, et al. The mtDNA legacy of the Levantine Early Upper Palaeolithic in Africa. Science, 2006, 314: 1767–1770.
47. Palanichamy M G, Zhang C-L, Mitra, et al. Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. BMC Evolutionary Biology, 2010, 10: 304. http://www.biomedcentral.com/1471-2148/10/304
48. Peng M.-S, Quang H H, Dang K P, et al. Tracing the Austronesian footprint in mainland Southeast Asia: A perspective from mitochondrial DNA. Molecular Biology and Evolution, 2010, 27: 2417–2430.
49. Perego U A, Angerhofer N, Pala M, et al. The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. Genome Research, 2010, 20: 1174–1179.
50. Razafindrazaka H, Ricaut F-X, Cox M P, et al. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. European Journal of Human Genetics, 2010, 18: 575–581.
51. Redd A J, Takezaki N, Sherry S T, et al. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Molecular Biology and Evolution, 1995, 12: 604–615.
52. Royal C D, Novembre J, Fullerton S M, et al. Inferring genetic ancestry: Opportunities, challenges, and implications. American Journal of Human Genetics, 2010, 86: 661–673.
53. Salihović M P, Barešić A, Klarić I M, et al. The role of the Vlax Roma in shaping the European Romani maternal genetic history. American Journal of Physical Anthropology, 2011, 146: 262–270.
54. Sekiguchi K, Imaizumi K, Fujii K, et al. Mitochondrial DNA population data of HV1 and HV2 sequences from Japanese individuals. Legal Medicine, 2008, 10(5): 284–286.
55. Soares P, Alshamali F, Pereira J B, et al. The expansion of mtDNA haplogroup L3 within and out of Africa. Molecular Biology and Evolution, 2012, 29: 915–927.
56. Starikovskaya E B, Sukernik R I, Derbeneva O A, et al. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Annals of Human Genetics, 2005, 69: 67–89.
57. Stefflova K, Dulik M C, Pai A A, et al. Evaluation of group genetic ancestry of populations from Philadelphia and Dakar in the context of sex-biased admixture in the Americas. PLoS ONE, 2009, 4: e7842.
58. Tanaka M, Cabrera V M, González A M, et al. Mitochondrial genome variation in eastern Asia and the peopling of Japana. Genome Research, 2004, 14: 1832–1850.
59. Torroni A, Schurr T G, Cabell M F, et al. Asian affinities and continental radiation of the four founding Native American mtDNAs. American Journal of Human Genetics, 1993, 53: 563–590.
60. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation, 2009, 30: E386-94. http://www.phylotree.org. Build 14 (dated April 5, 2012).
61. Wang W-Z, Wang C-Y, Cheng Y-T, et al. Tracing the origins of Hakka and Chaoshanese by mitochondrial DNA analysis. American Journal of Physical Anthropology, 2010, 141: 124–130.
62. Yu D, Jia X, Zhang A-M, et al. Mitochondrial DNA sequence variation and haplogroup distribution in Chinese patients with LHON and m.14484T > C. PLoS ONE, 2010, 5: e13426. ■