J Hum Genet (2006) 51:429–439
DOI 10.1007/s10038-006-0381-x

ORIGINAL ARTICLE

Miguel A. Alfonso-Sánchez
Cristina Martínez-Bouzas · Azucena Castro
Jose A. Peña · Isabel Fernández-Fernández
Rene J. Herrera · Marian M. de Pancorbo

Sequence polymorphisms of the mtDNA control region in a human isolate: the Georgians from Swanetia

Received: 5 October 2005 / Accepted: 20 January 2006 / Published online: 1 April 2006

Abstract In this work, we analyzed the sequence diversity of the mtDNA control region (HVI and HVII) in a sample of 48 individuals from Swanetia (Georgia), using direct fluorescent-based sequencing methods. We identified 43 different mtDNA haplotypes resulting from 78 polymorphic sites (46 in HVI and 32 in HVII). Most of the variable positions identified in both HVI and HVII were transitions (82.6 and 71.9%, respectively). The frequency of length heteroplasmy in the homopolymeric C-stretch regions was the same for both segments (10.4%). The sequence diversity increased markedly when both hypervariable regions were analyzed jointly (HVI: 0.985, HVII: 0.975, HVI + HVII: 0.994). Accordingly, the probability of two randomly selected sequences matching (random match probability, RMP) decreased from 3.4% (HVI) to 2.6% (HVI + HVII), despite which the RMP values in Georgians remained higher than estimated in most Europeans. This suggests that the variability of maternal lineages tends to be lower in traditional human isolates and, therefore, the potential of discrimination of mtDNA in forensic analysis is more limited in this type of population. The incorporation of HVII data also contributed to the refinement of results regarding the genetic relationships among the samples included in the analyses, which stress the importance of considering HVII in both population and forensic genetics.

Keywords mtDNA haplogroups · Sequence diversity · D-loop · HVI · HVII · Caucasus · Georgian population · Genetic relationships

M.A. Alfonso-Sánchez and C. Martínez-Bouzas contributed equally to this work and are listed alphabetically

M.A. Alfonso-Sánchez · J.A. Peña
Departamento de Genética y Antropología Física,
Facultad de Ciencia y Tecnología, Universidad del País Vasco,
Bilbao, Bizkaia, Spain

C. Martínez-Bouzas · M.M. de Pancorbo (✉)
Servicio de Genómica: Banco de ADN y Departamento de Zoología y Biología Celular Animal, Facultad de Farmacia,
Universidad del País Vasco, 48940 Leioa, Bizkaia, Spain
E-mail: marian.mdepancorbo@ehu.es
Tel.: +34-94-5013045
Fax: +34-94-5013014

A. Castro · I. Fernández-Fernández
DataGene, Edificio BEAZ, Sangroniz 6, 48150 Sondika,
Bizkaia, Spain

R.J. Herrera
Molecular Biology and Human Diversity Laboratory,
Department of Biological Sciences,
Florida International University,
Miami, FL 33199, USA

Introduction

The Caucasus is a region located in the border between Europe and western Asia (between the Black and Caspian Seas). This region exhibits a high degree of linguistic diversity, with four major language families represented (North Caucasian, South Caucasian, Indo-European, and Altaic), which are spoken by around 50 autochthonous groups. A notable geographic barrier, the Caucasus Mountains, divides the northern and southern sub-regions of the Caucasus. Georgians are the main ethnic group in Swanetia, a province of the Republic of Georgia on the southern slopes of the Central Caucasus. Its rough, mountainous terrain has helped make Swanetia a very isolated region, and this has been decisive in the preservation of its cultural and linguistic patrimony, and therefore of its genetic background. The local language (Svan) belongs to the Kartvelian or South Caucasian language family and is non Indo-European in origin (Gamkrelidze and Ivanov 1990; Renfrew 1991).

The D-loop region of human mitochondrial DNA (mtDNA), which is approximately 1,150 base pairs (bp) long, is located between the mitochondrial $tRNA^{Pro}$ and

$tRNA^{Phe}$ genes. This is the most variable region in the mitochondrial genome, and the most polymorphic nucleotide sites within this loop are concentrated in two 'hypervariable segments', HVI (positions 16024–16365) and HVII (positions 69–340) (Aquadro and Greenberg 1983).

Mitochondrial DNA is maternally inherited and evolves rapidly. The average number of base pair differences between two human mitochondrial genomes is estimated to be between 9.5 and 66 (Zeviani et al. 1998). The high mutation rate has resulted in the accumulation of a wide range of neutral, population-specific base substitutions in mtDNA. These have accumulated sequentially along radiating maternal lineages that have diverged approximately on the same time-scale as human populations have colonized different geographical regions of the world (Torroni and Wallace 1994; Wallace 1995). In the last two decades, genetic data derived from mtDNA studies have helped greatly in elucidating human evolution, estimating a time-scale for events in human prehistory, and detecting past demographic movements (Vigilant et al. 1991; Schurr and Wallace 2002; Tajima et al. 2004; Starikovskaya et al. 2005). The study of female lineages has provided authentic evidence on the African origin and subsequent dispersal of our species (Cann et al. 1987).

On the other hand, dysfunction in mitochondrial processes has been related to several pathologies. Mutations in mtDNA are now recognized as major contributors to human pathologies and possibly to normal aging (reviewed in Wallace 1997, 1999; Zeviani et al. 1998). A large number of rearrangements and point mutations in protein-coding and tRNA genes have been identified in patients with mitochondrial disorders (Holt et al. 1988; Reardon et al. 1992; Procaccio and Wallace 2004; Sudo et al. 2004).

Finally, analysis of mtDNA is a very useful forensic tool (Parson et al. 1998). When forensic cases arise where there is insufficient biological material for nuclear DNA typing, mtDNA analysis can provide valuable supplemental information, even from such limited samples as 0.5-cm-long hair fragments or single teeth. Because of its usefulness when limited biological material is available, and due to its unique pattern of maternal inheritance, mtDNA plays a key role in personal identity testing (Budowle et al. 2003).

In recent years, many databases of the mtDNA control region have been published to permit mtDNA forensic casework in particular countries or geographic regions (Parson et al. 1998; Rousselet and Mangin 1998; Budowle et al. 1999, 2002; Pfeiffer et al. 1999; Imaizumi et al. 2002; Vanecek et al. 2004; Zupanic Pajnic et al. 2004). However, the bulk of mtDNA sequence data published to date have been centered in the HVI segment. The strength of the mtDNA evidence that can be reported is very often limited by the lack of relevant database information. For this reason, it is important that mtDNA sequence databases continue to be generated to extend mtDNA typing capability to additional populations and to increase the size of existing databases (Imaizumi et al. 2002).

The genetic characterization of so-called 'isolated' human populations is relevant to accomplishing several scientific goals such as those related to disease mapping, human demographic history and forensic identification (Arcos-Burgos and Muenke 2002; Pérez-Miranda et al. 2004, 2005). Previous studies addressing the genetic characterization of mtDNA polymorphism in Georgians have been limited to the HVI segment (Comas et al. 2000; Nasidze and Stoneking 2001). The aim of the present study was to analyze HVI and HVII sequences of the mtDNA control region in a sample of autochthonous individuals from Swanetia (Georgia) using direct fluorescent-based sequencing. Our findings on mtDNA diversity in Georgia are then viewed in a comparative context, using previously published data on mtDNA haplogroups for other human populations worldwide. With this integrative approach, we seek to assess population affinities and phylogenetic relationships of the Georgian group in a broader geographical context. The genetic information presented herein may be highly useful in both evolutionary genetic studies and forensic analyzes, especially data on the polymorphism of hypervariable region HVII, which is analyzed here for the first time in Georgians.

Materials and methods

Genomic DNA was extracted from peripheral blood using a standard phenol-chloroform procedure (Sambrook et al. 1989). Blood samples were collected by venipuncture from 48 unrelated healthy men living in the Caucasus region (Georgia). Only native people were included in the sample; Georgian ancestry (based on surnames and birthplaces) was ascertained for three generations back in order to define autochthony for each donor. The autochthonous individuals were carefully selected to avoid the bias caused by sample heterogeneity. Specifically, the Georgian population sample was collected in the region of Swanetia to avoid any admixture with Turkish and Russian lineages, which are most frequently found in the lowlands of the country. Ethical guidelines were adhered to as stipulated by each of the institutions involved in the study. All blood donors gave their informed consent prior to their inclusion in the study sample.

The HVI and HVII segments of the mtDNA D-loop region were amplified by polymerase chain reaction (PCR). The primers used to amplify HVI and HVII segments were L15996/H16401 and L29/H408, respectively, as described by Vigilant et al. (1991). PCRs were performed in a Thermal Cycler GeneAmp PCR System 9600 (PE Applied Biosystems, Foster City, CA) by hot start at 85°C and 30 cycles of 94°C 45 s, 66°C 60 s and 72°C 60 s, and a final extension step at 72°C for 10 min. PCR products were sequenced with the dRhodamine Terminator Cycle Sequencing Ready Reaction kit on an

ABI Prism 310 Genetic Analyzer (PE Applied Biosystems).

The mtDNA sequences reported in this paper have been deposited in the DDBJ/EMBL/GenBank nucleotide database (accession numbers AM181177–AM181225).

Statistical analysis

The individual sequences were aligned and compared to the revised Cambridge Reference Sequence for human mtDNA, rCRS (Andrews et al. 1999), by means of the Clustal W program package (Version 1.5). The rCRS is a modified version of the original CRS of Anderson et al. (1981), and has 18 annotated nucleotides. We chose the Multiple Sequence Alignment option (Higgins et al. 1992) to determine the specific nucleotide substitutions. The HVI segment was examined for positions 16033–16362 (according to rCRS), while analysis of the HVII segment included positions 66–370. The number of constant sites was calculated using Puzzle (Version 4.02) software (Strimmer and Von Haeseler 1995). The genetic variability of the mtDNA sequences in the Georgian population was measured by means of three parameters: nucleotide diversity (Nd), sequence diversity, and haplogroup diversity (Nei 1987). These basic diversity parameters were computed using the Arlequin program, version 2.000 (Schneider et al. 2000). The random match probability (RMP) was estimated as the solution to the equation

$$RMP = \sum_{i=1}^{h} p_I^2,$$

where p_I is the sample frequency of the I-th haplotype, and h is the number of distinct haplotypes observed in the sample (Jones 1972).

For interpopulation comparison purposes, published data on both HVI haplogroups and sequence variations in the hypervariable regions HVI and HVII from many other human groups worldwide were compiled. Thus, mtDNA haplogroup data in Georgians were compared to the following 23 populations: Andalusian, Catalonian, Portuguese, Algerian (Côrte-Real et al. 1996), British (Piercy et al. 1993), North Portuguese (Pereira et al. 1999), Tuscan (Francalacci et al. 1996), South German (Lutz et al. 1998), North-West German (Pfeiffer et al. 1999), Austrian, Spanish, Egyptian, Japanese, sub-Saharan African, African-Americans (USA), European Americans (USA), Hispanic, Navajo (Budowle et al. 1999), Basque (unpublished data), Kung (Vigilant et al. 1991), French (pooled from Rousselet and Mangin 1998; Budowle et al. 1999), Korean (pooled from Pfeiffer et al. 1998; Budowle et al. 1999), and Taiwanese (pooled from Melton et al. 1995; Budowle et al. 1999).

Genetic distances between pairs of populations were calculated using the mismatch-intermatch means of the pairwise difference distribution. Then, $D = d_{ij} - (d_{ii} + d_{jj})/2$, where d_{ij} is the intermatch mean (i.e., the mean number of nucleotide differences) between populations I and J, and d_{ii} and d_{jj} are the mismatch values within populations I and J, respectively (Rao 1982; Nei 1987). In order to depict the mismatch–intermatch distance matrix in a two-dimensional genetic map, nonmetric multidimensional scaling (MDS) analysis (Kruskal 1964) was performed using the SPSS (version 13.0; SPSS, Chicago, IL) statistical package. Furthermore, phylogenetic trees based on the neighbor-joining (NJ) method (Saitou and Nei 1987) were generated from the resultant distance matrix, using programs in Phylip version 3.2 (Felsenstein 1989). The reliability of the dendrogram (robustness of the branch nodes) was estimated by bootstrap resampling methods (Felsenstein 1985). Finally, population trees based solely on the sequence variations for HVI, on the one hand, and on the sequence variations identified when a joint analysis of HVI + HVII was carried out on the other hand, were compared. In this latter analysis, only 14 populations (including Georgia) could be considered, due to the paucity of HVII data. Populations included were those compiled in Budowle et al. (1999) and Basques (unpublished data).

Results and discussion

Sequence variability

Sequencing data on the hypervariable region HVI, corresponding to nucleotide positions 16033–16362 in the revised Cambridge Reference Sequence for human mtDNA (rCRS), are shown in Table 1. Sequence comparison between the Georgian individuals ($n = 48$) and the rCRS led to the identification of 37 different mtDNA haplotypes in HVI, defined by 46 polymorphic positions. Of them, 28 mtDNA types (75.7%) were unique, seven haplotypes were found twice, and two haplotypes were shared by three individuals. Only one mtDNA sequence showed no difference with regard to the rCRS (code GS-78, in Table 1).

The number of constant positions in HVI was 284, which represents 86.1% of the mtDNA fragment analyzed (330 bp in length). Most of the variable positions identified in the HVI segment were transitions, specifically 38, which is equivalent to 82.6% of the polymorphic sites. Furthermore, seven transversions and one insertion were also detected (see also Table 3).

The most frequent nucleotide substitutions were the transitions 16311 T > C in 16 individuals (33.3%) and 16223 C > T in 10 individuals (20.8%). Other common transitions observed were 16224 T > C and 16294 C > T, which appeared in eight individuals each (16.7%), whereas transitions 16126 T > C and 16189 T > C were both found in 14.6% of cases (seven individuals). Likewise, the most frequent transversion in the Georgian sample analyzed was 16183 A > C, which was identified in four individuals. Other transversions detected were 16085 C > G, 16146 A > C, 16071 A > C, 16214 C > A,

Table 1 Mitochondrial DNA diversity of the hypervariable region HVI in a Georgian population ($n = 48$). Polymorphic sites are given in comparison to the revised Cambridge reference sequence (Andrews et al. 1999), rCRS (bold type at the top)

No.	16071	16085	16093	16111	16126	16129	16146	16148	16163	16167	16171	16183	16184	16186	16189	16192	16193.1	16209	16213	16214	16217	16223	16224	16245	16246	16249	16260	16261	16263	16273	16278	16288	16290	16292	16294	16295	16296	16297	16311	16316	16325	16327	16343	16356	16359	16362
	C	**C**	**T**	**C**	**T**	**G**	**A**	**C**	**A**	**C**	**A**	**A**	**C**	**C**	**T**	**C**	**C**	**T**	**G**	**C**	**T**	**C**	**T**	**C**	**A**	**T**	**C**	**C**	**T**	**G**	**C**	**T**	**C**	**C**	**C**	**C**	**C**	**T**	**T**	**A**	**T**	**C**	**A**	**T**	**T**	**T**
GS-1	.	.	.	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.
GS-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.
GS-3	.	.	.	.	.	A	C	.	.	.	.	C	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-4	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.
GS-6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-10	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-14	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-15	.	.	.	.	C	A	.	.	G	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-16	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-21	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
GS-27	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
GS-28	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
GS-30	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-31	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
GS-33	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
GS-34	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-36	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
GS-37	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.
GS-38	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-39	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-40	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-42	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
GS-44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.
GS-45	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.	C	.	.	C	.	.	.	A	.	C	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
GS-46	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.
GS-47	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-48	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-56	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-58	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	C	.
GS-59	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-60	.	.	.	.	.	A	.	.	.	.	.	C	.	.	C	.	C	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
GS-61	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.
GS-64	.	.	.	.	C	.	.	.	G	.	.	.	.	T	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-73	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	C
GS-74	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-75	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-76	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
GS-77	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-78	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GS-79	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
GS-80	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.

16246 A > T, and 16359 T > G, each of them observed in one individual. In addition, an insertion in the region delimited by 16190–16193 bp (four individuals) was found, denoted as 16193.1C. There was length heteroplasmy in the homopolymeric C-stretch region (located between nucleotide positions 16184 and 16193 of HVI) in 10.4% of the sequences examined.

Analysis of mtDNA hypervariable region HVII included 305 bp (positions 66–370, Table 2). Sequence comparisons with rCRS permitted the identification of 31 different mtDNA variants defined by 32 polymorphic sites. Of the whole set of mitochondrial haplotypes, 22 DNA sequences (71.0%) were unique, five haplotypes were shared by three individuals, three haplotypes were found twice, and the most common haplotype was identified in five individuals.

A more in-depth analysis of mtDNA hypervariable region HVII revealed 273 constant positions (89.5%) in the sequences examined. Of the total number of nucleotide changes detected in this study, 23 (71.9%) were

Table 2 Mitochondrial DNA diversity of the hypervariable region HVII in a Georgian population ($n = 48$). Polymorphic sites are given in comparison to the rCRS (Andrews et al. 1999); bold type at the top

No.	73 **A**	81 **G**	93 **A**	114 **C**	119 **T**	142 **G**	146 **T**	150 **C**	152 **T**	153 **A**	186 **C**	189 **A**	194 **C**	195 **T**	196 **T**	196.1 **.**	197 **A**	199 **T**	204 **T**	207 **G**	250 **T**	257 **A**	263 **A**	279 **T**	283 **A**	285 **C**	295 **C**	298 **C**	309.1 **.**	309.2 **.**	309.3 **.**	315.1 **.**
GS-1	G	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	C	.	.	C
GS-2	G	.	.	.	.	.	.	.	.	.	.	G	T	C	.	.	.	.	C	A	.	.	G	.	.	.	.	.	C	.	.	C
GS-3	G	.	.	.	.	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	G	G	.	.	.	.	.	C	.	.	C
GS-4	G	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-6	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-7	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-8	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-10	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	C	.	.	.	.	C	.	.	C
GS-11	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-14	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	G	.	.	.	.	.	C	.	.	C
GS-15	G	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-16	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-21	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-24	G	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-27	G	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	C	.	.	C
GS-28	G	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	C	C	C
GS-30	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	G	.	.	.	.	.	.	.	.	C
GS-31	G	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-33	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-34	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	A	.	.	.	C
GS-36	G	.	.	.	.	.	C	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	.	.	.	.	C	.	.	C
GS-37	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-38	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-39	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-40	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-41	G	.	.	.	.	.	C	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-42	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-44	G	.	.	.	.	.	.	.	.	.	.	G	T	C	.	.	.	.	C	A	.	.	G	.	.	.	.	.	C	.	.	C
GS-45	G	.	.	.	.	.	C	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	C
GS-46	G	.	.	.	.	.	.	.	.	.	.	G	T	C	.	.	.	.	C	A	.	.	G	.	.	.	.	.	C	.	.	C
GS-47	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-48	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-51	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	C	.	C
GS-56	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-58	G	.	.	.	C	A	.	.	.	.	.	G	A	.	C	C	.	.	C	A	.	.	G	.	.	.	.	.	C	.	.	C
GS-59	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-60	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	C	C	C
GS-61	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-64	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-73	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A	.	.	.	.	C
GS-74	G	A	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.	C
GS-75	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	C
GS-76	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	A	.	C	.	.	C
GS-77	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-78	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C
GS-79	G	.	.	.	.	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	**C**
GS-80	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	C

transitions, 4 (12.5%) were transversions and 5 (15.6%) were insertions (see also Table 3). Interestingly, the transition 263 A > G was present in all the sequences analyzed (100%). Apart from this relevant nucleotidic change, the most frequent transitions in the HVII segment were 73 A > G and 195 T > C, which were found in 42 (87.5%) and 21 (43.8%) individuals, respectively. The four transversions observed in HVII were located at nucleotide positions 186 C > A, 194 C > A, 295 C > A (in two individuals) and 298 C > A.

There is also a poly-C tract between positions 303 and 315 in the HVII segment, which is interrupted at position 310 by a thymine according to the rCRS. In this study, we identified some sequences that presented between 1 and 3 insertions in the region of nucleotides 303–309. Specifically, we observed 32 cases (66.6%) showing one insertion (309.1C). A second C insertion (309.2C) was detected in three individuals (6.3%), whereas a third insertion (309.3C) was found in two sequences (4.2%). On the other hand, insertion 315.1C

Table 3 Sequence polymorphism of mitochondrial DNA hypervariable regions HVI and HVII in a sample of a Georgian population. *CP* Constant positions, *Ps* polymorphic sites, *Ts* transitions, *Tv* transversions, *Ins* insertions, *Nd* nucleotide diversity, *Sd* sequence diversity, *MNM* mean number of mismatches, *RMP* random match probability

	CP	Ps	Ts	Tv	Ins	Nd [a]	Sd	MNM	RMP
HVI	284 (86.1%)	46	38	7	1	0.017 ± 0.009	0.985 ± 0.007	5.56	0.034
HVII	273 (89.5%)	32	23	4	5	0.011 ± 0.006	0.975 ± 0.010	3.84	0.045
HVI-HVII	557 (87.7%)	78	61	11	6	0.014 ± 0.008	0.994 ± 0.006	9.47	0.026

[a]For the calculation of Nd, transition 263 (G) and insertion 315.1 (C) were not considered since they appeared in 100% of the sequences

was present in all individuals (100%) in the sample. Length heteroplasmy of the homopolymeric poly-C tract was observed in five sequences (10.4%). It is worth highlighting that the most frequent haplotypes in the Georgian population, i.e., 263 (G), 309.1 (C) and 315.1 (C) have been also reported as the most frequent in other European populations. This fact supports the hypothesis that these haplotypes might represent an ancient European common mtDNA sequence (Zupanic Pajnic et al. 2004).

The joint analysis of both segments (HVI and HVII) yielded 43 different haplotypic combinations. The number of constant positions was 557 (87.7%), whereas the number of polymorphic sites was 78 (see Table 3), with wide predominance of transitions (78.2%) over transversions (14.1%) and insertions (7.7%).

Diversity parameters

Some parameters characterizing within-population diversity of the mtDNA sequence, such as nucleotide diversity (Nd) and sequence diversity (Sd), are listed in Table 3. The Nd was higher in the hypervariable region HVI (0.017) than in HVII (0.011). The Nd estimated for HVI was slightly above the top edge of the range (0.0081–0.0155) reported in previous studies including several European populations (Salas et al. 1998; Zupanic Pajnic et al. 2004). This diversity parameter showed values similar to those observed in West Asian samples and in populations from the Anatolian Peninsula (e.g., Turkey), where figures of around 0.016 have been reported (Mergen et al. 2004). Interestingly, the Nd computed for HVII falls far below the variation range (0.0137–0.0187) reported for European populations in recent works (Salas et al. 2000; Zupanic Pajnic et al. 2004).

Consistent with the results described above for Nd, Sd was higher in the HVI region (0.985) than in HVII (0.975). The Sd value calculated for HVI is comparatively higher than levels reported in previous works for other European populations such as British (0.973) (Piercy et al. 1993) and Spaniards (0.939) (Côrte-Real et al. 1996). The Sd value obtained for the population examined (Georgians from Swanetia) was also slightly higher than that estimated in samples from other regions of Georgia, where values between 0.964 and 0.971 have been found (Comas et al. 2000; Nasidze and Stoneking 2001). Sequence diversity in Georgians was, however, very similar to that obtained in populations located in the Anatolian Peninsula, such as Turks (0.988) and Kurds (0.985) (Comas et al. 2000).

As expected, the Sd value increased up to 0.994 when both hypervariable regions (HVI and HVII) were considered jointly. This figure falls into the Sd range estimated by Budowle et al. (1999) for several worldwide populations (0.990–0.998), although it remains below those calculated in several European populations such as those of Austrian, French and USA-Europid samples (0.996). It must be emphasized that Sd was relatively high in Georgians solely in the case of the HVI segment. When we computed Sd including HVII, the global value (HVI + HVII) of Sd was comparatively lower than the estimated Sd in many other European populations analyzed to date (see Budowle et al. 1999; Zupanic Pajnic et al. 2004). These results are in agreement with previously discussed results regarding Nd.

Finally, we estimated the probability of two randomly selected sequences matching (RMP), which is inversely proportional to the power of discrimination of a given genetic marker. The RMP value was lower in hypervariable region HVI (3.4%) than in HVII (4.5%) (Table 3). In a logical reflection of the data on diversity parameters (Nd and Sd), the RMP value estimated for the Georgian population considering HVI and HVII (2.6%) was clearly above the variation range estimated for the Spanish regions of Aragon (1.3%) and Madrid (1.5%), and for other European populations such as those of Switzerland (0.9%), France (2.1%), Germany (0.9%), Austria and the United Kingdom (1.3%) (Dimo-Simonin et al. 2000). These findings suggest that the variability of maternal lineages tends to be lower in traditional human isolates, most likely because of extremely restricted gene flow and high inbreeding levels over long periods. Thus, the power of discrimination of the mtDNA in forensic identity testing is thought to be more limited in this type of population. This hypothesis is clearly fulfilled in another interesting 'human isolate' from Europe: the Basques (work in progress). The RMP estimated for HVI in the native Basque population stands out as having a notably high value (10.1%). The RMP for HVII, although lower than that of HVI, is also comparatively high (4.7%). Likewise, the RMP value calculated taking both HVI and HVII into account is practically twice as high (1.96:1.00) in Basques (5.1%) than in Georgians (2.6%), and these differences are even greater when compared to other Europeans.

MtDNA haplogroups and genetic relationships with other populations

MtDNA haplogroup-based analyses are very important in unveiling genetic relationships among different human groups. Haplogroup frequencies in the Georgian population are depicted in Fig. 1. MtDNA haplogroups were inferred from control region sequences (HVI + position 73 of HVII) according to the classification proposed by Richards et al. (1998). Within the set of mtDNA haplogroups identified in the study sample, only four showed frequencies above 10%. These were haplogroups U (25.0%), T (10.4%), K and H (12.5% each). As can be noted, haplogroup U was the more frequent in Georgians, in contrast with most European populations analyzed to date, where haplogroup H stands out as the most abundant. Another peculiarity of the sample analyzed herein is the absence of haplogroup U5. This haplogroup was identified in the Georgian collection examined by Comas et al. (2000). The haplogroup diversity estimate yielded a value of Hd = 0.891.

In order to assess the genetic relationships of Georgians with other populations based on mtDNA diversity, we compiled haplogroup data from previously published papers (see Materials and methods). First, we computed the mismatch-intermatch distances (Rao 1982; Nei 1987) between all pairs of populations (data not shown). A nonmetric MDS analysis was then used to represent the data generated by the distance matrix in a two-dimensional space.

Figure 2 illustrates the MDS plot generated from the distance matrix. Populations were grouped according to geography, and this topology was highly robust from the statistical viewpoint, bearing in mind that the two-dimensional representation accounted for 98.6% of the total variance. A major division can be observed along Dimension I between the Europid group (European and North African populations), which concentrated in the negative segment, and the Asian and sub-Saharan African samples (grouped in the positive portion of Dimension I). North African populations (Algeria, Egypt) occupied the most remote positions within the Europid cluster. Overall, the samples related to the African continent (Algerians, Egyptians, sub-Saharan Africans, Afro-Americans and Kung) tended to concentrate at the negative edge of Dimension II. Of these populations, the Kung sample seems to be the most differentiated genetically, as can be inferred from its remote position from the centroid of the distribution.

Worthy of attention is also the central position of the Georgian sample, which plotted close to the centroid, in a practically intermediate position between Europeans and Asians. Two plausible scenarios have been used in some previous studies to explain this interesting pattern: one scenario corresponds to a strictly phylogenetic interpretation of population trees and topogenetic maps, while other interprets these data strictly in terms of migration. The first scenario points out that Caucasus groups are derived from Near Eastern groups and are immediately ancestral to Europeans (Nasidze and Stoneking 2001). The second scenario argues that Caucasus groups are admixed and have experienced gene flow from both Europe and Asia. In this respect, although findings of some previous investigations have revealed a low degree of genetic divergence among Europid populations according to levels of differentiation observed in the sequence of the mtDNA control region, some authors indicate that there is a certain geographic patterning of the mtDNA variation, probably denoting a stepping-stone position of the Anatolian Peninsula

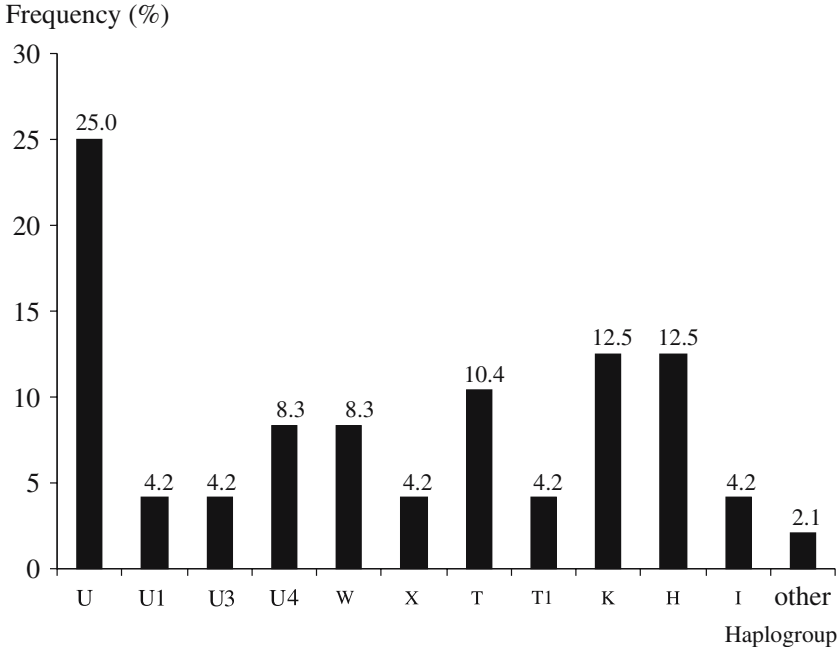

Fig. 1 Haplogroup frequencies of mitochondrial DNA (mtDNA) in a Georgian population (Swanetia). Haplogroups were inferred from control region sequences (HVI + position 73 of HVII) according to the classification proposed by Richards et al. (1998). Frequency values (%) are displayed. Haplogroups not represented in the histogram were not found

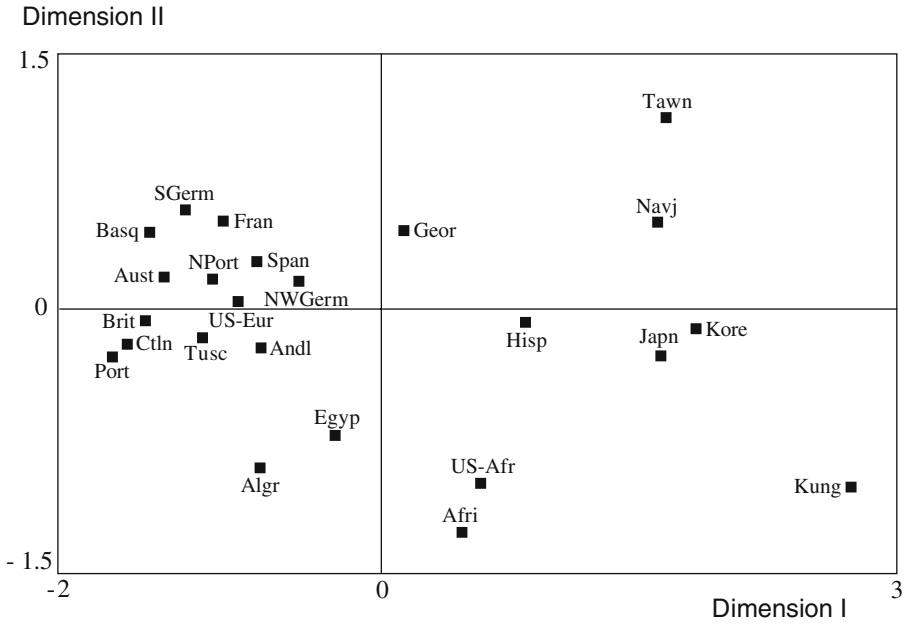

Fig. 2 Nonmetric multidimensional scaling (MDS) illustrating the genetic relationships derived from mtDNA haplogroup (according to Richards et al. 1998) data among Georgians and 23 other worldwide populations. The MDS plot was constructed based on the mismatch–intermatch means of the pairwise difference distribution (Rao 1982; Nei 1987). The total variance accounted for in the two-dimensional representation is 98.6%. *Geor* Georgia, *Andl* Andalusia, *Ctln* Catalonia, *Basq* Basques, *Span* Spain, *NPort* North Portugal, *Port* Portugal, *Tusc* Tuscany, *Fran* France, *SGerm* South Germany, *NWGerm* North-West Germany, *Aust* Austria, *Brit* British, *US-Eur* United States' Europids, *Egyp* Egypt, *Algr* Algeria, *Afri* sub-Saharan African, *US-Afri* Afro-Americans, *Kung* Kung, *Tawn* Taiwan, *Japn* Japan, *Kore* Korea, *Hisp* Hispanics, *Navj* Navajo

(Asian Turkey) and the Caucasus region (including Georgia) between the Middle East and Europe (Calafell et al. 1996; Comas et al. 1996). This explanation is based on the hypothesis of the replacement of Neanderthals in Europe by the arrival of anatomically modern humans during the cultural expansion of the Upper Paleolithic age, between 50,000 and 100,000 years ago.

To verify whether information on HVII sequence polymorphism could contribute to improving a phylogenetic classification based on mtDNA data, population trees were generated from the mismatch-intermatch distance matrix using the NJ method (Fig. 3). NJ trees were obtained using two different data sets: (1) mtDNA sequence variation solely in the hypervariable region HVI (Fig. 3a), and (2) mtDNA sequence variation for hypervariable regions HVI + HVII considered jointly (Fig. 3b). The thorough comparison of the population trees reveals an improvement in the phylogenetic classification when data on HVI and HVII are considered simultaneously. This improvement is reflected both in a substantial increase of the average bootstrap value and in a more adjusted distribution of the human groups, as expected bearing in mind the geographical and ethnohistorical characteristics of the populations included. Thus, the average bootstrap value of the NJ trees increased from 41.9 (HVI) up to 52.8 (approximately 1.3 times higher) when HVII data were incorporated into the analysis. Furthermore, the conspicuously doubtful locations of Austria (in an African cluster) on the one hand, and of Japan—separated from the rest of Asian populations and showing the greatest genetic affinity with Hispanics (see Fig. 3a)—were remarkably improved by including data on HVII polymorphic sites (see Fig. 3b). Consistent with the findings of the MDS, the Georgian population segregated alone, in an intermediate position among the main population clusters. This result reinforces the idea that the Caucasus region might have functioned as a migratory corridor for human displacements between Africa, Asia and Europe.

Both the MDS data and the NJ data failed to identify the remnants of a putative Paleolithic background shared between Georgians and Basques, as some authors have claimed based mainly on linguistic criteria (Lafon 1951). If the Caucasian and Basque languages really are related pre-Indo-European languages of Paleolithic antiquity, one might expect to see evidence of genetic affinity between both groups. Nevertheless, neither phylogenetic results (MDS and NJ data) derived from this study nor the forensic data (RMP) support such a controversial hypothesis. The genetic heterogeneity observed between Georgians and Basques with respect to mtDNA could be generated by differences in mutation rates, in demographic histories and differences regarding impact of the genetic drift. This situation would in turn be reinforced by the effects of a limited gene flow resulting from isolation by distance, all of which may have obscured any common mtDNA background between both human groups (Bertorelle et al. 1995; Comas et al. 2000).

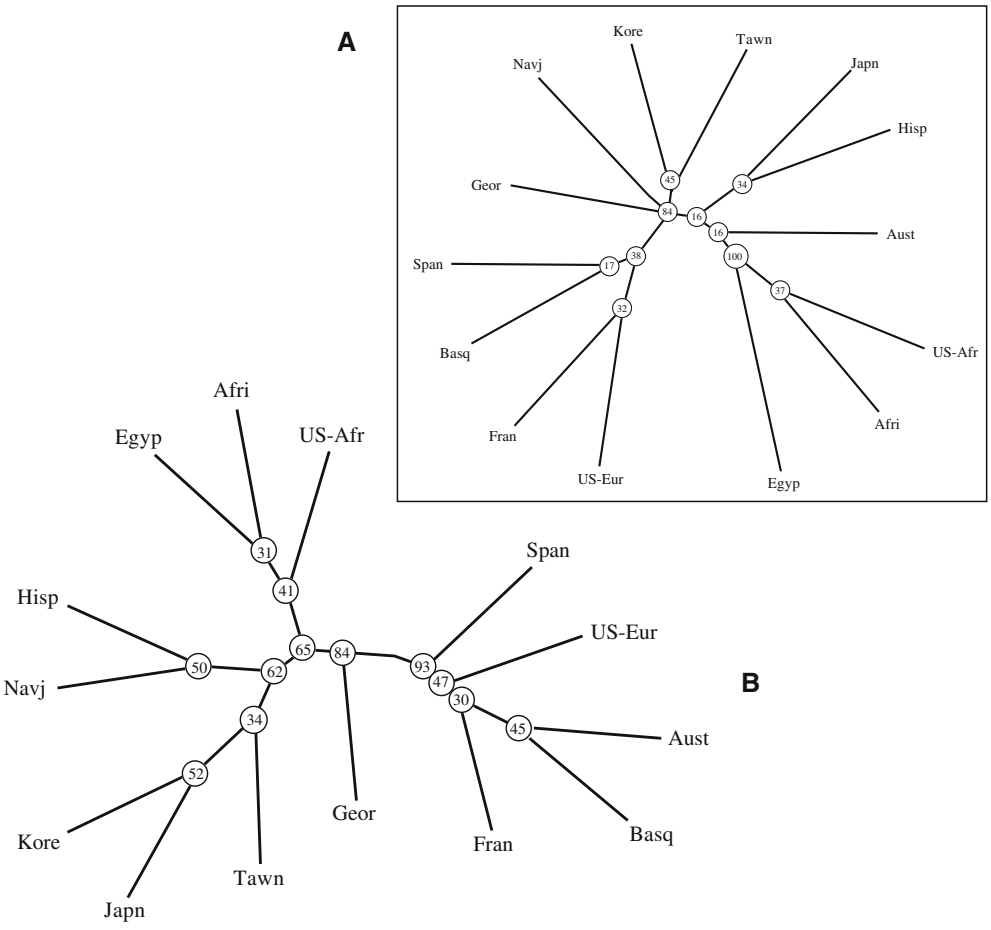

Fig. 3a,b Neighbor-joining (NJ) trees based on mismatch-intermatch distance (Rao 1982; Nei 1987) applied on mtDNA data for 14 populations worldwide. Figures in tree nodes are percentage bootstrap values, estimated from 1,000 iterations. **a** NJ tree obtained using data on sequence variation for HVI. **b** NJ tree obtained using data on sequence variation for HVI + HVII. *Geor* Georgia, *Basq* Basques, *Span* Spain, *Fran* France, *Aust* Austria, *US-Eur* United States' Europids, *Egyp* Egypt, *Afri* sub-Saharan African, *US-Afri* Afro-Americans, *Tawn* Taiwan, *Japn* Japan, *Kore* Korea, *Hisp* Hispanics, *Navj* Navajo

To summarize, the Sd increased markedly when both mtDNA segments were analyzed jointly (HVI: 0.985, HVII: 0.975, HVI + VII: 0.994) and, consequently, the RMP decreased from 0.034 (HVI) to 0.026 (HVI + HVII). In other words, incorporation of the HVII segment into the analysis increases the power of discrimination of the mtDNA. Similarly, incorporation of HVII data contributed to the refinement of results regarding the genetic relationships and population affinities among the samples included in the analyses. These results unquestionably stress the importance of considering HVII in both population and forensic genetics.

Acknowledgments This study was funded in part by Research Projects BOS 2002-01677 from Ministerio de Ciencia y Tecnología, Spain, and I/UPV 00076.125-14901/2002 and GIU 05/51 from Universidad del País Vasco (UPV/EHU). M.A. Alfonso-Sánchez was supported through a postdoctoral fellowship from the "Programa de Formación de Investigadores", Departamento de Educación, Universidades e Investigación (Basque Government).

References

Anderson S, Bankier AT, Barrell BG, Bruijn HL de, Coulson AR, Drouin J, Eperon IC Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

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 [5]. Nat Genet 23:147

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Arcos-Burgos M, Muenke M (2002) Genetics of population isolates. Clin Genet 61:233–247

Bertorelle G, Bertranpetit J, Calafell F, Nasidze IS, Barbujani G (1995) Do Basque- and Caucasian-speaking populations share non-Indo-European ancestors? Eur J Hum Genet 3:256–263

Budowle B, Wilson MR, Di Zinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL (1999). Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 103:23–25

Budowle B, Allard MW, Fisher CL, Isenberg AR, Monson KL, Stewart JEB, Wilson MR, Miller KWP (2002) HVI and HVII mitochondrial DNA data in Apaches and Navajos. Int J Legal Med 116:212–215

Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 4:119–141

Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60:35–49

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetit J (1996) Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and

its relationship to the European populations. Mol Biol Evol 13:1067–1077
Comas D, Calafell F, Bendukidze N, Fañanas L, Bertranpetit J (2000) Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. Am J Phys Anthropol 112:5–16
Côrte-Real HB, Macaulay V, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha SS, Bertranpetit J, Sykes B (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350
Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113:89–97
Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
Felsenstein J (1989) PHYLIP: Phylogeny inference package (version 3.2). Cladistics 5:164–166
Francalacci P, Bertranpetit J, Calafell F, Underhill P (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100:443–460
Gamkrelidze T, Ivanov V (1990) The early history of Indo-European languages. Sci Am 262:110–116
Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Computer Applications in the Biosciences (CABIOS) 8 (2):189–191
Holt IJ, Harding AE, Morgan-Hughes JA (1988) Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. Nature 331:717–719
Imaizumi K, Parson 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
Jones DA (1972) Blood samples: probability of discrimination. J Forensic Sci Soc 12:355–359
Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27
Lafon R (1951) Concordances morphologiques entre le basque et les langues caucasiques. Word 7:227–244
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
Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 57:403–414
Mergen H, Oner R, Oner C (2004) Mitochondrial DNA sequence variation in the Anatolian Peninsula (Turkey). J Genet 83:39–47
Nasidze I, Stoneking M (2001) Mitochondrial DNA variation and language replacements in the Caucasus. Proc R Soc Lond B 268:1197–1206
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
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
Pereira L, Gusmao L, Prata MJ, Moto P, Trovoada MJ, Amorim A (1999) Detection of additional structural variation at the FES/FPS system and population data from S. Tome e Principe and North Portugal. Int J Legal Med 112:204–206
Pérez-Miranda AM, Alfonso-Sánchez MA, Vidales MC, Calderón R, Peña JA (2004) Genetic polymorphism and linkage disequilibrium of the HLA-DP region in Basques from Navarre (Spain). Tissue Antigens 64:264–275
Pérez-Miranda AM, Alfonso-Sánchez MA, Kalantar A, García-Obregón S, De Pancorbo MM, Peña JA, Herrera RJ (2005) Microsatellite data support subpopulation structuring among Basques. J Hum Genet 50:403–414
Pfeiffer H, Steighner R, Fisher R, Mornstad H, Yoon CL, Holland MM (1998) Mitochondrial DNA extraction and typing from isolated dentin-experimental evaluation in a Korean population. Int J Legal Med 111:309–313
Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112:291–298
Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Leg Med 106:85–90
Procaccio V, Wallace DC (2004) Late-onset Leigh syndrome in a patient with mitochondrial complex I NDUFS8 mutations. Neurology 62:1899–1901
Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. Theor Popul Biol 21:24–43
Reardon W, Ross RJ, Sweeney MG, Luxon LM, Pembrey ME, Harding AE, Trembath RC (1992) Diabetes mellitus associated with a pathogenic point mutation in mitochondrial DNA. Lancet 340:1376–1379
Renfrew C (1991) Before Babel, speculations on the origins of linguistic diversity. Cambridge Archaeol J 1:3–23
Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260
Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. Int J Legal Med 111:292–298
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365–375
Salas A, Lareu V, Calafell F, Bertranpetit J, Carracedo A (2000) mtDNA hypervariable region II (HVII) sequences in human evolution studies. Eur J Hum Genet 8:964–974
Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York
Schneider S, Roessli D, Excoffier L (2000) A software for population genetics data analysis. Arlequin version 2.000. Genetics and Biometry Laboratory, University of Geneva, Switzerland
Schurr TG, Wallace DC (2002) Mitochondrial DNA diversity in Southeast Asian populations. Hum Biol 74:431–452
Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann Hum Genet 69:67–89
Strimmer K, von Haeseler A (1995) PUZZLE: Maximum likelihood analysis for nucleotide, aminoacid, and two-state data phylogenetic workshop. University of Bielefeld, Germany
Sudo A, Honzawa S, Nonaka I, Goto Y (2004) Leigh syndrome caused by mitochondrial DNA G13513A mutation: frequency and clinical features in Japan. J Hum Genet 49:92–96
Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. J Hum Genet 49:187–193
Torroni A, Wallace DC (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. J Bioenerg Biomembr 26:261–271
Vanecek T, Borrel F, Sip M (2004) Mitochondrial DNA D-loop hypervariable regions: Czech population data. Int J Legal Med 118:14–18
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991). African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wallace DC (1995) Mitochondrial DNA variation in human evolution, degenerative disease, and aging. Am J Hum Genet 57:201–223
Wallace DC (1997) Mitochondrial DNA in aging and disease. Sci Am 277:40–47
Wallace DC (1999) Mitochondrial diseases in man and mouse. Science 283:1482–1488
Zeviani M, Tiranti V, Piantadosi C (1998) Mitochondrial disorders. Medicine 77:59–72
Zupanic Pajnic I, Balazic J, Komel R (2004) Sequence polymorphism of the mitochondrial DNA control region in the Slovenian population. Int J Legal Med 118:1–4