European Journal of Human Genetics (1999) 7, 459–468

http://www.stockton-press.co.uk/ejhg

ARTICLE

Analysis of mtDNA HVRII in several human populations using an immobilised SSO probe hybridisation assay

David Comas[1], Rebecca Reynolds[2] and Antti Sajantila[1]

[1]*Department of Forensic Medicine, University of Helsinki, Finland*
[2]*Roche Molecular Systems, Alameda, CA, USA*

Several populations were typed for the hypervariable region II (HVRII) of the mitochondrial DNA (mtDNA) control region using immobilised sequence-specific oligonucleotide (SSO) probes. A total of 16 SSO probes was used to type 1081 individuals from eight different ethnic groups (African Americans, Somali, US Europeans, US Hispanics, Bosnians, Finns, Saami and Japanese). Data was compared with already published sequence data by analysis of principal components, genetic distances and analysis of the molecular variance (AMOVA). The analyses performed group the samples in several clusters according to their geographical origins. Most of the variability detected is assigned to differences between individuals and only 7% is assigned to differences among groups of populations within and between geographical regions. Several features are patent in the samples studied: Somali, as a representative East African population, seem to have experienced a detectable amount of Caucasoid maternal influence; different degrees of admixture in the US samples studied are detected; Finns and Saami belong to the European genetic landscape, although Saami present an outlier position attributable to a strong maternal founder effect. The technique used is a rapid and simple method to detect human variation in the mtDNA HVRII in a large number of samples, which might be useful in forensic and population genetic studies.

Keywords: mtDNA; HVRII; SSO hybridisation; principal component analysis; genetic distances; AMOVA analysis; population admixture

Introduction

Since the publication of the complete sequence of the human mitochondrial genome,[1] mitochondrial DNA (mtDNA) has been used in forensic identity testing[2–4] and especially in human population studies.[5–12] For these purposes mtDNA is particularly useful due to its unique properties, such as high number of copies per cell, maternal inheritance pattern, lack of recombination and its fast substitution rate. Different approaches for typing mtDNA have been used in these analyses: basically, low and high resolution restriction enzyme studies as well as DNA sequence analysis of the control region. A dot blot approach using sequence specific oligonucleotide (SSO) probes has also been described.[13] Most of the mtDNA sequence analysis

Correspondence: Antti Sajantila, Department of Forensic Medicine, PO Box 40, 00014 University of Helsinki, Finland. Tel: +358 9 19127472; Fax: +358 9 19127518; E-mail: antti.sajantila@helsinki.fi
Received 7 September 1998; revised 23 December 1998; accepted 19 January 1999

studies have focused on the first of the two hypervariable regions (HVRI and HVRII) within the control region. This is due to the slightly higher level of polymoprhism at HVRI compared to that of HVRII. Although fewer data are available regarding the diversity of HVRII at a population level, many of its positions are polymorphic and useful in establishing mtDNA diversity.

Whilst the restriction enzyme, DNA sequencing, and dot blot methods provide valuable diversity information, they screen less than 20% of the whole mtDNA sequence. To determine the diversity of the complete sequence of the mtDNA genome, approaches such as the microchip sequencing technique[14] will need to be employed. However, until this technique is available and validated, simple methods allowing rapid typing of mtDNA would be of utmost importance for forensic and molecular genetic laboratories that screen a large number of samples.

In this paper we demonstrate that a modified dot blot approach in which SSO probes are immobilised on membrane strips can be a rapid and powerful technique to detect the major mtDNA HVRII variants. This method was used to type mtDNA from several human populations and these data were compared to published mtDNA sequence data for HVRII. We also describe the variation observed among the populations analysed.

Materials and Methods

Population Samples

A total of 1081 individuals from eight populations were typed: 200 African American individuals, 200 individuals living in the US with European ancestry, 200 US Hispanic individuals, 90 Somali, 89 Japanese, 104 individuals from Bosnia, 104 Finns and 94 Saami from the Kola Peninsula. Genomic DNA was extracted from whole blood using standard methods.[15]

Published population data on HVRII mtDNA sequences was used in comparison with the present results: Mandenka,[16] British,[17] Tuscans,[18] Havik,[19] Indonesians,[20] Papua New Guineans,[20] Ngobe,[21] Huetar[22] and Mapuche.[23] This data set was selected in order to represent most of the major continents and avoid population studies with small numbers of individuals.

Analysis of mtDNA Polymorphism

MtDNA HVRII was amplified by PCR using the primers HVIIL (5'CACCCTATTAACCACTCACG3') and HVIIR (5'CTGTTAAAAGTGCATACCGCCA3'). Both primers were biotinylated at the 5' end. Amplifications were performed using 100 ng of genomic DNA in a final volume of either 30 μl or 60 μl. The cycling conditions were as follows: an initial denaturation stage of 92°C for 12 min, 34 cycles of 92°C for 30 s, 60°C for 30 s and 72°C for 30 s, and a final extension stage of 72°C for 10 min.

Sixteen sequence-specific oligonucleotides immobilised on nylon membranes were used. These probes covered five of the most variable regions of HVRII and were named A, B, C, D and E (Table 1) as described by Stoneking *et al*.[13] Complete mtDNA haplotype frequencies are available upon request from DC.

Membranes were prehybridised in 3 ml of hybridsation solution (5 × SSPE, 0.5% SDS) for 10 min at 55°C. After

Table 1 Nucleotides distinguished by each of the probes in the five mtDNA HVRII regions tested (A, B, C, D and E). In all regions, the '1' probe is specific for the reference sequence.[1] The HVRII position number refers to the nucleotide substitution from the reference sequence detected by the specific probe. All the substitutions consisted of transitions except probe E2 that detects the insertion of one or more Cs at position 309

Region	*Probe*	*Sequence variation*	*Position*
A	A1	– – – – – A – – – – – – – – – – – –	–
	A2	– – – – – G – – – – – – – – – – – –	73G
B	B1	– – – – – – – – T – – – C – T – – – – – – – – –	–
	B2	– – – – – – – – C – – – C – T – – – – – – – – –	146C
	B3	– – – – – – – – T – – – C – C – – – – – – – – –	152C
	B4	– – – – – – – – C – – – C – C – – – – – – – – –	146C/152C
	B5	– – – – – – – – T – – – T – T – – – – – – – – –	150T
	B6	– – – – – – – – T – – – T – C – – – – – – – – –	150T/152C
	B7	– – – – – – – – C – – – T – C – – – – – – – – –	146C/150T/152C
C	C1	– – – – A – – – – – T – – C – A – – – – – – –	–
	C2	– – – – A – – – – – C – – C – A – – – – – – –	195C
	C4	– – – – A – – – – – C – – T – A – – – – – – –	195C/198T
	C5	– – – – G – – – – – T – – C – G – – – – – – –	189G/200G
D	D1	– – – – – – – – – – – G – – – – – – – – – – – –	–
	D2	– – – – – – – – – – – A – – – – – – – – – – – –	247A
E	E2		309 > 7Cs

prehybridisation, 20 µl of the PCR product were mixed with the same amount of denaturation solution (1.6% NaOH), added directly to 3 ml of hybridisation solution, and hybridisation was carried out for 15 min at 55°C. Membranes were rinsed in wash solution (2 × SSPE, 0.5% SDS) at room temperature; 8 µl of horseradish peroxidase-streptavidin (HRP-SA) enzyme conjugate (PE Biosystems, Foster City, Ca, USA) were added to 3 ml of the hybridisation solution and incubated at 55°C for 5 min. Membranes were rinsed again in wash solution at room temperature, then soaked in wash solution at 55°C for 12 min followed by a final wash at room temperature. Membranes rinsed with 3 ml of citrate buffer (0.1 M sodium citrate, pH 5) and then 4µl 3% H_2O_2 and 150 µl of Chromogen TMB (3,3',5,5'-Tetramethylbenzidine) solution (PE Biosystems, Foster City, Ca, USA) were added to 3 ml of citrate buffer and incubated for 30 min at room temperature. Finally, membranes were washed three times in 3 ml of H_2O for 5 min.

In cases where two hybridisation signals were obtained in SSO typing, the samples were sequenced for HVRI and HVRII regions in order to exclude contamination or find evidence for heteroplasmy. Sequence reactions were performed using Big Dye sequencing kit (PE Biosystems, Foster City, Ca, USA) and same primers as for DNA amplification. Sequence reactions were resolved by capillary electrophoresis and sequences determined using software dedicated for this purpose according to manufacturer' instructions (CE310, PE Biosystems, Foster City, Ca, USA).

Statistical Analysis

Sequence variants and mtDNA haplotype frequencies were estimated by direct counting and genetic diversity[24] in each population was calculated according to the formula, $h = (1 - \Sigma x^2)\ n/(n-1)$, where x is the frequency of each mtDNA haplotype in the population and n is the sample size. The probability of two random individuals in a population having the same mtDNA haplotype is $P = \Sigma x^2$.

F_{ST}-related genetic distances based in the variant frequencies were computed between pairs of populations.[25] A neighbour-joining tree[26] was built from the genetic distance matrix by means of the PHYLIP 3.5c package,[27] and tree robustness was assessed through 1000 bootstrap iterations.[28]

Principal component (PC) analysis was performed on the correlation matrix of the frequencies of each of the probes of the five regions analysed using the SPSS package.

In order to ascertain the proportion of the genetic variance due to differences within or between populations, genetic variance was hierarchically apportioned through the analysis of molecular variance (AMOVA) performed with the Arlequin program[29] and the significance of the estimated variance components was tested by 1000 iterations.

Results

Frequency of mtDNA Sequence Variants

Most frequently one of the SSO probes in each region gives a positive signal. However, some individuals carry sequence variants that do not hybridise with any of the probes and a zero is assigned to that region for the mtDNA haplotype. Therefore, there are 720 possible mtDNA haplotypes and we found 121 in the present sample set. All individuals presented unambiguous specific hybridisations except for seven individuals, who presented hybridisation of two different probes for the same HVRII region. These ambiguous samples were sequenced to assure the correct typing. None of the samples, however, were included in the further statistical analyses performed. In all cases the sequence variants, with two nucleotides in particular positions were observed in HVRII region. Furthermore, no such signals were observed in the HVRI region.

The frequencies of the sequence variants at the five HVRII mtDNA regions for the eight populations tested and for the populations used for comparison are shown in (Table 2). Some of the HVRII variants and mtDNA haplotypes detected by the specific SSO probes seem to be specific for some groups of populations. An A at position 73 (detected by probe A1) is associated with European haplogroups H and V according to Torroni *et al*.[8,30] Position 73 is useful to classify phylogenetically the mtDNA lineages but unfortunately scanty knowledge of the HVRII variation compared with HVRI does not allow us to classify our samples in halplogroups already defined by HVRI positions. Samples presenting the CRS mtDNA haplotype seem also to belong exclusively to the European mtDNA pool. In our sample set African populations also present some specific HVRII variants such as 146C–150T–152C (probe B7), 195C–198T (probe C4) and 247A (probe D2), which could be useful to assign mtDNA lineages in admixture cases.

Sequence Diversity

Sequence diversity and the probability of two randomly selected individuals in a population having an identical mtDNA haplotype are shown in (Table 3). Somali (0.960) show a high sequence diversity compared with the rest of the populations tested. This is consistent with the high diversity described in African populations also for several other genetic markers such as microsatellites,[31–34] minisatellites,[35] nuclear haplotypes[36] and Alu insertions.[37] The other African sample, the Mandenka, presents a sequence diversity (0.925) similar to that which is present in the European and Asian populations. The Saami presented a very low sequence diversity compared with the rest of the European groups sampled, because of the high frequency of a single mtDNA haplotype. Native American populations used for comparison also presented very low sequence diversities consistent with other studies,

Table 2 Frequencies of the variants at five HVRII mtDNA regions for 17 populations

Probe	AFR (200)	EUR (200)	HIS (200)	SOM (90)	JAP (89)	FIN (104)	SAA (94)	BOS (104)	MAN (119)	BRI (100)	TUS (49)	HAV (48)	IND (24)	PNG (26)	NGO (47)	HUE (27)	MAP (39)
A1	0.080	0.395	0.075	0.258	–	0.365	0.223	0.510	0.017	0.510	0.429	0.021	–	–	0.106	0.370	–
A2	0.915	0.600	0.925	0.742	1	0.606	0.713	0.490	0.983	0.449	0.571	0.958	1	1	0.894	0.630	1
0	0.005	0.005	–	–	–	0.029	0.064	–	–	0.041	–	0.021	–	–	–	–	–
B1	0.182	0.645	0.449	0.337	0.517	0.683	0.415	0.692	0.059	0.620	0.551	0.500	0.292	0.231	0.319	0.296	0.462
B2	0.005	0.061	0.056	0.011	0.079	0.087	0.011	0.067	–	0.080	0.041	0.083	0.458	0.538	–	–	–
B3	0.176	0.127	0.066	0.157	0.169	0.077	0.021	0.096	0.311	0.120	0.204	0.271	0.083	–	–	–	0.179
B4	0.152	0.051	0.035	0.112	0.034	0.010	–	0.029	0.050	0.050	–	–	–	–	–	–	0.051
B5	0.182	0.041	0.056	0.157	0.169	0.106	0.489	0.038	0.059	0.040	0.082	0.063	–	–	–	–	–
B6	0.086	0.030	0.025	0.022	0.022	0.010	0.021	0.019	0.017	0.040	0.041	0.042	–	–	–	–	–
B7	0.051	0.005	0.010	–	–	–	–	–	0.361	–	–	–	–	–	–	–	–
0	0.167	0.041	0.303	0.202	0.011	0.029	0.043	0.058	0.143	0.050	0.082	0.042	0.167	0.231	0.681	0.704	0.308
C1	0.185	0.720	0.830	0.378	0.764	0.721	0.957	0.663	0.176	0.750	0.673	0.688	0.667	0.808	0.957	0.630	0.846
C2	0.355	0.145	0.070	0.222	0.034	0.125	0.032	0.163	0.294	0.160	0.143	0.042	0.125	0.115	–	–	–
C4	0.105	–	0.025	0.133	–	–	–	–	0.387	–	–	–	–	–	–	–	–
C5	0.100	–	0.030	0.022	0.011	–	–	–	0.017	0.010	–	–	–	–	–	–	–
0	0.255	0.135	0.045	0.244	0.191	0.154	0.011	0.173	0.126	0.080	0.184	0.271	0.208	0.077	0.043	0.370	0.154
D1	0.735	0.975	0.720	0.822	0.888	1	0.989	0.942	0.731	0.930	0.939	0.958	1	1	1	0.889	0.795
D2	0.240	–	0.025	0.144	–	–	–	–	0.261	0.010	–	–	–	–	–	–	–
0	0.025	0.025	0.255	0.033	0.112	–	0.011	0.058	0.008	0.060	0.061	0.042	–	–	–	0.111	0.205
E2	0.415	0.500	0.580	0.278	0.551	0.375	0.500	0.683	0.458	0.590	0.408	0.708	0.750	0.720	0.979	1	1
0	0.585	0.500	0.420	0.722	0.449	0.625	0.500	0.317	0.542	0.410	0.592	0.292	0.250	0.280	0.021	–	–

Numbers in parenthesis indicate the total number of individuals in each sample. Blanks are indicated under the '0' category. Present study: AFR (African American), EUR (US European), HIS (US Hispanic), SOM (Somali), JAP (Japanese), FIN (Finns), SAA (Kola Saami), BOS (Bosnian). Published sequence data used for comparison: MAN (Mandenka),[16] BRI (British),[17] TUS (Tuscan),[18] HAV (Havik),[19] IND (Indonesian),[20] PNG (Papua New Guinean),[20] NGO (Ngobe),[21] HUE (Huetar),[22] MAP (Mapuche).[23] Due to ambiguous positions for the A region only 89 Somali considered; and for the B region only 198 African American, 197 US European, 199 US Hispanic and 89 Somali were considered. Due to non-tested positions only 98 British were considered for the A region, and only 118 Mandenka and 25 Papua New Guinean for the E region were considered.

Table 3 Diversity parameters for the mtDNA haplotype defined by SSO probes at five HVRII mtDNA regions

	N	*n*	*h*	*P*
African Americans	198	68	0.973±0.004	0.031
US Europeans	198	47	0.932±0.009	0.071
US Hispanics	199	48	0.920±0.011	0.084
Somali	88	34	0.960±0.008	0.051
Japanese	89	25	0.928±0.013	0.083
Finns	104	29	0.929±0.012	0.080
Saami	94	14	0.812±0.030	0.196
Bosnians	104	34	0.925±0.014	0.083
Mandenka[16]	118	28	0.925±0.012	0.081
British[17]	98	37	0.927±0.016	0.083
Tuscans[18]	49	25	0.950±0.016	0.070
Havik[19]	48	18	0.912±0.022	0.107
Indonesians[20]	24	11	0.891±0.044	0.146
Papua New Guineans[20]	25	10	0.867±0.048	0.168
Ngobe[21]	47	6	0.598±0.059	0.415
Huetar[22]	27	5	0.809±0.032	0.220
Mapuche[23]	39	6	0.831±0.022	0.190

N: number of individuals; n: number of different mtDNA haplotype found; h: genetic diversity with its standard deviation; P: probability of mitotype identity between two randomly chosen individuals in the same population.

which might be attributed to a founder effect during the colonisation of the continent.

Genetic Distances Between Populations

Genetic distances were calculated and their representation in a neighbour-joining tree is shown in Figure 1. The tree shows several clusters according to the major geographical areas. Within the African group, the Somali reveal the shortest distances to the European populations compared with the rest of the African samples. One interesting feature is the position of the US samples: African Americans cluster clearly with the African samples and US Europeans cluster with the European groups, with robust supports in their nodes, whereas the position of the US Hispanics is much more ambiguous. The outlier position of the Saami within the European cluster is also noteworthy.

Principal Component Analysis

Principal component (PC) analysis was performed to characterise the extent of the differences in the frequencies of each probe for the five HVRII regions and to identify the variants responsible for the differences observed. A total of 16 a priori statistically

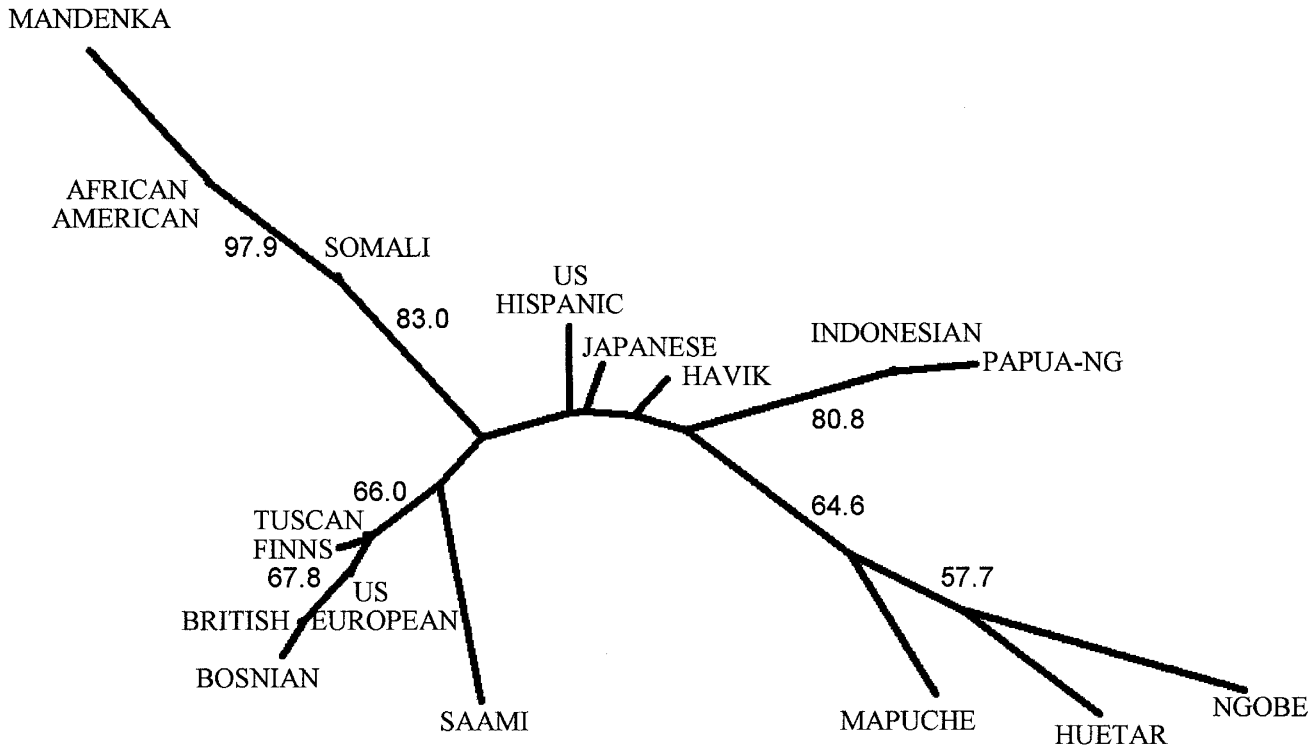

Figure 1 *Neighbour-joining tree of several worldwide populations for the mtDNA HVRII. Bootstrap supports percentages over 50% after 1000 iterations are shown in the nodes of the tree*

independent alleles for a set of 17 worldwide populations was used in the PC analysis. Figure 2 shows the two-dimensional plot of the first two PC axes, which accounts for 55.4% of the variance observed. Several clusters could be observed: African populations, European groups, Native Americans, and Asian and Pacific populations. The first principal component clearly separates the African groups from the rest of the populations studied. The singularity of the African groups is defined (with an absolute correlation higher than 0.7) by low frequencies of the probes C1 and D1, and high frequencies of B4, C2, C4, C5 and D2. In fact, the correlation of D2 with the first axis was greater than 0.95, showing the importance of position 247 of the HVRII in defining the African populations. The second principal component separates the European groups from the rest of the populations tested. European populations are defined (with an absolute correlation greater than 0.7) by high frequencies of A1 and B1, and low frequencies of A2.

In order to clarify the relationship between Finns and Saami, two of the outlier populations in Europe according to classic genetic markers,[38] PC analysis was performed using only the European samples tested and some other sequence data samples available for comparison (30 Bulgarians,[39] 200 Germans,[40] and 101 Austrians[41]). The two-dimensional plot of the first two PC axes is shown in Figure 3. These two axes account for 64.5% of the variance observed. The only population that seems to have special characteristics within the European samples is the Saami, whereas the Finns remain within the rest of the European populations.

AMOVA Analysis

In order to clarify the hierarchical apportioning of the genetic variance for the mtDNA haplotypes in the populations analysed, the AMOVA test[42] was performed. Most of the variation (93.8%) was due to differences between individuals within the populations, whereas the proportion of the variance due to differences among populations was 6.2% ($P < 0.0001$). This result is in agreement with other genetic markers such as classic markers[43,44] and microsatellites.[45] As an approximation to the hierarchical structure, the populations were split into five groups: African, comprising Mandenka, Somali and African American samples; European, comprising all the European populations and US Europeans; Native American including the US Hispanics; South-East Asia and Pacific populations, comprising Indonesians and Papua New Guineans; and the rest of the Asian populations. As previously shown, when all populations were treated as a single group, most of the variance (93.1%, $P < 0.0001$) remained within populations, whereas 3.8%. ($P < 0.0001$) was

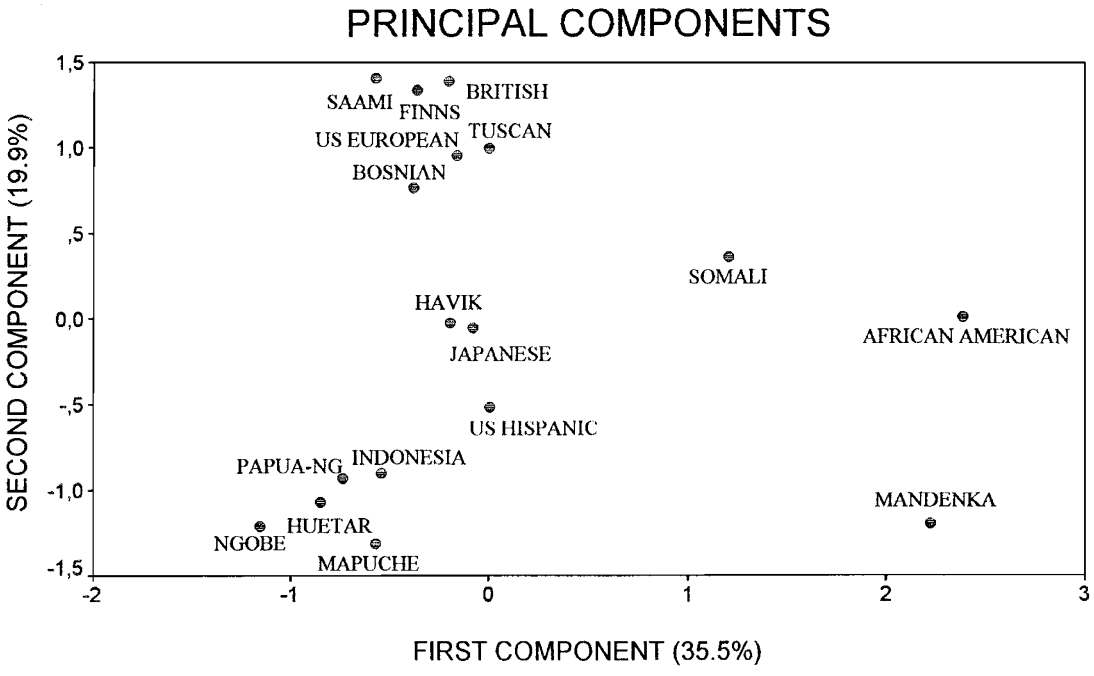

Figure 2 *Two-dimensional plot of the two first principal components (PC) axes in 17 worldwide populations for the mtDNA HVRII. The horizontal axis encompasses 35.5% of the variance observed and the vertical axis encompasses 19.9%*

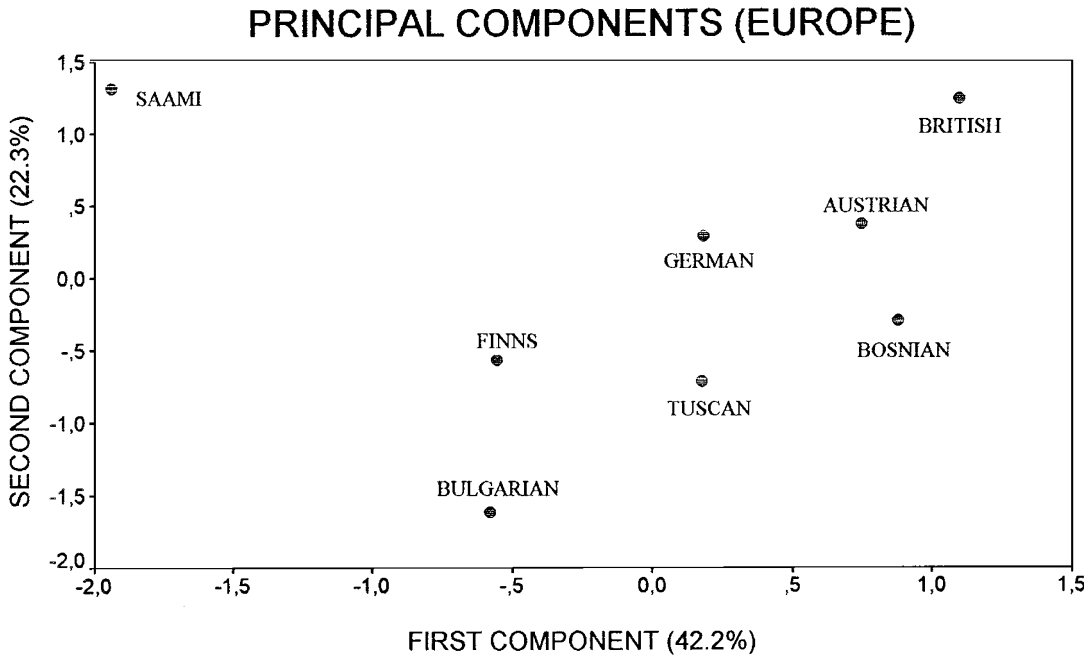

Figure 3 *Two-dimensional plot of the two first principal components (PC) axes in eight European populations for the mtDNA HVRII. The horizontal axis encompasses 42.2% of the variance observed and the vertical axis encompasses 22.3%*

due to differences among populations within the same group, and a 3.1% ($P < 0.0001$) was due to differences among the five established groups of populations.

Discussion

In the present study, seven of the 1081 individuals showed positive hybridisation for two probes of the same HVRII region. The presence of a second signal could be due to cross-hybridisation, contamination of the sample with mtDNA from a second individual, or heteroplasmy. Cross-hybridisation as the source of the second signal can be ruled out because the assay conditions have been set to avoid cross-hybridisation and other strips typed at the same time with the same reagents did not show additional signals. The HVRII region of these seven samples was sequenced, and the ambiguous positions detected by SSO probe hybridisation assay showed two different nucleotides. The HVRI region of these samples was also sequenced to reject the possibility of contamination of the samples and there was no evidence of a second contributor at any position in HVRI. Therefore, the likelihood that the second signal is due to contamination is extremely low. The most likely explanation for these results is heteroplasmy, the presence of more than one mtDNA type in the same individual, which may be more common than generally believed.[46–50] In the present study, approximately 0.7% of individuals appear to be heteroplasmic at positions in HVRII that are detected by these SSO probes.

mtDNA Diversity in East Africa

No population data are available for the mtDNA HVRII in East Africa except for the present Somali sample. It has been suggested that East African populations, mainly Ethiopians and Somali, have been affected by a strong Caucasoid influence. Cavalli-Sforza *et al*,[38] estimated from classic autosomal genetic markers that the East African genetic pool is approximately of 60% sub-Saharan and 40% Caucasoid origin. Recent data, based on mtDNA RFLPs and Y chromosome polymorphisms,[51] suggest that the Caucasoid influence in Ethiopians occurred predominantly through males, although a proportion from 5.4% to 27.0% of the mtDNA haplotypes could be attributed to Caucasoid lineages. Moreover, in a neighbour-joining tree of 15 sub-Saharan, two North African and two European populations for the mtDNA HVRI sequences[52] a Somali sample[12] appears in an intermediate position between sub-Saharan groups and the cluster composed by European and North African populations. Our Somali sample presents features that clearly locate it close to the African samples, but European features are also evident. First, high frequency of A at position 73

(probe A1) has been suggested to be an exclusively European marker;[53] secondly, the presence of CRS mtDNA haplotypes (with different lengths of Cs at position 309) are typical European mtDNA haplotypes; thirdly, high frequencies of 146T–150C–152T (probe B1) and 189A–195T–198C–200A (probe C1) occur compared with other African populations; fourthly, its closer genetic distances to the European samples compared with the rest of the African populations; finally, its position in the two-dimensional PC plot is between African and European samples. For a simple approach to measure the Caucasoid influence in East Africa, the triangle method described by Cavalli-Sforza *et al*[38] was used to compute the proportion of admixture from the genetic distance matrix. Taking the British as a representative Caucasoid sample and the Mandenka as a sub-Saharan population, the proportion m of caucasoid lineages in the Somali is $m = 0.46$. This value is similar to the estimate based on autosomal studies[38] ($m = 0.40$), and clearly higher than the estimates for the mtDNA found in Ethiopians[51] ($m = 0.05$–0.27). Our results agree with the hypothesis of a maternal influence of Caucasoid lineages in East Africa, although its contribution seems to be higher than previously reported in mtDNA studies.

Admixture in US Samples

During historical times, significant admixture between populations has occurred in the United States. Native Americans, Europeans and African populations contributed mainly to the genetic gene pool of the existing populations of the US.

The degree of admixture seems particularly high in the US Hispanic group compared with African Americans and US Europeans. Our results show that US Hispanics mtDNA lineages share some basic features with Native American populations, as suggested by previous studies,[54] such as the same frequency pattern of most of the variants and the high frequency of some uncommon mtDNA haplotypes. However, some specific variants from Africa and Europe found in US Hispanics, some specific mtDNA haplotypes from Europe and Africa found in this group, and its uncertain position in the neighbour-joining tree and PC analysis suggest that the US Hispanic sample is a very heterogeneous group and the level of admixture between several groups might have been very important in historical times.

Through all the present analyses, African American and US European samples display African and European features respectively, showing that the degree of admixture of these groups with the surrounding populations has not been enough to dilute their original mtDNA characteristics. In order to make a rough estimation of the level of admixture in these populations, the triangle distance matrix approach[38] was used. Using British and Mandenka as ancestors for the European and African lineages, respectively, the proportion m of African lineages in the present US European sample was $m = 0.08$, whilst the proportion of European lineages in African Americans was $m = 0.22$. This value is similar to previous estimates based on polymorphic protein-coding genes[55] where the European genetic component in African Americans was around 25%.

Finns and Saami within the European Genetic Landscape

Studies of allele frequencies of classical markers have shown that the Finns are an outlier in the homogeneous genetic European landscape.[38] But analyses of the mtDNA[10,11,30,56] and microsatellites[57] have shown that their maternal lineages belong clearly to the European gene pool. In the present analysis, the Finns cluster with the rest of the European populations and as shown in previous mtDNA studies no distinct features differentiate them from the rest of the European samples. On the other hand, the mtDNA lineages of the Saami, another European genetic outlier for gene frequencies studies,[38] seem to have a more controversial origin. Some mtDNA studies suggest that the Saami have a distinct history from that of other groups in Europe,[10] whereas others suggest that the Saami mtDNA gene pool appears to be basically European with a small influence of Western Central Asian populations.[30] In the present mtDNA analysis, the Saami seem to share common characteristics with the rest of the European populations, but have some distinct features in the frequencies of the sequence variants and an outlier position in the neighbour-joining tree of populations, and in the PC analysis centred on Europe. The outlier position of the Saami could be explained basically by the high incidence of the 73G–150T–309(7Cs) mtDNA haplotype that might be attributed to a strong founder effect in the maternal lineage. There are no data available for the mtDNA HVRII in Western and Central Asian populations, thus it is difficult to infer the possible maternal influence of those populations in the present Saami mtDNA gene pool.

To conclude, immobilised SSO probes are useful, rapid and simple means of detecting human mtDNA HVRII variation for population studies and forensic

identification. Sequence analysis and high-resolution RFLP analysis have been proved to be very powerful in detecting human mtDNA variaiton but these methods are not simple and rapid to perform when a large number of samples has to be typed. Although it may be worthwhile to design additional probes to minimise further the frequency of blanks, the present technique is easy to perform in population studies and large-scale forensic cases where large numbers of samples have to be typed.

Acknowledgements

We wish to thank Tiiu Anttinen, Tarja Formisto, Riitta Kauppila, Kaisa Lalu, Antti Penttilä, Helena Ranta and Antti Vuorinen for help and support in various stages of the project. Bruce Budowle, Francesc Calafell, Arpita Pandya and Ann-Christine Syvänen made fruitful comments and suggestions. We also thank the DNA laboratory staff of the Forensic Department of Medicine, University of Helsinki, especially Kirsti Höök, for technical assistance. A portion of this work was supported by NIJ grant 95-IJ-CX-0014. DC was financially supported by the Finnish Ministry of Education. We especially thank Mònica Bayés and Bru Cormand for continuous encouragement and permanent support.

References

1 Anderson S, Bankier AT, Barrell BG *et al*: Sequence and organization of the human mitochondrial genome. *Nature* 1981; **290**: 457–465.
2 Gither C, Issel-Tarver L, King MC: Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat Genet* 1992; **2**: 135–138.
3 Holland MM, Fisher DL, Mitchell LG *et al*: Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci* 1993; **38**: 542–553.
4 Gill P, Kimpton C, Aliston-Greiner R *et al*: Establishing the identity of Anna Anderson Manahan. *Nat Genet* 1995; **9**: 9–10.
5 Cann RL, Stoneking M, Wilson AC: Mitochondrial DNA and human evolution. *Nature* 1987; **325**: 31–36.
6 Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC: African populations and the evolution of human mitochondrial DNA. *Science* 1991; **253**: 1503–1507.
7 Di Rienzo A, Wilson AC: Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* 1991; **88**: 1597–1601.
8 Torroni A, Huoponen K, Francalacci P *et al*: Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996; **144**: 1835–1850.
9 Bertranpetit J, Sala J, Calafell, F, Underhill PA, Moral P, Comas D: Human mitochondrial DNA variation and the origin of the Basques. *Ann Hum Genet* 1995; **59**: 63–81.
10 Sajantila A, Lahermo P, Anttinen T *et al*: Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 1995; **5**: 42–52.
11 Richards M, Corte-Real H, Forster P *et al*: Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 1996; **59**: 185–203.
12 Watson E, Bauer K, Aman K, Weiss G, von Haeseler A, Pääbo S: mtDNA sequence diversity in Africa. *Am J Hum Genet* 1996; **59**: 437–444.
13 Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA: Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 1991; **48**: 370–382.
14 Chee M, Yang R, Hubbell E *et al*: Accessing genetic information with high-density DNA arrays. *Science* 1996; **274**: 610–614.
15 Sambroock J, Fritsch EF, Maniatis T: *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1989.
16 Graven L, Passarino G, Semino O *et al*: Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 1995; **12**: 334–345.
17 Piercy R, Sullivan KM, Benson N, Gill P: The application of the mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 1993; **106**: 85–90.
18 Francalacci P, Bertranpetit J, Calafell F, Underhill PA: Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthrop* 1996; **100**: 443–460.
19 Mountain JL, Hebert JM, Bhattacharyya S *et al*: Demographic history of India and mtDNA sequence diversity. *Am J Hum Genet* 1995; **56**: 979–992.
20 Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M: Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 1995; **12**: 604–625.
21 Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guinonneau-Sinclair F: Reduced mtDNA diversity in the Ngobe Amerinds of Panama. *Genetics* 1995; **140**: 275–283.
22 Santos M, Ward RH, Barrantes R: mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. *Hum Biol* 1993; **66**: 693–977.
23 Ginther C, Corach D, Penacino GA *et al*: Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. *EXS* 1993; **67**: 211–219.
24 Nei M: *Molecular Evolutionary Genetics*. Columbia University Press: New York, 1987.
25 Reynolds J, Weir BS, Cockerman CC: Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics* 1983; **105**: 767–779.
26 Saitou N, Nei M: The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406–425.
27 Felsenstein J: PHYLIP- Phylogeny Inference Package. *Cladistics* 1989; **5**: 164–166.

28 Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **35**: 785–791.
29 Schneider S, Kueffer JM, Roessli D, Excoffier L: *Arlequin (ver. 1.0): a Software Environment for the Analysis of Population Genetics Data*. Genetics and Biometry Laboratory, University of Geneva: Switzerland, 1996.
30 Torroni A, Bandelt HJ, D'Urbano L *et al*: mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 1998; **62**: 1137–1152.
31 Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL: High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 1994; **368**: 455–457.
32 Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J: Microsatellite variation and the differentiation of modern humans. *Hum Genet* 1997; **99**: 1–7.
33 Jorde LB, Rogers AR, Bamshad MJ *et al*: Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA* 1997; **94**: 3100–3103.
34 Calafell F, Shuter A, Speed WC, Kidd JR, Kidd KK: Short tandem repeat polymorphism in humans. *Eur J Hum Genet* 1998; **6**: 38–49.
35 Armour JA, Anttinen T, May CA *et al*: Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet* 1996; **13**: 154–160.
36 Tishkoff SA, Dietzch E, Speed W *et al*: Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 1996; **271**: 1380–1387.
37 Knight A, Batzer MA, Stoneking M *et al*: DNA sequences of the Alu elements indicate a recent replacement of the human autosomal genetic complement. *Proc Natl Acad Sci USA* 1996; **93**: 4360–4364.
38 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton University Press; Princeton, 1994.
39 Calafell F, Underhill PA, Tolun A, Angelicheva D, Kalaydjieva L: From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 1996; **60**: 35–49.
40 Lutz S, Weisser HJ, Heizmann J, Pollak S: Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 1998; **111**: 67–77.
41 Parson W, Parsons TJ, Scheithauer R, Holland MM: Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* 1998; **111**: 124–132.
42 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
43 Lewontin RC: The apportionment of human diversity. *Evol Biol* 1972; **6**: 381–398.
44 Ryman N, Chakraborty R, Nei M: Differences in the relative distribution of human gene diversity between electrophoresis and red and white cell antigen loci. *Hum Hered* 1983; **33**: 93–102.
45 Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL: An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* 1997; **94**: 4516–4519.
46 Gill P, Ivanov PL, Kimpton C *et al*: Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 1994; **6**: 130–135.
47 Comas D, Bertranpetit J, Pääbo S: Heteroplasmy in the control region of human mitochondrial DNA. *Genome Res* 1995; **5**: 89–90.
48 Bendall KE, Macaulay V, Baker JR, Sykes B: Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet* 1996; **59**: 1276–1287.
49 Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ: Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 1996; **12**: 417–420.
50 Parsons TJ, Muniec DS, Sullivan K: A high observed substition rate in the human mitochondrial DNA control region. *Nat Genet* 1997; **15**: 363–368.
51 Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti S: Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 1998; **62**: 420–434.
52 Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Avade A, Bertranpetit J: A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* 1997; **61**: 507–518.
53 Wilkinson-Herbots HM, Richards MB, Forster P, Sykes BC: Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. *Ann Hum Genet* 1996; **60**: 499–508.
54 Merriwether DA, Huston S, Iyengar S *et al*: Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* 1997; **102**: 153–159.
55 Chakraborty R, Kamboh MI, Nwanko M, Ferrell RE: Caucasian genes in American blacks: new data. *Am J Hum Genet* 1992; **50**: 145–155.
56 Lahermo P, Sajantila A, Sistonen P *et al*: The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 1996; **58**: 1309–1322.
57 Sajantila A, Pääbo S: Language replacement in Scandinavia. *Nat Genet* 1995; **11**: 359–360.