Mitochondrial DNA Variability in the Czech Population, with Application to the Ethnic History of Slavs

BORIS A. MALYARCHUK,[1] TOMAS VANECEK,[2] MARIA A. PERKOVA,[1] MIROSLAVA V. DERENKO,[1] AND MIROSLAV SIP[3]

Abstract Mitochondrial DNA (mtDNA) variability was studied in a sample of 179 individuals representing the Czech population of Western Bohemia. Sequencing of two hypervariable segments, HVS I and HVS II, in combination with screening of coding-region haplogroup-specific RFLP markers revealed that most Czech mtDNAs belong to the common West Eurasian mitochondrial haplogroups (H, pre-V, HV*, J, T, U, N1, W, and X). However, about 3% of Czech mtDNAs encompass East Eurasian lineages (A, N9a, D4, M*). A comparative analysis with published data showed that different Slavonic populations in Central and Eastern Europe contain small but marked amounts of East Eurasian mtDNAs. We suggest that the presence of East Eurasian mtDNA haplotypes is not an original feature of the gene pool of the proto-Slavs but rather may be mostly a consequence of admixture with Central Asian nomadic tribes, who migrated into Central and Eastern Europe in the early Middle Ages.

Polymorphisms of the mitochondrial genome and nonrecombining regions of the Y chromosome are now the focus of many population and evolutionary genetics studies. Analysis of the variation in mitochondrial DNA (mtDNA), which is maternally inherited and is not affected by recombination, is widely used in studying the population genetic history and the peopling of different regions of the world. Previous mtDNA studies have shown that Slavonic- and Baltic-speaking populations share a common genetic substratum characteristic of Central and Eastern European populations, such as German and western Finno-Ugric populations (Malyarchuk et al. 2002; Pliss et al. 2006). Malyarchuk et al. (2003) suggested that this genetic substratum also penetrates southeastern European populations (such as the Bosnians and Slovenians), thus reaching territory as far as the Western Balkans. Clustering of most Slavonic populations (Poles, Russians, Slovenians, Bosnians, Herzegovinians, Macedonians, Serbians, Croatians), with the

[1]Institute of Biological Problems of the North, Russian Academy of Sciences, Portovaya str. 18, 685000 Magadan, Russia.
[2]Department of Pathology, Medical Faculty Hospital, Charles University, Pilsen, Czech Republic.
[3]Faculty of Health and Social Studies, University of South Bohemia, Ceske Budejovice, Czech Republic.

Human Biology, December 2006, v. 78, no. 6, pp. 681–696.

KEY WORDS: MITOCHONDRIAL DNA, HYPERVARIABLE SEGMENT (HVS I AND II), CZECH POPULATION, MOLECULAR PHYLOGEOGRAPHY.

exception of the island populations of Croatians, has also been revealed in the mtDNA study of southeastern Europeans (Cvjetan et al. 2004).

Meanwhile, Slavonic population samples have been analyzed in different ways. Some studies covered only mtDNA noncoding hypervariable segment I (HVS I) or, in addition, HVS II (Malyarchuk et al. 1995; Calafell et al. 1996; Orekhov et al. 1999; Vanecek et al. 2004; Zupanic Pajnic et al. 2004); other studies also included coding-region RFLP markers diagnostic of all major mtDNA clusters present in human populations (Richards et al. 2000; Tolk et al. 2000; Malyarchuk and Derenko 2001; Belyaeva et al. 2003; Cvjetan et al. 2004; Malyarchuk et al. 2004). Combined HVS I and II sequencing and RFLP analysis appears to be useful for phylogeographic interpretations of mtDNA variation data. However, only a limited number of Slavonic population samples—Russians, Poles, Bosnians, and Slovenians (Malyarchuk et al. 2002, 2003)—have been investigated using this high-resolution approach. Thus many populations of the Slavs remain underexplored at the present time. The aim of the present study is to characterize the mtDNA variation (based on variation of the HVS I and HVS II sequences typed for the presence of major Eurasian haplogroup-specific RFLP markers) in a Czech population representing Western Slavs. This study allowed us to obtain a better characterization of Slavonic mtDNA variability and to extend conclusions about phylogenetic relationships between European populations and the ethnic history of the Slavs.

Materials and Methods

Population Samples. A population sample of 179 Czech individuals from Western Bohemia was studied. All participating individuals were maternally unrelated and originated from the area considered for this study. Appropriate informed consent was obtained from all participants.

mtDNA HVS I and II Sequencing. The total DNA was isolated from blood samples with a QIAamp DNA Blood Mini kit (QIAgen, Hilden, Germany) according to the manufacturer's protocol. Ten nanograms of DNA was amplified using primers L15926 and H580, as described previously (Allen et al. 1998). Sequencing reactions were carried out using the Big Dye Terminator sequencing kit (Applied Biosystems, Foster City, California) and primers L15996, L29, H7, H271, H408, and H580 (Allen et al. 1998; Vanecek et al. 2004). Sequencing products were analyzed on an ABI Prism 310 automatic sequencer (Applied Biosystems). The nucleotide sequences from position 15997 to 16400 and from position 30 to 407 were determined and compared with the revised Cambridge Reference Sequence (Anderson et al. 1981; Andrews et al. 1999).

mtDNA RFLP Analysis. To determine the haplogroup status of the control region sequences, we performed RFLP typing using a restriction endonuclease

analysis of PCR-amplified mtDNA fragments using the same primer pairs and amplification conditions as described by Torroni et al. (1996) and Finnilä et al. (2000). The samples were typed for a set of RFLPs that were diagnostic of all major Eurasian clusters, on the basis of the hierarchical mtDNA RFLP scheme (Macaulay et al. 1999; Richards et al. 2000; Yao et al. 2002; Loogväli et al. 2004). The table reporting 33 RFLP markers analyzed in this study can be provided upon request from B. Malyarchuk. HVS I and II sequences, belonging to haplogroups N1a, N1b, N9a, U1, U2c, U3, and U5, were identified on the basis of the mtDNA classification (Kong et al. 2003; Palanichamy et al. 2004; Achilli et al. 2004, 2005).

Phylogeographic and Statistical Analysis. Genetic variation was analyzed using methods implemented in Arlequin, version 2.0 (Schneider et al. 2000). The statistical significance of F_{ST} values was estimated by permutation analysis using 10,000 permutations. Multidimensional scaling analysis of pairwise interpopulation F_{ST} values was performed by means of the software package Statistica (StatSoft Inc., Tulsa, Oklahoma).

For statistical analysis we used data from the following populations: 200 southern Germans (Lutz et al. 1998); 101 Austrians (Parson et al. 1998); 150 western Germans (Baasner et al. 1998; Baasner and Madea 2000); 104 Slovenians and 144 Bosnians (Malyarchuk et al. 2003); 436 Poles and 201 Russians (Malyarchuk et al. 2002); and 50 Finns, 47 Estonians, and 83 Karelians (Sajantila et al. 1995). For purposes of phylogeographic analysis we used all available published data on HVS I mtDNA variability or RFLPs in Slavonic-speaking populations (in addition to the aforementioned population samples): Bulgarians (Calafell et al. 1996; Richards et al. 2000); Russians (Malyarchuk et al. 1995; Orekhov et al. 1999; Richards et al. 2000; Malyarchuk et al. 2002; Belyaeva et al. 2003; Malyarchuk et al. 2004); Ukrainians (Malyarchuk and Derenko 2001); Belorussians (Belyaeva et al. 2003); Poles (Richards et al. 2000; Malyarchuk et al. 2002); Czechs (Richards et al. 2000; Vanecek et al. 2004); Bosnians (Malyarchuk et al. 2003; Cvjetan et al. 2004); Slovenians (Malyarchuk et al. 2003; Zupanic Pajnic et al. 2004); Croatians (Tolk et al. 2000; Cvjetan et al. 2004); and Serbians, Herzegovinians, and Macedonians (Cvjetan et al. 2004). We should note that previously published control region mtDNA sequences in 93 Czech individuals (Vanecek et al. 2004) do not overlap with the data of the present study.

Results

The analysis of HVS I and II variation in combination with RFLP typing of the coding-region haplogroup-diagnostic sites allowed detection of 146 different mtDNA haplotypes in a total sample of 179 Czech individuals from Western Bohemia (Table 1; the full data reporting different control region haplotypes in

Table 1. mtDNA Haplotypes in the Czech Population[a]

HVS I (Minus 16000)	*HVS II*	*Haplogroup*	*N*
223-290-319-362	73-152-235-263-309iC-315iC	A	1
188-223-362	73-263-309iCC-315iC	D4	1
CRS	263-315iC	H*	6
CRS	263-309iC-315iC	H*	1
CRS	263-309iCC-315iC	H*	4
CRS	150-262-263-315iC	H*	1
CRS	152-263-315iC	H*	1
CRS	152-200-263-309iC-315iC	H*	1
93	242-263-309iC-315iC	H*	1
93-129-316	152T/C-263-315iC	H*	1
111-129-256	93-153-263-309iC-315iC	H*	1
114	263-309iCC-315iC	H*	1
114-209	263-309iCC-315iC	H*	1
126-192	263-315iC	H*	1
129	263-315iC	H*	1
188-212-245	263-309iC-315iC	H*	1
218-278	93-263-315iC	H*	1
311	263-315iC	H*	1
311	263-309iC-315iC	H*	1
324	263-315iC	H*	1
CRS	263-315iC	H1*	2
CRS	263-309iC-315iC	H1*	2
CRS	263-309iCC-315iC	H1*	1
CRS	152C/T-263-315iC	H1*	1
CRS	200-263-309iCC-315iC	H1*	1
93	263-315iC	H1*	2
93-263-327-362	259G/A-263-315iC	H1*	1
129	263-315iC	H1*	1
153-189-193iC-295	263-309iCC-315iC	H1*	1
168	263-315iC	H1*	1
183C-189-193iC	146-263-309iCC-315iC	H1*	1
209	263-315iC	H1*	1
239	263-315iC	H1*	1
51-162	73-263-315iC	H1a	1
51-162-291-304	73-263-315iC	H1a	1
162	73-263-309iC-315iC	H1a	2
162	73-263-309iCC-315iC	H1a	1
80-183C-189-193iC-356	263-309iCC-315iC	H1b	1
80-189-193iC-223-356	151-263-315iC	H1b	1
183C-189-193iC-356	263-315iC	H1b	1
183C-189-193iC-356-362	263-309iCC-315iC	H1b	2
185-189-356-362	263-315iC	H1b	1
189-193iC-356	152-263-315iC	H1b	1
189-193iC-356-362	146-263-309iCC-315iC	H1b	1
CRS	249D-263-309iC-315iC	H2*	1
CRS	263-309iC-315iC	H2*	1
CRS	309iCC-315iC	H2*	2
274	263-315iC	H2*	2
193-354	93-207-263-309iC-315iC	H2a1	1

Table 1. (Continued)

HVS I (Minus 16000)	*HVS II*	*Haplogroup*	*N*
193-354	150-263-309iC-315iC	H2a1	1
193-354	263-315iC	H2a1	1
260	263-315iC	H4	1
271-304	263-309iC-315iC	H5*	1
304	146-263-309iCC-315iC	H5*	1
304	150-263-309iC-315iC	H5*	1
304	263-309iCC-315iC	H5*	1
304	152-263-315iC	H5a	1
304	263-309iC-315iC	H5a	1
362	239-263-315iC	H6	2
193-219-362	204-239-263-309iC-315iC	H6	1
261	194-195-263-309iC-315iC	H7	1
265	263-309iCC-315iC	H7	1
293-311	143-195-263-309iC-315iC	H11a	1
293-311	143-195-263-309iCC-315iC	H11a	1
293-311	195-263-315iC	H11a	1
93-104-311	152-263-309iCC-315iC	HV*	1
111-311	146-263-309iC-315iC	HV*	1
311	131-152-263-309iCC-315iC	HV*	1
86-129-223-391	73-152-199-204-207-239-250-263-309iC-315iC	I	1
129-172-223-287-311-391	73-199-203-204-245C/T-250-263-309iC-315iC	I	1
129-172-223-311-391	73-189-199-203-204-250-263-315iC	I	1
129-172-223-311-319-391	73-189-199-202-203-204-250-263-309iC-315iC	I	1
129-223-391	73-152-188-199-204-207-250-263-309iCC-315iC	I	1
69-126-145-222-235-261-271	73-263-295-315iC	J1b*	1
69-126-145-222-235-261-271	73-263-295-309iC-315iC	J1b*	4
69-126-145-172-222-261	73-242-263-295-315iC	J1b1	1
69-126-145-172-222-261	73-146-200-242-263-295-309iC-315iC	J1b1	1
69-126	73-228-263-295-309iC-315iC	J1c	1
69-126	73-146-185-188-222-228-263-295-309iC-315iC	J1c	1
69-126-186-209	73-185-188-228-263-295-309iC-315iC	J1c	1
69-126-186-224	73-152-185-228-263-295-315iC	J1c	2
69-126-192	73-185-188-228-263-295-315iC	J1c	1
69-126-222	73-146-185-228-263-295-309iC-315iC	J1c	1
69-126-287	73-185-228-263-295-315iC	J1c	1
69-126-293A/G-311	73-228-263-295-315iC	J1c	1
69-126-295G	73-185-204-228-263-295-315iC	J1c	1
69-126-355	73-185-188-228-263-295-315iC	J1c	1
69-126-390	73-185-228-263-295-315iC	J1c	1
69-126-145-231-261	73-150-152-195-215-263-295-309iC-315iC-319	J2a	1
69-126-145-231-261	73-150-152-195-215-263-295-309iCC-315iC-319	J2a	1
93-224-311	73-152-263-315iC	K	1
93-224-311	73-195-263-315iC	K	1
224-311	73-263-315iC	K	1
224-311	73-146-152-263-315iC	K	3
224-311-320	73-146-152-263-315iC	K	1
172-189-193iC-223-311	73-146-263-315iC	M*	2

Table 1. (Continued)

HVS I (Minus 16000)	*HVS II*	*Haplogroup*	*N*
147A-172-223-248-320-355	73-152-199-204-263-315iC	N1a	1
145-176G-223-390	73-152-263-315iC	N1b	1
145-176G-183C-189-193iCC-223-390	73-152-263-315iC-338	N1b	1
129-223-257A-261	73-150-263-309iC-315iC	N9a	1
37-126-294-296-304	73-146-263-315iC	T*	1
86-126-294-296	73-263-315iC	T*	1
126-153-294	73-150-263-315iC	T*	1
126-153-294	73-150-263-309iC-315iC	T*	2
126-153-294-296	73-150-200-263-309iC-315iC	T*	1
126-172-294-304	73-200-263-309iC-315iC	T*	1
126-175G/A-294-296-304	73-263-315iC	T*	1
126-182C-183C-189-193iC-294-296-298	73-195-207-263-315iC	T*	1
126-234-294-296-304-357	73-146-263-315iC	T*	4
126-292-294-296	73-146-263-309iC-315iC	T*	1
126-294-304	73-263-315iC	T*	2
126-294-296-304	73-263-309iCC-315iC	T*	1
126-163-186-189-294	73-152-263-315iC	T1	1
126-163-186-189-294	73-152-195-263-309iC-315iC	T1	3
126-163-186-189-294-362	73-152-195-263-309iC-315iC	T1	1
51-145-172-184-189-192-234-294-342	73-152-199-315iC	U2c	1
343-362	73-150-263-309iC-315iC	U3	1
343-390	73-150-263-315iC	U3	1
343-390	73-150-263-309iC-315iC	U3	2
134-221-234-356	73-152-195-263-315iC	U4a	1
CRS	73-195-263-310	U4a	1
295-356	73-195-263-310	U4a	1
183C-189-193iC-356	73-152-185-189-195-215-263-315iC	U4*	1
265-356-362	73-195-247-263-315iC	U4*	1
147G-256-270-311	73-263-315iC	U5a	2
176-192-256-270-278-399	73-263-315iC	U5a	1
189-191iC-192-256-270-311-362	73-151-152-263-309iC-315iC	U5a	1
192-256-270	73-263-309iC-315iC	U5a	1
192-256-270-399	73-263-309iC-315iC	U5a	2
192-256-270-320-399	73-195-207-263-315iC	U5a	1
256-270-399	73-152-263-309iC-315iC	U5a	1
93-189-193iC-270	73-150-263-315iC-384	U5b	1
189-192-270	73-150-263-315iC	U5b	1
189-192-270-398	73-150-263-315iC	U5b	1
189-193iC-270	73-150-152-263-315iC	U5b	1
192-311	73-150-263-309iC-315iC	U5b	1
270-304	73-150-228-263-315iC	U5b	1
146-342	73-263-282-309iC-315iC	U8a	1
298	72-199C/T-263-309iC-315iC	V	1
298	72-263-309iCC-315iC	V	1

Table 1. (Continued)

HVS I (Minus 16000)	*HVS II*	*Haplogroup*	*N*
298	263-309iCC-315iC	V	1
37-298-311	72-195-263-309iC-315iC	pre*V1	1
93-223-291-292-325	73-143-152-189-195-204-207-263-315iC	W	1
183C-189-193iC-223-278	73-152-153-195-263-295-309iC-315iC	X	1
183C-189-193iC-223-278	73-153-195-225-226-263-315iC	X	1
189-193iC-223-278	73-153-195-225-226-263-309iC-315iC	X	1

a. Mutations are shown by indicating positions relative to the revised Cambridge Reference Sequence (Andrews et al. 1999). The nucleotide positions in HVS I and II sequences correspond to transitions; transversions are further specified. Haplogroup names are given according to the mtDNA classification (Macaulay et al. 1999; Richards et al. 2000; Palanichamy et al. 2004; Achilli et al. 2005). The presence of insertions (i) or deletions (D) follows the nucleotide position. Heteroplasmic mutations are designated by a slash.

combination with RFLP variants can be provided upon request from B. Malyarchuk). Comparison to the revised Cambridge Reference Sequence (Andrews et al. 1999) showed that 84 and 44 nucleotide positions were polymorphic in the HVS I and HVS II regions, respectively. To determine the phylogenetic status of the control region sequences, we performed a restriction analysis of the coding-region haplogroup-specific sites. As a result, we found that most of the mitochondrial haplotypes in our Czech population from Western Bohemia are clustered into West Eurasian haplogroups HV (H, pre-V, HV*), J (J1, J2), T (T1, T*), U (K, U2c, U3, U4, U5a, U5b, U8a), N1 (I, N1a, N1b), N2 (W), and X (Table 2). The frequency of East Eurasian–specific haplotypes in Czechs is low (2.8%). These haplotypes belong to macrohaplogroups M (D4 and M*) and N (A and N9a).

The main mitochondrial haplogroup in the Czech population is H, present in 44% of the Czech samples. Haplogroup H is characterized by a considerable branching substructure, with several large subclusters (Finnilä et al. 2001; Herrnstadt et al. 2002; Achilli et al. 2004; Loogväli et al. 2004; Pereira et al. 2005). Among the Czechs, 12 H subclusters were found using the RFLP approach described by Loogväli et al. (2004). Table 3 shows the distribution of subclusters H*, H1*, H1a, H1b, H2*, H2a1, H4, H5*, H5a, H6, H7, and H11a revealed in Czechs and in other European populations. As can be seen, the most frequent H subgroups (besides the paragroups H* and H1*) in the Czech population are H1b, H2, H5, and H1a. Recent studies have shown that these subgroups appear to be more frequent in populations of Eastern and North-Central Europe (Loogväli et al. 2004; Pereira et al. 2005). Subclades within H1 subgroups H1a and H1b are more frequent in Eastern and Central Europe. For instance, subgroup H1a is characteristic of Germans [especially those from northeastern Germany (Poetsch et al. 2003)], different Slavonic populations (except for some Balkan populations, such as Bosnians and Croats), and Finno-Ugric populations (Table

Table 2. Haplogroup Distribution in Czechs Compared to Other Slavonic Populations[a]

Haplogroup	*Czechs (179), N (%)*	*Bosnians (144), N (%)*[b]	*Slovenians (104), N (%)*[b]	*Poles (436), N (%)*[c]	*Russians (201), N (%)*[c]
H	79 (44.13)	69 (47.92)	49 (47.12)	197 (45.18)	85 (42.29)
HV*	3 (1.68)	1 (0.69)	0	4 (0.92)	4 (1.99)
pre-V	4 (2.23)	9 (6.25)	7 (6.73)	21 (4.82)	11 (5.47)
pre-HV	0	2 (1.39)	0	0	1 (0.50)
J	21 (11.73)	10 (6.94)	10 (9.62)	34 (7.80)	16 (7.96)
T*	17 (9.5)	5 (3.47)	5 (4.81)	41 (9.40)	18 (8.96)
T1	5 (2.79)	2 (1.39)	1 (0.96)	9 (2.06)	4 (1.99)
K	7 (3.91)	6 (4.17)	4 (3.85)	15 (3.44)	6 (2.99)
U1	0	2 (1.39)	0	0	2 (1.00)
U2	1 (0.56)	0	1 (0.96)	4 (0.92)	3 (1.49)
U3	4 (2.23)	1 (0.69)	2 (1.92)	2 (0.46)	2 (1.00)
U4	5 (2.79)	8 (5.56)	6 (5.77)	22 (5.05)	7 (3.48)
U5a	9 (5.03)	10 (6.94)	8 (7.69)	23 (5.28)	15 (7.46)
U5b	6 (3.35)	7 (4.86)	3 (2.88)	15 (3.44)	6 (2.99)
U7	0	0	0	1 (0.23)	1 (0.50)
U8	1 (0.56)	0	0	2 (0.46)	0
U*	0	0	0	1 (0.23)	0
I	5 (2.79)	4 (2.78)	2 (1.92)	8 (1.83)	5 (2.49)
W	1 (0.56)	2 (1.39)	5 (4.81)	16 (3.67)	4 (1.99)
X	3 (1.68)	2 (1.39)	1 (0.96)	8 (1.83)	7 (3.48)
N1a	1 (0.56)	0	0	0	0
N1b	2 (1.12)	1 (0.69)	0	1 (0.23)	0
N1c	0	0	0	1 (0.23)	0
N9a	1 (0.56)	0	0	0	0
R*	0	0	0	2 (0.46)	1 (0.50)
L1b	0	1 (0.69)	0	0	0
L3	0	0	0	1 (0.23)	0
M	3 (1.68)	2 (1.39)	0	8 (1.83)	3 (1.49)
A	1 (0.56)	0	0	0	0
$h \pm$ SE	0.78 ± 0.03	0.75 ± 0.04	0.75 ± 0.04	0.77 ± 0.02	0.8 ± 0.03
$i \pm$ SE	4.92 ± 2.41	4.07 ± 2.04	4.07 ± 2.05	4.60 ± 2.26	4.70 ± 2.31

a. Diversity values based on mtDNA haplogroup frequencies (h) and mean number of pairwise differences between HVS I sequences (i) are shown.

b. Data for Bosnians and Slovenians are from Malyarchuk et al. (2003).

c. Data for Poles and Russians are from Malyarchuk et al. (2002).

3). Subgroup H1b is also typical of Central and Eastern Europeans, being found at the highest frequencies (more than 3%) in Estonians, Latvians, Czechs, Poles, and South Germans (present data; Lutz et al. 1998; Malyarchuk et al. 2002; Pliss et al. 2006).

Haplogroup J is characterized by high frequency in the Czech population (11.7%) and is represented by subgroups J1b, J1c, and J2a. Among J1b haplotypes one specific branch, J1b*, is remarkable, because this haplotype is characterized by the HVS I motif 16069-16126-16145-16222-16235-16261-16271,

Table 3. Frequencies of Haplogroup H Subgroups in Czechs and Other European Populations[a]

Subgroup	*Czechs*	*Slovaks*	*Eastern Slavs*	*Estonians*	*Latvians*	*Volga-Ural Region Finno-Ugric Population*	*Balkan Population*
H*	14.5	18.5	15.2	18.4	15.1	16.8	24.3
H1*	8.94	2.52	6.4	9.65	5.7	11.2	3.6
H1a	2.79	3.36	3.2	1.75	1.0	2.4	0
H1b	4.46	1.68	1.6	5.26	6.7	0	1.8
H2*	3.35	0	0.8	0.88	0.7	0	0
H2a	1.68	0.84	7.2	2.63	2.0	0	0
H4	0.56	1.68	0	0	0.7	1.6	0
H5*	2.23	3.36	1.6	0.88	1.0	0	2.7
H5a	1.12	0.84	0.8	0.88	5.7	0	4.5
H6	1.68	2.52	0.8	1.75	1.3	1.6	3.6
H7	1.12	0.84	0	0.88	1.3	3.2	2.7
H8	0	0.84	0	0	0	0	0
H11	1.68	5.04	1.6	0.88	3.3	3.2	1.8
H sample	79	50	50	50	133	50	50
H frequency (%)	44.1	42.0	40.0	44.0	44.5	40.0	45.0
Total sample size	179	119	125	114	299	125	111

a. Data on distribution of haplogroup H subgroups in European populations were taken from Loogväli et al. (2004), except for Latvians (Pliss et al. 2006) and Czechs (present data).

which is rare in European populations but is typical in Polish and Spanish Roma (Gresham et al. 2001; Malyarchuk et al. 2006). Among Europeans this haplotype has previously been revealed at highest frequency in Czechs (2.3%; Vanecek et al. 2004). It is noteworthy that the U2c haplotype found in the Czech population (see Table 1) is also of Asian ancestry, because it has been shown that the distribution of haplogroup U2c is essentially restricted to Indo-Pakistani regions (Quintana-Murci et al. 2004). In addition, Roma-specific mtDNA haplogroup M5 has been observed in the previously published Czech sample (Vanecek et al. 2004). Moreover, it is curious that the Roma-specific M5 lineages were also observed in other Slavonic populations: Poles (Malyarchuk et al. 2002), Bosnians (Malyarchuk et al. 2003), and Bulgarians (Richards et al. 2000).

Several haplotypes (2.8%) found in our Czech population belong to East Eurasian haplogroups A, D4, N9a, and M*. It is noteworthy that almost all Slavonic populations studied to date show the presence of East Eurasian mtDNA haplogroups. They were found in 1.5% of Russians and 1.6% of Poles (Malyarchuk et al. 2002); in 1.4% of Bosnians (Malyarchuk et al. 2003; Cvjetan et al. 2004); and in 2% of Croatians, 1.5% of Herzegovinians, 0.9% of Serbians, and 2.1% of Macedonians (Cvjetan et al. 2004).

To study differentiation of European populations and to determine the position of the Czech population among Europeans, we performed an analysis of molecular variance (AMOVA) on the level of HVS I sequences. For analysis, populations belonging to different language groups—Slavonic-speaking (Czechs, Slovenians, Bosnians, Poles, Russians), German-speaking (West Germans, South Germans, Austrians), and Finno-Ugric-speaking (Finns, Karelians, Estonians)—were used. Analysis of between-population differentiation revealed that only 0.165% of variation was due to differences among populations ($p = 0.026$). Significant pairwise F_{ST} differences ($p < 0.05$) were found only between some Finno-Ugric populations (Finns and Karelians) and Slavonic and German populations (Table 4). Therefore the AMOVA results point to a low but statistically significant level of mtDNA differentiation in the European populations compared. The results of within-group F_{ST} analysis indicate that all populations within linguistically determined groups appear to be homogeneous (with F_{ST} values close to 0 and $p > 0.1$). Significant differences were observed between populations when they were treated as three separate groups in accordance with their linguistic affiliation ($F_{ST} = 0.22\%$ and $F_{CT} = 0.14\%$ with $p < 0.03$ in both cases; $F_{SC} = 0.08\%$, $p = 0.186$) (Table 5). The multidimensional scaling analysis performed on the basis of pairwise F_{ST} values (see Table 4) revealed that there is an apparent differentiation only between Karelians and Finns and the remaining European populations (Figure 1).

Discussion

According to the AMOVA results based on HVS I variation data, the Czech population showed no significant differences with neighboring European populations, except for the most northeastern European populations (see Table 4). Detailed analysis of mtDNA structure in the Czech population based on sequencing

Table 4. Between-Population Differences Based on Pairwise F_{ST} Distances Inferred from mtDNA HVS I Variation Data

Population	*Czechs*	*Poles*	*Slovenians*	*Bosnians*	*Russians*	*Austrians*	*South Germans*	*West Germans*	*Estonians*	*Finns*
Poles	0.0013									
Slovenians	0.0031	0.0009								
Bosnians	0.0032	0.0006	0.0003							
Russians	−0.0028	−0.0008	0.0019	−0.0011						
Austrians	−0.0031	0.0010	0.0015	−0.0007	−0.0046					
South Germans	0.0007	0.0001	0.0004	−0.0012	−0.0014	−0.0008				
West Germans	0.0010	−0.0024	−0.0010	0.0021	−0.0026	0.0003	0.0005			
Estonians	−0.0007	−0.0031	0.0017	−0.0016	−0.0006	−0.001	−0.0007	−0.0082		
Finns	0.0085[a]	0.0038	0.0076[a]	0.0017	0.0087[a]	0.0082[a]	0.0043	0.0026	−0.0018	
Karelians	0.0078[a]	0.0052[a]	0.0089[a]	0.0034	0.0056	0.0087[a]	0.0062[a]	−0.0009	−0.0063	0.0010

a. Significant differences ($p < 0.05$).

Table 5. AMOVA Results According to Linguistic Classification of European Populations, Based on mtDNA Variation

	Percentage of Variation		
Classification	*Among Groups*	*Among Populations, Within Groups*	*Within Populations*
Linguistic[a]	0.32	0.06	99.62

a. The following language groups were used for classification: Slavonic (represented by Czechs, Slovenians, Bosnians, Poles, and Russians), German (represented by West Germans, South Germans, and Austrians), and Finno-Ugric (represented by Finns, Karelians, and Estonians).

of two hypervariable segments in combination with RFLP screening of coding-region markers revealed that the vast majority of Czech mtDNAs belong to common West Eurasian haplogroups (H, pre-V, HV*, J, T, U, N1, W, and X). An important finding of this study is that about 3% of mtDNA haplotypes found in the Czech population from Western Bohemia are identified as East Eurasian specific, belonging to haplogroups D4, A, N9a, and M*. Population screening of East Eurasian mtDNAs observed in Czechs revealed that the closest mtDNA haplotypes can be found mainly among Bashkirs from the South Ural region and

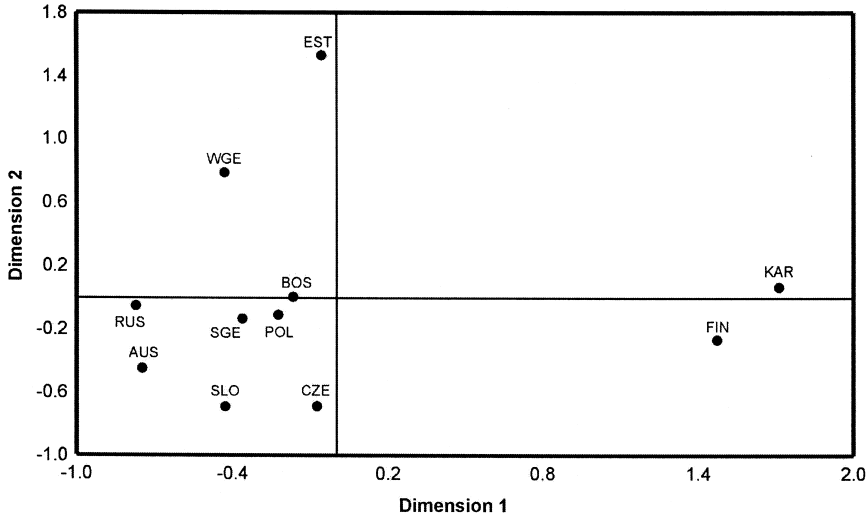

Figure 1. Multidimensional scaling plot of F_{ST} distances between Czechs and the surrounding European populations based on mtDNA HVS I variation data (stress value 0.054). CZE, Czechs; POL, Poles; SLO, Slovenians; BOS, Bosnians; RUS, Russians; AUS, Austrians; SGE, South Germans; WGE, West Germans; EST, Estonians; FIN, Finns; KAR, Karelians.

Table 6. Frequency of East Eurasian mtDNAs in Gene Pools of Different European Ethnolinguistic Population Groups

Population	*Sample Size*	*Frequency (%)*	*Haplogroup Composition*	*Reference*
Germanic-speaking populations				
Germans	333	0.6	D	Pliss et al. (2006)
Norwegians	397	0.5	Z	Pliss et al. (2006)
Finno-Ugric-speaking populations				
Estonians	409	0.2	D	Pliss et al. (2006)
Karelians	83	4.8	D	Sajantila et al. (1995)
Finns	580	1.8	Z, D	Pliss et al. (2006)
Hungarians	98	1.7	B, M	Semino et al. (2000)
Mari	136	7.4	A, M*, C, Z, D	Bermisheva et al. (2002)
Mordva	102	2.9	C, D	Bermisheva et al. (2002)
Baltic-speaking populations				
Latvians	299	0.3	G	Pliss et al. (2006)
Lithuanians	180	0.6	A	Pliss et al. (2006)
Slavonic-speaking populations				
Eastern Slavs (Russians)	526	1.3	M*, C, D, G	Malyarchuk et al. (2002, 2004)
Western Slavs (Poles, Czechs)	708	1.8	A, N9a, M*, C, D, G	Malyarchuk et al. (2002); present study; Vanecek et al. (2004)
Southern Slavs (Bosnians, Croatians, Herzegovinians, Serbians, Macedonians, Slovenians)	1,705	1.2	A, F, D, M*	Cvjetan et al. (2004); Malyarchuk et al. (2003)

among Nogays from the North Caucasus [according to data of Bermisheva et al. (2002, 2004)]. Matching HVS I haplotypes have also been observed in South Siberian populations, such as the Buryats and Khakassians (Derenko et al. 2003).

Comparative analysis with published data shows that gene pools of different Slavonic populations contain small but definite amounts of East Eurasian mtDNA lineages (Table 6). It is unclear, however, whether this is due to common ancestry of Slavonic groups or to recent immigration of East Asian mtDNAs. According to linguistic and archeological data, the split within the Balto-Slavonic language group occurred at about 1700 B.C., resulting in the proto-Baltic and proto-Slavonic language groups (Šavli et al. 1996). It is noteworthy that Baltic populations (Latvians, Lithuanians, and Estonians) have avoided a marked influence of maternal lineages of East Eurasian origin (see Table 6) (Pliss et al. 2006).

Only two East Eurasian mtDNA haplogroups, Z1 and D5, are present in the gene pools of North European Finnic populations, such as Saami, Finns, and Karelians. Unlike these populations, Slavonic populations in general are characterized by much more heterogeneous mtDNA composition, defined, in addition to Z1 and D5, by haplogroups A, C, D4, G2a, M*, N9a, F, and Y. Therefore different scenarios of female-mediated East Eurasian genetic influence on Northern and Eastern Europeans should be highlighted: (1) The most ancient influence, probably originating in the early Holocene, was an influx of Asian tribes, which brought a few selected East Asian mtDNA haplotypes to Fennoscandia (Tambets et al. 2004); and (2) gradual gene flows during historical times, mostly in the Middle Ages, were due to migrations of nomadic peoples (such as Huns in the 4th century, Avars in the 6th century, and Mongols in the 13th century) to Eastern and Central European territories inhabited mainly by Slavonic tribes.

In Central Europe archeologists have found traces of frequent invasions from the East European steppes dating even to approximately 1500 B.C. (Sedov 1979). In the early Middle Ages Europeans experienced several invasions, the most important of which was the Avarian one. The Avars were a nomadic Turkic-speaking people of Eurasia who migrated into Central and Eastern Europe in the early 6th century. The Avars reached the territories of Wallachia, Pannonia, Transylvania, and Bohemia, where they established a powerful state (the Avar Khaganate), which ruled over areas of Eastern and Central Europe and controlled the Slavs, who had lived in the area before the Avar arrival. Avar rule persisted over much of the Pannonian plain up to the early 9th century. From that time, the Avars likely merged with the Slavs (Gimbutas 1971; Sedov 1979).

Therefore it is possible that after Avar rule was destroyed, the Avar maternal lineages, including East Eurasian ones, were assimilated mainly by the Slavs. Hence further spreading of Avar mtDNAs in Europe was probably connected with the Slavonic migrations. Thus it seems that the occurrence of East Eurasian mtDNA haplotypes in the Slavs is not an original feature of their gene pool but rather a trace of relatively recent interactions between expanding Slavonic tribes and their nomadic neighbors in Europe. Meanwhile, the presented data do not allow us to discriminate between the age of the mtDNA molecules and the age of their movements into the Czech and other Slavonic populations. Therefore a much more detailed phylogeographic analysis, requiring the mtDNA variation data in both the source Asian (including populations from still poorly investigated Siberian and Kazakhstan regions) and the recipient European populations, is necessary to attain this goal.

Acknowledgments This study was supported by a grant ("Dynamics of Gene Pools and Biodiversity") from the Program of Basic Research of the Russian Academy of Sciences.

Received 6 June 2006; revision received 13 October 2006.

Literature Cited

Achilli, A., C. Rengo, V. Battaglia et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* 75:910–918.

Achilli, A., C. Rengo, C. Magri et al. 2005. Saami and Berbers: An unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* 76:883–886.

Allen, M., A. S. Engstrom, S. Meyers et al. 1998. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: Sensitivity and matching probabilities. *J. Forensic Sci.* 43:453–464.

Anderson, S., A. T. Bankier, B. G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.

Andrews, R. M., I. Kubacka, P. F. Chinnery et al. 1999. Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA. *Nat. Genet.* 23:147.

Baasner, A., and B. Madea. 2000. Sequence polymorphisms of the mitochondrial DNA control region in 100 German Caucasians. *J. Forensic Sci.* 45:1343–1348.

Baasner, A., C. Schäfer, A. Junge et al. 1998. Polymorphic sites in human mitochondrial DNA control region sequences: Population data and maternal inheritance. *Forensic Sci. Int.* 98:169–178.

Belyaeva, O., M. Bermisheva, A. Khrunin et al. 2003. Mitochondrial DNA variations in Russian and Belorussian populations. *Hum. Biol.* 75:647–660.

Bermisheva, M. A., I. A. Kutuev, T. Y. Korshunova et al. 2004. Phylogeographic analysis of mitochondrial DNA in the Nogays: A strong mixture of maternal lineages from Eastern and Western Eurasia. *Mol. Biol. (Moscow).* 38:516–523.

Bermisheva, M., K. Tambets, R. Villems et al. 2002. Diversity of mitochondrial DNA haplogroups in ethnic populations of the Volga-Ural region. *Mol. Biol. (Moscow)* 36:990–1001.

Calafell, F., P. Underhill, A. Tolun et al. 1996. From Asia to Europe: Mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann. Hum. Genet.* 60:35–49.

Cvjetan, S., H. V. Tolk, L. Barac Lauc et al. 2004. Frequencies of mtDNA haplogroups in Southeastern Europe: Croatians, Bosnians, and Herzegovinians, Serbians, Macedonians, and Macedonian Romani. *Colleg. Antropol.* 28:193–198.

Derenko, M. V., T. Grzybowski, B. A. Malyarchuk et al. 2003. Diversity of mitochondrial DNA lineages in South Siberia. *Ann. Hum. Genet.* 67:391–411.

Finnilä, S., I. E. Hassinen, L. Ala-Kokko et al. 2000. Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. *Am. J. Hum. Genet.* 66:1017–1026.

Finnilä, S., M. S. Lehtonen, and K. Majamaa. 2001. Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* 68:1475–1484.

Gimbutas, M. 1971. *The Slavs.* New York: Praeger.

Gresham, D., B. Morar, P. A. Underhill et al. 2001. Origins and divergence of the Roma (Gypsies). *Am. J. Hum. Genet.* 69:1314–1331.

Herrnstadt, C., J. L. Elson, E. Fahy et al. 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.* 70:1152–1171.

Kong, Q.-P., Y.-G. Yao, C. Sun et al. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* 73:671–676.

Loogväli, E.-L., U. Roostalu, B. A. Malyarchuk et al. 2004. Disuniting uniformity: A pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.* 21:2012–2021.

Lutz, S., H.-J. Weisser, J. Heizmann et al. 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int. J. Legal Med.* 111:67–77.

Macaulay, V., M. Richards, E. Hickey et al. 1999. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* 64:232–249.

Malyarchuk, B. A., and M. V. Derenko. 2001. Mitochondrial DNA variability in Russians and Ukrainians: Implication to the origin of the Eastern Slavs. *Ann. Hum. Genet.* 65:63–78.

Malyarchuk, B., M. Derenko, T. Grzybowski et al. 2004. Differentiation of mitochondrial DNA and Y chromosome in Russian populations. *Hum. Biol.* 76:877–900.

Malyarchuk, B. A., M. V. Derenko, and L. L. Solovenchuk. 1995. Types of mitochondrial DNA control region in the Eastern Slavs. *Rus. J. Genet.* 31:723–727.

Malyarchuk, B. A., T. Grzybowski, M. V. Derenko et al. 2002. Mitochondrial DNA variability in Poles and Russians. *Ann. Hum. Genet. 6*6:261–283.

Malyarchuk, B. A., T. Grzybowski, M. V. Derenko et al. 2003. Mitochondrial DNA variability in Bosnians and Slovenians. *Ann. Hum. Genet.* 67:412–425.

Malyarchuk, B. A., T. Grzybowski, M. V. Derenko et al. 2006. Mitochondrial DNA diversity in the Polish Roma. *Ann. Hum. Genet.* 70:195–206.

Orekhov, V., A. Poltoraus, L. A. Zhivotovsky et al. 1999. Mitochondrial DNA sequence diversity in Russians. *FEBS Lett.* 445:197–201.

Palanichamy, M. G., C. Sun, S. Agrawal et al. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: Implications for the peopling of South Asia. *Am. J. Hum. Genet.* 75:966–978.

Parson, W., T. J. Parsons, R. Scheithauer et al. 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., M. Richards, A. Goios et al. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* 15:19–24.

Pliss, L., K. Tambets, E.-L. Loogvali et al. 2006. Mitochondrial DNA portrait of Latvians: Towards the understanding of the genetic structure of Baltic-speaking populations. *Ann. Hum. Genet.* 70:439–458.

Poetsch, M., H. Wittig, D. Krause et al. 2003. Mitochondrial DNA diversity of a northeast German population sample. *Forensic Sci. Int.* 137:125–132 (corrigendum, 2004, *Forensic Sci. Int.* 145:73–77).

Quintana-Murci, L., R. Chaix, R. S. Wells et al. 2004. Where West meets East: The complex mtDNA landscape of the southwest and central Asian corridor. *Am. J. Hum. Genet.* 74:827–845.

Richards, M. B., V. A. Macaulay, E. Hickey et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67:1251–1276.

Sajantila, A., P. Lahermo, T. Anttinen et al. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.* 5:42–52.

Šavli, J., M. Bor, and I. Tomažic. 1996. *Veneti: First Builders of European Community—Tracing the History and Language of Early Ancestors of Slovenes.* Vienna: Editiones Veneti.

Schneider, S., D. Roessli, and L. Excoffier. 2000. *Arlequin, Ver. 2.0: A Software for Population Genetics Data Analysis.* Geneva: Genetics and Biometry Laboratory, University of Geneva.

Sedov, V. V. 1979. *Origin and Early History of Slavs.* Moscow: Nauka (in Russian).

Semino, O., G. Passarino, L. Quintana-Murci et al. 2000. mtDNA and Y-chromosome polymorphisms in Hungary: Inferences from the Palaeolithic, Neolithic and Uralic influences on the modern Hungarian gene pool. *Eur. J. Hum. Genet.* 8:339–346.

Tambets, K., S. Rootsi, T. Kivisild et al. 2004. The Western and Eastern roots of the Saami: The story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* 74:661–682.

Tolk, H. V., M. Pericic, L. Barac et al. 2000. mtDNA haplogroups in the populations of Croatian Adriatic Islands. *Colleg. Antropol.* 24:267–279.

Torroni, A., K. Huoponen, P. Francalacci et al. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850.

Vanecek, T., F. Vorel, and M. Sip. 2004. Mitochondrial DNA D-loop hypervariable regions: Czech population data. *Int. J. Legal Med.* 118:14–18.

Yao, Y.-G., Q.-P. Kong, H.-J. Bandelt et al. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* 70:635–651.

Zupanic Pajnic, I., J. Balazic, and R. Komel. 2004. Sequence polymorphism of the mitochondrial DNA control region in the Slovenian population. *Int. J. Legal Med.* 118:1–4.