Int J Legal Med (2004) 118: 294–306
DOI 10.1007/s00414-004-0466-z

ORIGINAL ARTICLE

Anita Brandstätter · Christine T. Peterson · Jodi A. Irwin · Solomon Mpoke · Davy K. Koech · Walther Parson · Thomas J. Parsons

Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database

Received: 16 February 2004 / Accepted: 4 June 2004 / Published online: 10 July 2004

Abstract Large forensic mtDNA databases which adhere to strict guidelines for generation and maintenance, are not available for many populations outside of the United States and western Europe. We have established a high quality mtDNA control region sequence database for urban Nairobi as both a reference database for forensic investigations, and as a tool to examine the genetic variation of Kenyan sequences in the context of known African variation. The Nairobi sequences exhibited high variation and a low random match probability, indicating utility for forensic testing. Haplogroup identification and frequencies were compared with those reported from other published studies on African, or African-origin populations from Mozambique, Sierra Leone, and the United States, and suggest significant differences in the mtDNA compositions of the various populations. The quality of the sequence data in our study was investigated and supported using phylogenetic measures. Our data demonstrate the diversity and distinctiveness of African populations, and underline the importance of establishing additional forensic mtDNA databases of indigenous African populations.

A. Brandstätter · W. Parson
Institute of Legal Medicine, University of Innsbruck,
Müllerstrasse 44,
6020 Innsbruck, Austria

S. Mpoke · D. K. Koech
Immunology Laboratory, CBRD Kenya Medical Research Institute,
Mbagathi Way,
PO BOX 54840, 00200 Nairobi, Kenya

C. T. Peterson · J. A. Irwin · T. J. Parsons (✉)
Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology,
Building 101 1413 Research Boulevard,
Rockville, MD, 20850 , USA
e-mail: parsons@afip.osd.mil

C. T. Peterson
The Institute for Genomic Research,
Rockville, Maryland, USA

Keywords Mitochondrial DNA · Control region sequences · Africa · AMOVA

Introduction

Mitochondrial DNA analysis as a forensic tool is now well established (for a review see e.g., [1]). At a rapid pace, mtDNA typing is being implemented in laboratories worldwide, using standard approaches for sequencing HVI and HVII of the mtDNA control region [2]. The laboratory aspects of basic forensic mtDNA sequencing may be considered fully established, and advances continue at a rapid pace in the development of quick single nucleotide polymorphism-based (SNP) assays [3, 4], increased power of exclusion [5], and haplogroup determination [6]. However, an area that has lagged behind the laboratory development of mtDNA typing, and that continues to restrict the utility of mtDNA forensic testing, is the establishment of appropriate global population databases. This is especially true for populations outside of western Europe and the United States.

These statements relating to mtDNA database limitations may seem somewhat surprising in the light of the vast literature that exists on global mtDNA variation that comes from molecular anthropological investigations [7, 8, 9], and the great effort that forensic scientists have expended in establishing both local [10, 11, 12, 13] and large, combined databases suitable for forensic applications [14, 15, 16]. However, published anthropological mtDNA sequences are often too short for forensic purposes, usually restricted to HVI, and lack standardization from study to study in the sequence range that is reported. Some attempts have been made to establish standardization of sequence ranges and quality in forensic science [17, 18, 19]. However, the typical forensic target range of 16024–16365 for HVI and 73–340 for HVII is also rather arbitrary, and omits adjacent highly variable sites. Our present study targets the entire control region sequence, accessing much additional variation and providing a reference database useful for any boundaries of HVI/

HVII. The published quality standards for forensic mtDNA sequence analysis should be sufficient to ensure generation of correct sequences, if followed with care and expertise. Unfortunately, however, recent analyses have shown that mtDNA sequences published in forensic journals are no less prone to suffer widely from "phantom mutations" than those of anthropological studies [20, 21, 22, 23]. A principle reason for this is that most errors are not those of laboratory processing or primary data analysis, but of post-sequencing data transcription and tabulation. The recent extensive discussion as to the causes of mtDNA database errors should be taken greatly to heart, and we feel that quality standards for publication of forensic mtDNA databases should now be quite high.

The effort required to produce high quality mtDNA databases is large enough that appropriate forensic databases are not available for many world populations. As law enforcement becomes more of a global cooperative undertaking, and as mtDNA is increasingly used in international cases of missing persons identification and mass disasters (e.g. Shanksville, Pennsylvania, Pentagon, and World Trade Center terrorist attacks of Sept. 11, 2002; Armed Forces DNA Identification Laboratory [AFDIL] and New York City Office of Chief Medical Examiner, unpublished), establishment of global mtDNA databases for forensic purposes should be accelerated and systematically pursued.

In order for forensic mtDNA testing to be used, it is necessary to have an appropriate population database to determine the rarity of sequence types encountered in casework [1]. Ideally, the comparative database would represent the mtDNA haplotype frequencies from an assemblage of individuals who represent the potential pool of contributors for the particular questioned sample. It is well established that different ethnic, or population groups (see [24] for useful distinction between these terms) differ significantly in their mtDNA distributions [25, 26]. This basic observation precludes the pooling of global databases, but leaves incompletely answered the question of how finely divided mtDNA databases should be within ethnic or racial groups, at regional or local levels. This question has been most thoroughly examined in Europeans, where there has been some evidence for population homogeneity on a large scale [27]. However, even within Europeans, a recent detailed study revealed that despite the lack of statistically significant substructure at the population level, reflecting broad dispersal of a majority of mtDNA types, fluctuations in the frequencies of particular mtDNA types was detectable at the local level [28].

The high evolutionary rate and maternal inheritance of mtDNA makes it extremely useful for forensic testing and studies of population history. In the last decade, African mtDNA data have been examined closely in an effort to elucidate the history of human population expansion across the world. Nonetheless, the sampling of mtDNA variation in Africa is incomplete, particularly in the southeastern portion of the continent. To date, forensic studies from sub-Saharan African populations were restricted to variation revealed by sequence-specific oligonucleotide (SSO) probes [29, 30]. In addition, several studies of control region (CR) sequences [31, 32, 33] or a combination of RFLPs and CR sequences [34, 35, 36] have examined African populations. Together, these studies provide strong evidence that modern *Homo sapiens* originated in sub-Saharan Africa some 150,000–200,000 years ago, with only particular branches of the mtDNA lineage following human migrations out of Africa to the rest of the world.

It is now clear that human control region sequences evolve according to a complex pattern. Base composition is not uniform, transitions occur in greater frequencies than transversions, and the number of pyrimidine transitions in the L-strand exceeds the number of purine transitions [37, 38]. Another striking feature of mtDNA sequence evolution is the great variation of evolutionary substitution rates among sites [38, 39, 40, 41]. A model that most accurately describes the complex substitution process of mtDNA is necessary for accurate inference in evolutionary analysis. For example, accounting for rate heterogeneity to obtain unbiased estimates of the transition-transversion ratio is particularly important for impartial estimation of measures of genetic diversity and other parameters of population history [42]. Accurate mtDNA databases are extremely useful for evaluating these specific measures, as well as examining other phylogenetic and evolutionary parameters associated with mtDNA. Such analyses can reveal the presence of systematic errors in databases [20, 21, 22, 23], as well as provide information on the extent to which databases should be subdivided at the population and geographic levels.

Nairobi, the capital of Kenya, was founded in the late 1890s as a British railroad camp, in a region that was historically inhabited by pastoral tribal groups. In 1905 the city became the capital of the British East Africa Protectorate (called Kenya Colony from 1920 to 1963) and in 1963 Nairobi became the capital of independent Kenya. Presently, Nairobi can be considered a melting pot of tribal individuals who have, due to population pressures, lack of sustainability of ancestral lifestyles, and/or search for higher paying "white collar" jobs, immigrated to the city. Due to the large influx of people from rural areas, the population of Nairobi has grown from about half a million in 1969 to the present population of about 3 million. The immigrants represent many tribal elements from Kenya, a country with more than 70 ethnic groups divided into 3 linguistic groups namely Bantu (comprising tribes such as Embu, Kamba, Kikuyu, Kisii, Luhya, Meru, Kuria, Bukusu), Nilotes (Luo, Iteso, Nandi, Kipsigis, Marakwet, Maasai) and Cushites (Boran, Rendille, Somali). These ethnic groups have originated from diverse countries of Africa such as Zaire, Ethiopia, Tanzania, Uganda, Somalia and the Sudan. In addition to mtDNA variation reflecting this African history, the possibility also exists for genetic mixture with Europeans (primarily British colonialists), Arabs, and various Asian immigrants who arrived through British influence to work on the railroads.

We report here a database of entire mtDNA control region sequences from urban Nairobi. This population sample was chosen in order:

1. To begin to address the notable lack of forensic mtDNA databases from continental Africa
2. To permit evaluation of the evidentiary significance of mtDNA testing results from forensic casework performed specifically in Nairobi
3. To investigate issues regarding combination (pooling) of forensic databases in sub-Saharan African populations.

With regards to the latter, we tested for significant differences in the distribution of mtDNA types between the Nairobi population sample and other African population samples, and examined the variation and relationships of the Nairobi mtDNA sequences within the larger context of continental African mtDNA haplogroup distributions.

Materials and methods

DNA samples and extraction

DNA was extracted from bloodstained filter paper cards using either the Qiagen QIAamp DNA Mini kit or a modified buccal swab protocol on the Qiagen 9604 robot. The anonymous samples were obtained from 100 consenting volunteers from urban Nairobi (Kenya); known maternal relatives were excluded. All individuals appeared to be, and attested to be, of indigenous East African ancestry.

Amplification and sequencing of mtDNA

PCR amplifications were conducted using primers F15878 (AAA TGG GCC TGT CCT TGT AG) and R649 (TTT GTT TAT GGG GTG ATG TGA) or F15971 (TTA ACT CCA CCA TTA GCA CC) and R599 (TTG AGG AGG TAA GCT ACA TA). Automated PCR set-up was performed on the MWG RoboAmp 4200 robotic work-

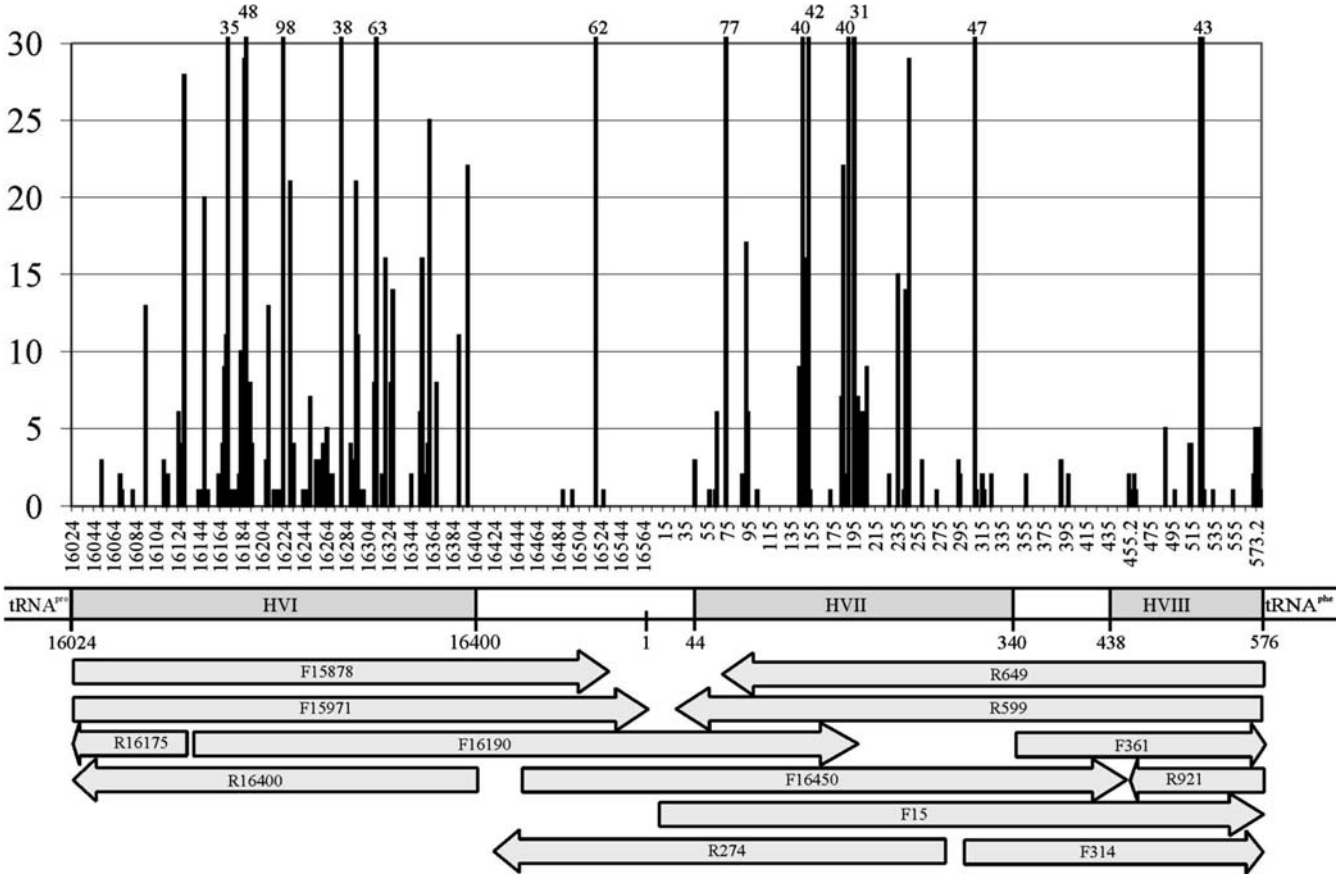

Fig. 1 Number and distribution of polymorphic positions in the human mitochondrial control region among 100 maternally unrelated persons from Nairobi. *X-axis* Nucleotide position in the control region. Nomenclature is in accordance with [43]. Positions 263 and 315.1 were removed from the diagram, as they were not variable in the Nairobi population sample (position 315.1 was shared by all individuals from the population sample; position 263 was found in 99% of all samples). *Y-axis* Number of individuals who show certain deviations from the reference sequence. The scale was limited to n=30 in order to allow for a better comparison of low values. Values above n=30 are given above the respective bars. Below the histogram, the recommended ranges for the hypervariable regions are depicted (see Results & Discussion section), together with a diagram representing the sequence ranges of the different sequencing primers

station using a 96-well format. Reactions consisting of 1× AmpliTaq Gold reaction buffer (Applied Biosystems), 200 μM each dNTP, 200 nM each primer and either 5 units (F15878/R649) or 2.5 units (F15971/R599) AmpliTaq Gold, were carried out in a volume of 50 μl. Cycling conditions for primers F15878/R649 were 95°C for 10 min and then 36 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2 min. Thermal cycler conditions for primers F15971/R599 were 95°C for 10 min and then 36 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min. PCR products were purified with exonuclease I (2.5 units) and shrimp alkaline phosphatase (5 units), and then sequenced with the Big Dye Terminator Cycle sequencing kit (Applied Biosystems, Foster City, CA) on the MWG RoboAmp 4200. Sequencing reactions were analyzed on either an Applied Biosystems 377 or an Applied Biosystems 3100 DNA Analyzer, following purification with AGTC columns (Edge Biosystems, Gaithersburg, MD). Primers used for sequencing, in addition to the amplification primers, were: F16190 (CCC CAT GCT TAC AAG CAA GT), F16450 (GCT CCG GGC CCA TAA CAC TTG), F15 (CAC CCT ATT AAC CAC TCA CG), F314 (CCG CTT CTG GCC ACA GCA CT), R274 (TGT GTG GAA AGT GGC TGT GC), R16400 (GTC AAG GGA CCC CTA TCT GA), and R16175 (TGG ATT GGG TTT TTA TGT A) (Fig. 1).

The 5' end of the control region of 5 samples (Nai040, Nai050, Nai059, Nai060, and Nai084) had C-stretch length heteroplasmy that degraded the quality of sequence reads for some sequencing primers. To confirm the last few bases of the control region, an additional amplification was performed with F361 (ACA AAG AAC CCT AAC ACC AGC) and R2216 (TGT TGA GCT TGA ACG CTT TC), and the amplicon sequenced with R921 (ACT TGG GTT AAT CGT GTG ACC).

The nucleotide positions in the mtDNA control region sequences considered for the analysis of the samples were 16024–16569 and 1–576. Nucleotides are numbered according to the revised Cambridge Reference Sequence (rCRS [43, 44]). The sequences were aligned using Sequencher software (GeneCodes, Ann Arbor, Michigan) according to a series of rules developed for the consistent placement of gaps [18].

Quality assurance

In order to ensure data quality, a redundant approach to data generation and analysis was used. At the Armed Forces DNA Identification Laboratory (AFDIL), duplicate amplifications were sequenced in both the forward and reverse directions. Consensus sequences were generated independently by two scientists, and the results compared to generate a final consensus sequence. The consensus sequence was then exported directly from Sequencher into a locked master database. The raw sequence lane files were transmitted to the Institute of Legal Medicine (ILM), Innsbruck, where they were again independently aligned and edited for incorporation into the EMPOP (EDNAP mtDNA Population Database) [45, 46]. As a final check, the AFDIL and ILM databases were compared; apart from nomenclature issues regarding C-stretches, there were no discrepancies. Sequence data in electronic form are available from the authors upon request, and have been submitted to GenBank, with accession numbers AY632902–AY632953, AY632955–AY632958, AY632960–AY632983, AY632985–AY633004.

To detect potential "phantom mutations" [21], the data were checked using the computer program SPECTRA ([21], available at http://www.stats.gla.ac.uk/~vincent/fingerprint/index.html). After preparing a binary matrix of the data in the input format for the phylogenetic network software package NETWORK (from the Fluxus Engineering Web site), the HVI and HVII "speedy transitions" listed in [21] were removed. A reduced median network was then constructed using NETWORK 4.0 [47], with a reduction threshold set to 99. The resulting network showed many reticulations (cubes) involving a number of additional sites. SPECTRA was run under the DOS prompt, computing the cube (*f*) and incompatibility (*s*) spectra of the filtered data. The output of SPECTRA consists of the two vectors *f* and *s*, where *f* counts the number of (nodes, links, squares, cubes, hypercubes,...) in the network, and *s* counts the number of sets of *k* positions that are "incompatible", which means that they are not consistent with a tree, without assuming homoplasious mutation. Using NETWORK, the number of links corresponding to each site was counted in the network. This measure was used in an ad hoc sequential decision process of checking for correctness and then removing sites until a perfect tree was obtained and the checking process stopped. In addition to the sites recommended for removal [21], the following were excluded according to the process described above: 16069, 16166, 16168, 16169, 16207, 16224, 16230, 16344, 16368, 16390, 16399, 16519, 64, 73, 89, 183, 198, 200, 228, 236, 247, 294, 489, 511, 513. Additionally, we disregarded all polymorphisms in the C-stretches of HVI (around np 16189), HVII (around np 310) and HVIII (around np 573), as well as variations in an AC-short tandem repeat in HVIII [48].

Pairwise comparisons

Pairwise comparisons within and between populations were performed using the program LookADat (J. Irwin, AFDIL, unpublished), that permits hypervariable length differences in C-stretches of HVI, HVII, and HVIII to be ignored in the comparisons.

Random match probability

The random match probability was calculated as the sum of the squares of the haplotype frequencies [25]. C-stretch length variants were ignored in distinguishing haplotypes for calculation of random match probability.

Population genetic and molecular evolution analysis

Molecular diversity indices and an analysis of molecular variance (AMOVA) were calculated in ARLEQUIN (Version 2.0 [49]). The Nairobi data were compared with 109 sequences from Mozambique [36], 109 sequences from Sierra Leone [50] and 1,148 sequences from African Americans [50]. Permutation tests (1,000 replicates) were used to evaluate the significance of calculated genetic distances between populations. For this alignment, sequences were trimmed to fit the greatest common range 16024–16365 and 73–340.

Classification of mtDNA

mtDNA haplotypes from Nairobi were classified into haplogroups and sub-haplogroups based on patterns of shared haplogroup-specific or haplogroup-associated polymorphisms, as reported in [8, 9, 36, 51, 52, 53, 54, 55].

Results and discussion

We have determined the nucleotide sequence of the complete mitochondrial DNA control region of 100 persons from Nairobi who are unrelated (as far as we know) in their maternal lineage (Table 1). Sequence comparisons led to the identification of 93 mitochondrial lineages as defined by 198 variable sites. The mean pairwise difference between individuals was 15.31 nucleotides (95% CI 15.16–15.47), compared to a mean pairwise difference of 9.32 nucleotides in an Austrian West-Eurasian population (273 unrelated individuals, data not shown).

The distribution of variable sites across the mtDNA control region in the Nairobi population sample is depicted in Fig. 1. Our data underline the existence of an additional hypervariable region (HVIII) located around positions 438–576 as postulated by [56]. The highest density of polymorphic sites was obtained for hypervariable region I (HVI), which contained 84 variable positions throughout the 342 bp region 16024–16365 (24.56%), and hypervariable region II (HVII) displayed 43 mutable sites within 268 bp (73–340, 16.05%). HVIII exhibited a slightly lower variability with 18 polymorphic sites distributed over 137 bp (13.14%). In contrast, the segments located between the hypervariable regions showed nucleotide substitution rates of only 3.45% (positions 16366–16569), 5.48% (positions 1–72) and 3.13% (positions 31–437). To take 16024–16365 as the range for HVI follows forensic traditions, however, sites adjacent to this range such as 16368, 16390, 16391, and 16399 are also quite variable, and would be useful to include within the range of HVI. There is a rather conservative 21 bp region in between, but such a relatively conservative region is also observed for 16024–16050. It would therefore be reasonable to follow population genetics practice and use 16024–16400 as the full HVI range also for forensic investigations. An analogous situation is found with HVII, which traditionally spans positions 73–340. However, there is a relatively unstable region around np 44, so 44–340 might constitute a more useful range for HVII (Fig. 1).

In five of the Nairobi samples, Nai040, Nai050, Nai059, Nai060, and Nai084, we observed an unstable polycytosine stretch (C-stretch) at the extreme 5' end of the control region, just a few bases prior to the start of the tRNAphe gene. This is characterized by the insertion of additional C residues into what is, in most people as well as the rCRS, a run of six C residues. As has been documented for the C-stretch regions of HVI and HVII [57, 58], this run of six or more Cs is apparently associated with a heteroplasmic mixture of multiple length variants (Fig. 2). This C-stretch has been encountered in other populations [56, 59], and, as with the Nairobi samples, the heteroplasmy is almost always so pronounced that it is difficult to determine even a majority length variant (but see the method we adopted). Sequence data from reverse sequencing primers show that the terminal ACA bases of the control region are retained just prior to the beginning of the tRNAphe gene, so we agree with others [18, 56, 59] in treating this condition as involving C insertions rather than multiple A-C transversions. We chose to designate length variation in this region by counting the number of cytosines present in the majority molecule type and listing those as insertions at 573 (Fig. 2). The majority molecule was inferred by checking the highest "virtual" peak in any of the downstream non-cytosine nucleotides. For example, the nucleotides after the Cs are ACAG. Due to the preceding length heteroplasmy, the single G appears as multiple virtual G peaks. The highest virtual G peak corresponds to the majority molecule type, and, logically, the C-stretch ends four nucleotides before that majority G peak. We have not observed a simple duplication of six Cs, as previously reported [18].

Random match probability

The mtDNA diversity present in the Nairobi population sample is extremely high and 88 of the 93 haplotypes were seen only once. The most common haplotype in the Nairobi database found was a profile assigned to haplogroup L3b at a frequency of 4%. The probability of a random match between two unrelated individuals was calculated to be 1:83 (1.2%) for the entire CR and 1:81 (1.24%) for HVI+HVII (Table 2). The latter is lower than the random match probabilities computed for similar sized HVI+HVII databases from Sierra Leone (1:52 [50]) and Mozambique (1:28 [36]). The Sierra Leone and Mozambique populations may reflect a more geographically localized sampling.

Comparison with other African populations

In order to evaluate the diversity observed in Nairobi in relation to global African mtDNA variation, we compared

Table 1 mtDNA control region haplotypes in Nairobi, Kenya

HG	Sample	HVI																HVII															
J1	Nai093	69	126	266	519													73	185	188	228	263	295	309.1C	315.1C	462	489								
L0a	Nai012	93	148	172	187	188G	189	223	230	311	320	519						64	89	93	152	189	236	247	263	315.1C	523d	524d							
L0a	Nai015	129	148	168	172	187	188G	189	223	230	311	320						64	93	185	189	200	236	247	263	309.1C	315.1C	523d	524d						
L0a	Nai021	93	129	148	168	172	187	188G	189	223	230	278	293	311	320			93	95C	185	189	236	247	263	315.1C	523d	524d								
L0a	Nai027	148	172	187	188A	189	223	230	311	320	497	519	527					93	95C	195	236	247	263	315.1C											
L0a	Nai030	129	148	168	172	187	188G	189	223	230	278	293	311	320				93	95C	185	189	236	247	263	309.1C	315.1C	523d	524d							
L0a	Nai031	129	144	148	168	172	187	188G	189	207	223	230	311	320	344			93	146	185	189	200	236	247	263	309.1C	315.1C	523d	524d						
L0a	Nai052	129	148	168	172	187	188G	189	223	230	311	320						64	93	152	185	189	200	236	247	263	315.1C	523d	524d						
L0a	Nai058	148	172	187	188G	189	223	230	311	320	519							64	93	146	152	185	189	236	247	263	309.1C	315.1C	523d	524d					
L0a	Nai069	51	129	148	172	187	188G	189	223	230	311	320						93	95C	185	189	236	247	263	309.1C	315.1C									
L0a	Nai075	93	148	172	173	187	188G	189	223	230	311	320	399	519				93	152	236	247	263	309.1C	315.1C	523d	524d									
L0a	Nai082	129	148	168	172	187	188G	189	223	230	311	320						64	93	152	185	189	236	247	263	309.1C	315.1C	523d	524d						
L0a	Nai101	129	148	168	172	187	188G	189	223	230	278	293	311	320	519			89	93	95C	185	189	236	247	263	309.1C	315.1C	523d	524d						
L0a	Nai106	129	148	166d	172	187	188G	189	223	230	311	320	519					93	152	189	204	207	236	247	263	315.1C	523d	524d							
L0a	Nai108	129	148	168	172	187	188G	189	223	230	287	293	294	311	519			93	151	152	185	189	236	247	263	315.1C	523d	524d	553						
L0a	Nai109	148	170	172	187	188G	189	223	230	311	320	519						93	152	189	236	247	263	315.1C	523d	524d									
L0f	Nai041	52	114A	169	187	189	223	230	278	290	311	325	327	368	519			143	146	152	185	189	247	263	315.1C	523d	524d								
L0f	Nai051	129	147A	169	172	187	189	223	230	278	311	327	368	390	399	519		44.1	146	185	189	204	207	247	309.1C	315.1C	511								
L0f	Nai084	169	172	187	189	223	230	265	278	311	319	327	354	368				93	95C	151	152	173	185	189	247	263	315.1C	523d	524d	573.1-5CCCCC					
L0f	Nai085	129	169	172	187	189	223	278	311	327	368	519						44.1	146	152	189	204	207	247	263	315.1C	511								
L0f	Nai089	52	129	169	187	189	223	230	278	290	311	325	327	344	354	368	519	143	146	152	185	189	247	263	309.1C	315.1C	523d	524d							
L0f	Nai096	129	169	172	187	189	223	278	311	327	368	519						44.1	146	152	189	204	207	247	263	315.1C	511								
L0f	Nai099	129	169	172	187	189	223	230	278	311	325	327	354	368				73	146	152	185	189	199	207	247	263	315.1C								
L0f	Nai103	52	129	169	187	189	223	230	278	290	311	325	327	354	360	368	519	143	146	152	189	247	263	315.1C	523d	524d									
L1b	Nai042	187	189	223	264	270	278	311	519									73	151	152	182	185T	195	198	247	263	315.1C	357	523d	524d					
L1b	Nai080	93	126	187	189	223	264	270	278	293	311	519						73	152	182	185T	189	195	247	263	309.1C	315.1C	357	523d	524d					
L1c	Nai036	129	163	187	189	209	223	278	293	294	311	360	519					73	151	152	182	183	186A	189C	247	263	315.1C	316A	523d	524d					
L1c	Nai077	129	163	187	189	209	223	278	293	294	298	311	360	519				73	152	182	186A	189C	194	198	247d	263	309.1C	315.1C	316A	523d	524d				
L1e	Nai023	129	148	166	183C	186	189	223	278	311	355	362						73	152	182	195	247	263	315.1C	455.1T	455.2T	459.1C	523d	524d						
L1e	Nai068	93	129	148	166	183d	187	189	223	278	311	355	362					73	151	152	182	195	247	263	315.1C	455.1T	455.2T	459.1C	513	523d	524d				
L1e	Nai107	111	129	148	166	172	187	189	223	254	278	360						73	189	195	247d	263	309.1C	315.1C	535										
L2	Nai083	183C	189	193.1	223	224	278	390	519									73	146	152	182	183	263	309.1C	315.1C	511									
L2a	Nai017	189	192	223	249	278	292	294	309	390								73	143	146	152	195	263	315.1C	523d	524d									
L2a	Nai018	81	189	192	223	234	278	294	309	390								73	146	152	195	263	295	309.1C	315.1C										
L2a	Nai020	189	192	223	278	294	309	390	519									73	146	152	195	263	315.1C												
L2a	Nai034	223	234	249	278	294	295	390	519									73	143	146	152	195	263	309.1C	315.1C										
L2a	Nai063	189	223	278	294	309	390											73	143	146	152	195	263	315.1C											
L2a	Nai079	93	223	256	278	294	309	390	519									73	143	146	152	195	263	315.1C											
L2a	Nai088	189	223	245	278	294	390	519										73	143	146	152	195	263	315.1C											
L2a	Nai112	145	182C	183C	189	223	224	278	294	390								73	143	146	152	189	195	199	263	309.1C	309.2C	315.1C							
L2d1	Nai013	93	129	172	189	278	300	354	390	399	519							73	146	150	195	263	315.1C	456											
L3b1	Nai008	124	223	278	362	519												73	263	315.1C	523d	524d													
L3b1	Nai011	93	124	223	278	362	519											73	263	315.1C	523d	524d													
L3b1	Nai032	124	223	278	362	519												73	263	315.1C	523d	524d													
L3b1	Nai033	93	124	223	278	362	519											73	263	315.1C	523d	524d													
L3b1	Nai091	93	124	223	278	362	519											73	263	315.1C	523d	524d													
L3b1	Nai095	93	124	223	278	362	519											73	263	309.1C	315.1C	523d	524d												
L3e1	Nai026	223	325d	327														73	150	185	189	263	309.1C	315.1C											
L3e1	Nai061	223	325d	327														73	150	185	189	263	309.1C	315.1C											
L3e1	Nai111	223	325d	327														73	150	152	185	189	263	309.1C	315.1C										
L3e1a	Nai029	185	209	223	327													73	150	152	189	195	200	207	263	309.1C	315.1C								

Table 1 (continued)

HG	Sample	HVI											HVII										
L3e1a	Nai087	185	223	311	327								73	150	185	189	263	315.1C					
L3e2b	Nai049	172	183C	189	223	320	519						73	150	195	198	263	315.1C					
L3e2b	Nai100	172	183C	189	223	320	519						73	150	195	241.1-3TAA	263	315.1C					
L3e3	Nai040	223	265T	519									73	150	195	263	309.1C	315.1C	523d	524d	573.1-5CCCCC		
L3e3	Nai059	223	265T	519									73	150	195	263	309.1C	315.1C	523d	524d	573.1-4CCCC		
L3e3	Nai060	82	223	265T	519								73	150	195	198	263	309.1C	315.1C	523d	524d	573.1-5CCCCC	
L3f	Nai010	186	209	223	309	310	311	327					73	263	315.1C								
L3f	Nai064	209	223	234	311	519							73	189	263	309.1C	315.1C	523d	524d				
L3fl	Nai053	129	209	223	292	295	311	519					73	189	200	263	315.1C						
L3fl	Nai071	209	223	292	311	519							73	146	189	200	263	315.1C					
L3fl	Nai074	129	209	223	292	295	311	519					73	150	263	272	309.1C	315.1C					
L3g	Nai022	172	223	293T	311	355	362	399					73	146	244	263	315.1C						
L3g	Nai028	223	261	293T	311	355	362	399	519				73	146	244	257	263	315.1C	497				
L3g	Nai037	192	223	311	399								73	146	263	309.1C	315.1C						
L3g	Nai039	183	223	261	287	293T	311	355	362	399	519		73	146	189	244	257	263	294	309.1C	315.1C		
L3g	Nai047	183	223	261	287	293T	311	355	362	399	519		73	146	244	257	263	294	309.1C	315.1C			
L3g	Nai048	172	209	223	293T	311	355	399					73	146	244	263	315.1C						
L3g	Nai050	189	223	287G	293T	311	355	362	399				73	146	152	244	263	315.1C	513	573.1-6 CCCCCC			
L3g	Nai054	172	223	293T	311	355	362	399					73	146	244	263	315.1C						
L3g	Nai056	223	311	354	399								73	146	153	263	315.1C						
L3g	Nai057	111	223	293T	311	355	362	399	519				73	146	244	263	315.1C	398	513	523d	524d		
L3g	Nai072	172	223	293T	311	355	362	399	519				73	146	244	263	309.1C	315.1C	391				
L3g	Nai086	111	153	223	293T	311	355	362	399	519			73	146	244	263	315.1C	398	513	523d	524d		
L3g	Nai090	172	192	223	249	293T	311	355	362	399	488	519	73	146	244	263	309.1C	315.1C	391				
L3g	Nai097	71	172	223	293T	311	355	362	399	519			73	146	244	263	315.1C						
L3g	Nai098	172	223	293T	311	355	399						73	146	244	263	315.1C						
L3g	Nai102	172	223	266	293T	311	355	362	399	519			73	146	244	263	309.1C	315.1C	391				
M1	Nai035	129	183C	189	193.1	223	249	311	359	519			73	195	263	309.1C	315.1C	489					
M1	Nai055	183C	189	193.1	223	249	311	519					73	195	263	309.1C	315.1C	489	524.1A	524.2C			
M1	Nai065	129	189	223	249	311	359	519					63	73	150	195	198	263	309.1C	315.1C	489	523d	524d
M1	Nai070	129	182	189	223	249	265C	311	519				73	195	263	315.1C	489						
(pre-HV)1	Nai078	114	126	362	519								57.1C	64	263	309.1C	315.1C						
U6a	Nai073	172	183C	189	193.1	219	278						73	103	146	195	263	309.1C	315.1C				
?	Nai014	207T	220	223	260	261	311	318	362	519			73	195	198	263	315.1C	325					
?	Nai110	93Y	169	207C	223	260	311	362	519				73	195	198	263	315.1C	325					
?	Nai016	165	189	192	223	311	399						73	146	152	207	263	315.1C					
?	Nai019	148	172	192	214	223	234	292	311	399			73	93	146	152	263	315.1C					
?	Nai024	129	172	174	192	218	223	256A	311	362			73	151	152	189C	195	263	294	315.1C	523d	524d	
?	Nai009	171	178	209	223	309	325						73	263	315.1C								
?	Nai025	93	223	260	278	311	519						73	150	152	189	263	315.1C	523d	524d			
?	Nai038	209	223	295	309								73	152	263	309.1C	315.1C						
?	Nai043	223	254	316									73	152	189	263	309.1C	315.1C	523d	524d	573		
?	Nai044	223	254	316									73	152	189	263	309.1C	315.1C	523d	524d	573		
?	Nai066	176	209	223	256								73	152	189	195	263	309.1C	315.1C	318			
?	Nai076	183C	189	223	243	311	362	519					73	195	263	315.1C							
?	Nai081	169	223	278	519								73	150	204	207	263	315.1C	523d	524d			
?	Nai094	126	209	223	519								73	228	263	315.1C							
?	Nai104	148	223	292	311	399							73	146	152	263	315.1C						

Table 1 (continued)

HG	Sample	HVI								HVII									
		69	169	171	223	278	519	73	150	204	207	263	315.1C	523d	524d				
?	Nai105																		

Sequence positions are numbered according to the rCRS [43, 44]. Numbers in the motifs refer to transitions unless otherwise designated, a deletion is indicated by *d*, insertions by *Site.1*, *Site.2*, etc..., and transversions are in bold with the variant indicated.
Left panel lists polymorphisms from 16024–16569 (less 16000) including HV1.
Right panel lists polymorphisms from 1–576 including HV2.

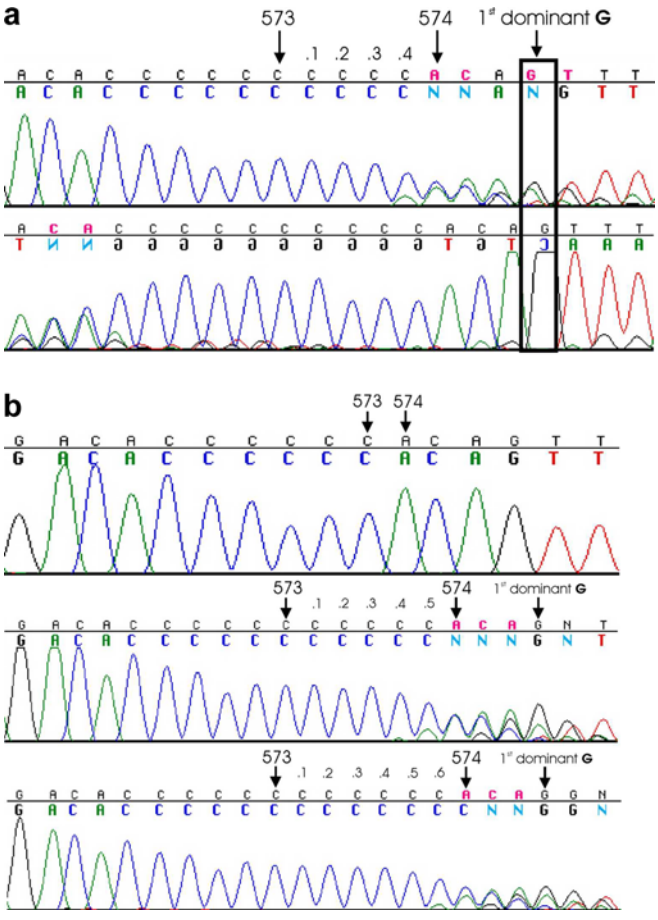

Fig. 2a-b Length heteroplasmy at the 5'-end of the control region. The indicated insertions (.1, .2, etc...) follow base position 573. Examples show differing C-stretch lengths and degrees of length heteroplasmy. **a** Example from one individual showing four cytosine insertions as the majority sequence type. *Upper panel* L-strand sequence, *lower panel* H-strand sequence. **b** Examples from three individuals showing no C-insertions (*upper panel*), 5 C-insertions (*middle panel*), and 6 C-insertions (*lower panel*)

the Nairobi sample set to databases of other African populations (Mozambique, Sierra Leone and African Americans Table 2). Only HVI and HVII data were considered, since only the Nairobi database included entire control region sequences. Pairwise comparisons showed a considerable number of matches between Nairobi and African American sequences: 19 individuals (15 out of 93 haplotypes) from Nairobi were found in African Americans (67 individuals, 15 out of 747 haplotypes). Additionally, 5 individuals (5 haplotypes) from Nairobi were found in Mozambique (13 individuals, 5 out of 64 haplotypes), and 2 individuals (1 out of 93 haplotypes) from Nairobi were found in Sierra Leone (2 individuals, 1 out of 81 haplotypes). Pairwise comparisons between populations revealed that 19% of the profiles from Nairobi matched individuals in the African American database, while 21% of the profiles (23 individuals, 9 out of 81 haplotypes) from Sierra Leone and 35% of the profiles (38 individuals, 18 out of 64 haplotypes) from Mozambique were also present in the African American database (33 individuals, 9 out of 747 African American haplotypes

Table 2 HVI + HVII (positions 16024–16365 and 73–340) molecular diversity measures for the analyzed African population databases

Population	Sample Size	Haplotypes[1]	Segregating sites[2]	Random matching probability	Mean pairwise differences (95% Confidence interval)	Reference
Nairobi	100	93 (93.0)	130 (21.0)	1.24%	15.31 (15.16–15.47)	
Mozambique	109	64 (58.7)	97 (15.7)	3.61%	14.41 (14.22–14.60)	[36]
Sierra Leone	109	81 (74.3)	89 (14.4)	1.91%	13.07 (12.93–12.22)	[50]
African Americans	1148	747 (65.1)	227 (36.7)	0.25%	14.10 (14.08–14.11)	[50]

The lower calculated random match probability of the African American population is due mainly to the fact that small sample sizes (e.g. n ? 100) give unique sequences an artificially high apparent frequency.
[1]Number of distinct haplotypes in sample (percentage of sample size).
[2]Number of sites variable in the sample (percentage of all sites).

found in Sierra Leone; 97 individuals, 18 out of 747 African American haplotypes found in Mozambique). Interestingly however, there was not a single match between individuals from Sierra Leone and Mozambique.

AMOVA was used to test for significant variation in the mtDNA distributions among the various populations (Table 3). Of the variance observed among the 4 populations, 97.04% is attributable to differences within populations, and 2.96% ($p<0.05$) represents differences among populations. This result was also noted previously [30], and it is the extreme intra-population variability of the mtDNA locus that renders it so valuable as a forensic tool. The comparison of Fst values from pairs of population samples (Tables 4 and 5) revealed that all of the samples in these African-derived population databases displayed highly significant differences in mtDNA distributions.

Haplogroup distribution

We categorized most of the Nairobi sequences to accepted sub-Saharan African haplogroups. One sample belonged to the north African haplogroup U6a and a few haplotypes from the Near and Middle East were also present in the Nairobi dataset. Although individuals included in this study were thought to be of East African ancestry, the introgression of mtDNA from other ethnic groups is not extremely surprising. For example, in African Americans, European mtDNA sequences are not uncommon and the most common European HVI/HVII type alone is seen at a frequency of 0.6% in a forensic African American database [50].

Table 3 AMOVA design and results

Source of variation	d.f.	Sum of squares	Variance components	Percent of variation
Among populations	3	178.775	0.28766 Va	2.96
Within populations	1448	13664.271	9.43665 Vb	97.04
Total	1451	13843.045	9.72431	

Table 4 Population pairwise FSTs

	Mozambique	Nairobi	Sierra Leone	African Americans
Mozambique	0.00000			
Nairobi	0.02938	0.00000		
Sierra Leone	0.04639	0.06141	0.00000	
African Americans	0.02667	0.04076	0.01730	0.00000

Table 5 FST*P*-values

	Mozambique	Nairobi	Sierra Leone	African Americans
Mozambique	*			
Nairobi	0.00000	*		
Sierra Leone	0.00000	0.00000	*	
African Americans	0.00000	0.00000	0.00000	*

Many previous surveys of aboriginal populations have demonstrated that the branches of the mtDNA tree (composed of groups of related haplotypes or haplogroups) are continent-specific, with virtually no mixing of mtDNA haplogroups from the different geographic regions [60]. The four most ancient mtDNA haplogroups (L0, L1, L2, and L3) are specific for sub-Saharan Africa. It has been shown that for African populations there is a close correspondence between the grouping identified by CR sequences and those identified by HR-RFLP haplotypes [35, 61]. In our case, 84% of the sequences could be assigned with reasonable confidence to accepted mtDNA haplogroups, based on shared constellations of polymorphisms previously shown to be diagnostic for, or associated with, the named haplogroups (Table 6). However, the high prevalence of reversion mutations (homoplasy) present in the mtDNA control region in many cases makes unambiguous haplotype assignment difficult, and we did not categorize 16 of the samples (Table 1). The latter may, at least in part, correspond to additional African haplogroups or sub-haplogroups that have yet to be characterized.

Table 6 Characteristic control region motifs found for the haplogroups in the Nairobi sample

Haplogroup	16024–16569 Sequence motif (np-16000)	1–576 Sequence motif
J1	69-126	73-263-295-489
L0a	(129)-148-172-187-188G-189-223-230-311-320	93-(152)-(185)-189-236-247-263-315.1C-523d-524d
L0f	(129)-169-(172)-187-189-223-(230)-278-311-327-368	146-152-(185)-189-247-263-315.1C
L1b	126-187-189-223-264-270-278-311-519	73-152-182-185T-195-247-263-315.1C-357-523d-524d
L1c	129-163-187-189-209-223-278-293-294-311-360-519	73-152-182-186A-189C-247-263-315.1C-316A-523d-524d
L1e	129-148-166-187-189-223-278-(311)-(355)-(362)	73-152-182-195-247-263-315.1C-455.1T-455.2T-459.1C-523d-524d
L2a	(189)-223-278-294-(309)-390	73-(143)-146-152-195-263-315.1C
L2c	223-278-390	73-146-152-182-263-315.1C
L2d1	129-278-300-354-390-399	73-146-263-315.1C
L3b1	(93)-124-223-278-362-519	73-263-315.1C-523d-524d
L3e1	223-325d-327	73-150-185-189-263-309.1-315.1C
L3e1a	185-223-327	73-150-189-263-315.1C
L3e2b	172-183C-189-223-320-519	73-150-195-263-315.1C
L3e3	223-265T-519	73-150-195-263-309.1-315.1C-523d-524d-573.1C-573.2C-573.3C
L3f	209-223-311	73-263-315.1C
L3f1	(129)-209-223-292-(295)-311-519	73-189-200-263-315.1C
L3g	223-293T-311-355-362-399	73-146-244-263-315.1C
M1	(129)-189-223-249-311-519	73-195-263-315.1C-489
pre-HV	126-362	263-315.1C
U6a	172-189-219-278	73-263-315.1C

Those positions that have been found in the majority, but not in all samples from a respective haplogroup are in parentheses.

Haplogroup L0 was represented by 23 individuals belonging to L0a and L0f (Table 1). Haplogroup L1 appeared in our dataset as three haplotype clusters (L1b, L1c, L1e, 7% of all samples) and haplogroup L2 comprised 10% of the Nairobi samples (L2a, L2c, L2d). The CR sequences of haplogroup L3 amounted to 37% of the Nairobi population sample (L3b, L3e1, L3e1a, L3e2b, L3e3, L3f, L3f1, L3g) and 4% of the Kenyan population sample belonged to the mitochondrial haplogroup M1 [51]. Haplogroups J1, (pre-HV)1, and U6a were each represented by one individual, haplogroups J and (pre-HV) 1 are also prominent in the Near East and the Arabian peninsula [62]. The particular (pre-HV)1 HVI sequence has a match in a sample from Israeli Druze [62].

Sequence range

We have targeted the entire control region for this database to permit access to additional discriminatory variation that resides outside of HVI-HVII in the control region [5, 48]. In the case of this Nairobi database, the entire control region discriminates only one additional haplotype as compared to HVI-HVII sequencing, so it may seem to give only a marginal return for the additional effort. However, we have found that in high volume mtDNA casework, especially in cases of mass fatalities, situations are routinely encountered where multiple references match in HVI-HVII, which can be distinguished by variation from the entire control region. Accordingly, robust, short amplicons suitable for degraded samples have been designed to target these regions [63]. Moreover, most laboratories require two or more differences between sequences before exclusions can be reported [64, 65, 66], and targeting additional variation can make the difference between "inconclusive" and "exclusion." To have this additional variation available as a forensic tool in its most useful form requires databases that cover the entire control region. We would like to urge that, if possible, the entire control region be sequenced for population reference databases of the future.

Sequence data quality assurance

Phylogenetic methods have been put to use recently in identifying errors in published databases [20, 21, 22, 23]. We investigated this approach [21] as a tool for assessing the quality of our database as well. The method makes use of the known patterns of variation in mtDNA, especially the fact that particular sites evolve quickly and vary extensively in populations, while others are far less rapidly mutating or invariant. Phylogenetic analysis maps the patterns of evolutionary mutations represented in an mtDNA database. Departures from expectations in these patterns—i.e., a site known to evolve slowly appearing to evolve quickly—can flag systematic errors in a database. In a high quality database, exclusion of "speedy transitions," the fastest sites, should permit the construction of a phylogenetic network with few reticulations involving parallel mutations, so that the remaining sites involved in reticulations can either be checked (as correct) or identified as errors. In our treatment, this was a greedy sequential process. Partly because we analyzed the entire

control region, our phylogenetic network showed many reticulations even after the "standard" speedy transitions [21] had been removed. It wasn't until we had gone through several cycles of removing sites involved in network reticulations from the analysis (each time checking definitively that the discarded data were correct) that we obtained an incompatibility spectrum of (1 49) and a cube spectrum of (50 49), pointing to the fact that all characters were pairwise compatible. The ratio between the numbers of speedy and weighty transitions in our data was calculated as 1.5. The ratio of the number of weighty transitions to the number of transversions plus insertions and deletions (again disregarding C-runs), denoted as the "WTTI ratio" by [21], was estimated as 1.6, similar to the WTTI ratio of 1.5 [21] found for HVI (16051–16365) of another high quality African dataset [67].

Conclusions

Errors in mtDNA databases, including many published in the forensic literature, have received much attention in the last few years [20, 21, 22, 23], and it can safely be said that both the fields of forensic genetics and molecular anthropology would be well served at this point by a strong emphasis on producing, compiling, and distributing mtDNA databases with as few errors as possible. We are not aware of any convincing argument that sporadic errors in mtDNA databases pose a substantial threat to the interpretation of mtDNA evidence, but clearly every effort should be taken to minimize errors, and databases with widespread and/or systematic errors should be purged. As has been pointed out (e.g., [20, 22]), many errors are introduced not in the laboratory, but in processes of data tabulation and transposition. (With this in mind, although effort was made to ensure the correctness of Table 1, we advise anyone wishing to analyze or use the data reported in this paper to contact us for direct transfer in electronic format or to download the sequences from GenBank). Our phylogenetic examination of the quality of the Nairobi database confirmed the absence of major systematic errors (for example, a mix-up of site designations). While the phylogenetic approach has proven extremely beneficial in uncovering database errors, it constitutes a search tool for discovering particular types of errors, rather than a defined test for database correctness. Ultimately, our confidence in the quality of the Nairobi database resides in the careful and redundant approach we have taken to generating and analyzing the data, in crosschecking the compiled database, and in the use of bioinformatic tools to avoid manual manipulations of the data.

We report a new database of entire mtDNA control region sequences from Nairobi. The low average random match probability indicates a strong utility for mtDNA testing, and the database will now permit some assessment of the relative rarity of mtDNA types encountered in casework. Such a small sample, however, will not provide accurate estimates of the actual population frequencies of haplotypes, and certainly will miss a large number of haplotypes that are present [68]. Our AMOVA results clearly indicate that various African populations differ significantly in the distribution of mitochondrial DNA control region sequences. Therefore, it would be inappropriate to pool the databases of Nairobi, Mozambique and Sierra Leone for the purposes of establishing a single, large pan-African database. In these population samples, unbiased estimates of haplotype frequencies can only be obtained if the appropriate database is used, as evidenced by the fact that samples seen only once in the Nairobi database occurred at significantly higher frequencies in the similarly sized Mozambique database.

While the number of mtDNA sequences from European populations in forensic databases is now quite large, the effort to increase the number of forensically established African mtDNA population databases is only beginning. The results of our study highlight the distinctiveness of various African populations, and consequently, the need for additional databases from indigenous African populations. Further studies of indigenous sub-Saharan populations would add valuable information to the currently small pool of existing data.

Acknowledgements The authors wish to thank Hans-Jürgen Bandelt (University of Hamburg, Germany) for many helpful comments and Vincent Macaulay (University of Glasgow, UK) for valuable assistance with SPECTRA. The manuscript benefited greatly from comments made by two anonymous reviewers. We thank Jennifer O'Callaghan and Rebecca Just (AFDIL) for assistance with database confirmation, and James Ross (AFDIL) and Jon Norris (Future Technologies, Inc., Fairfax, VA, USA) for bioinformatics assistance. We thank James C. Canik, Kevin S. Carroll, and Brion C. Smith (AFDIL) for logistical, administrative, and moral support. 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.

References

1. Holland MM, Parsons TJ (1999) Mitochondrial DNA sequences analysis—validation and use for forensic casework. Forensic Sci Rev 11:21–49
2. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B (1995) Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74
3. Andreasson H, Asp A, Alderborn A, Gyllensten U, Allen M (2002) Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. Biotechniques 32:124–133
4. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med 118:147–157
5. Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ (2003) Single nucleotide polymorphisms over the entire mtDNA genome that increase the forensic power of mtDNA testing in Caucasians. Int J Legal Med 118:137–146
6. Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. Int J Legal Med 117:291–298

7. Richards M, Rengo C, Cruciani F et al. (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. Am J Hum Genet 72:1058–1064
8. Bandelt H-J, Alves-Silva J, Guimaraes PE et al. (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65:549–563
9. Salas A, Richards M, De la Fe T et al. (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111
10. Imaizumi K, Parsons TJ, Yoshino M, Holland MM (2002) A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. Int J Legal Med 116:68–73
11. Derenko MV, Grzybowski T, Malyarchuk BA et al. (2003) Diversity of mitochondrial DNA lineages in south Siberia. Ann Hum Genet 67:391–411
12. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66:261–283
13. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobnic K, Miscicka-Sliwka D (2003) Mitochondrial DNA variability in Bosnians and Slovenians. Ann Hum Genet 67:412–425
14. Röhl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med 115:29–39
15. Allard MW, Miller K, Wilson M, Monson K, Budowle B (2002) Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific Working Group on DNA Analysis Methods. J Forensic Sci 47:1215–1223
16. Miller KW, Budowle B (2001) A compendium of human mitochondrial DNA control region: development of an international standard forensic database. Croat Med J 42:315–327
17. Miller KW, Dawson JL, Hagelberg E (1996) A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. Int J Legal Med 109:107–113
18. Wilson MR, Allard MW, Monson K, Miller KW, Budowle B (2002) Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Int 129:35–42
19. Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124–132
20. Bandelt H-J, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115:64–69
21. Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150–1160
22. Forster P (2003) To err is human. Ann Hum Genet 67:2–4
23. Dennis C (2003) Error reports threaten to unravel databases of mitochondrial DNA. Nature 421:773–774
24. Forster P, Cali F, Röhl A et al. (2002) Continental and subcontinental distributions of mtDNA control region types. Int J Legal Med 116:99–108
25. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370–382
26. Budowle B, Allard MW, Fisher CL et al. (2002) HVI and HVII mitochondrial DNA data in Apaches and Navajos. Int J Legal Med 116:212–215
27. Melton T, Wilson M, Batzer M, Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of European populations. J Forensic Sci 42:437–446
28. Pfeiffer H, Forster P, Ortmann C, Brinkmann B (2001) The results of an mtDNA study of 1,200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. Int J Legal Med 114:169–172
29. Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M (2001) Diversity and heterogeneity in mitochondrial DNA of North American populations. J Forensic Sci 46:46–52
30. Melton T, Ginther C, Sensabaugh G, Soodyall H, Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of sub-Saharan African populations. J Forensic Sci 42:582–592
31. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507
32. Soodyall H, Jenkins T (1993) Mitochondrial DNA polymorphisms in negroid populations from Namibia: new light on the origins of the Dama, Herero and Ambo. Ann Hum Biol 20:477–485
33. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90:19–30
34. Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704
35. Graven L, Passarino G, Semino O et al. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345
36. Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet 65:439–458
37. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526
38. Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613–623
39. Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16:1357–1368
40. Parsons TJ, Muniec DS, Sullivan K et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368
41. Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37:347–354
42. Misawa K, Tajima F (1997) Estimation of the amount of DNA polymorphism when the neutral mutation rate varies among sites. Genetics 147:1959–1964
43. Anderson S, Bankier AT, Barrell BG et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
44. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147
45. Parson W, Brandstätter A, Pircher M, Steinlechner M, Scheithauer R (2004) EMPOP—The EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database. In: Doutrempuich C, Morling N (eds) Progress in forensic genetics 10, international congress series 1261, pp 106–108
46. Parson W, Brandstätter A, Alonso A et al. (2004) The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. Forensic Sci Int 139:215–226
47. Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753
48. Lutz S, Weisser HJ, Heizmann J, Pollak S (1997) A third hypervariable region in the human mitochondrial D-loop. Hum Genet 101:384
49. Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.0: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

50. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2003) The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Science Communications Online
51. Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441
52. Brehm A, Pereira L, Bandelt H-J, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. Ann Hum Genet 66:49–60
53. Torroni A, Rengo C, Guida V et al. (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356
54. Mishmar D, Ruiz-Pesini E, Golik P et al. (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A 100:171–176
55. Salas A, Richards M, Lareu MV et al. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465
56. Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111:67–77
57. Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248–256
58. Stewart JE, Fisher CL, Aagaard PJ et al. (2001) Length variation in HV2 of the human mitochondrial DNA control region. J Forensic Sci 46:862–870
59. Brandstätter A, Niederstätter H, Parson W (2004) Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region. Int J Legal Med 118:47–54
60. Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. Gene 238:211–230
61. Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000) mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. Am J Hum Genet 66:1362–1383
62. Macaulay V, Richards M, Hickey E et al. (1999) The emerging tree of west Eurasian mtDNAs: a synthesis of control region sequences and RFLPs. Am J Hum Genet 64:232–249
63. Edson SM, Ross JP, Coble MD, Parsons TJ, Barritt SM (2004) Naming the dead—Confronting the realities of rapid identification of degraded skeletal remains. Forensic Sci Rev 16:63–90
64. Bär W, Brinkmann B, Budowle B et al. (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196
65. Carracedo A, Bär W, Lincoln P et al. (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
66. Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124:83–91
67. Alves-Silva J, Silva SM da, Guimaraes PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461
68. Pereira L, Cunha C, Amorim A (2004) Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. Int J Legal Med 118:132–136