Brief Communications

Mitochondrial Haplogroup U2d Phylogeny and Distribution

BORIS MALYARCHUK,[1] MIROSLAVA DERENKO,[1] MARIA PERKOVA,[1] AND TOMAS VANECEK[2]

Abstract The sequencing of the entire mitochondrial DNA belonging to haplogroup U2d reveals that this clade is defined by four coding-region mutations at positions 1700, 4025, 11893, and 14926. Phylogenetic analysis suggests that western Eurasian haplogroup U2d appears to be a sister clade with the Indo-Pakistani haplogroup U2c. Results of a phylogeographic analysis of published population data on the distribution of haplogroup U2d indicate that the presence of such mtDNA lineages in Europe may be mostly a consequence of medieval migrations of nomadic tribes from the Caucasus and eastern Europe to central Europe.

Mitochondrial haplogroup U2, which is characterized by a transition at nucleotide position (np) 16051 in the hypervariable segment 1 (HVS1), has been subdivided into two branches: the "European" U2e, defined by transversion 16129C; and the "Indian" U2i, which lacks such a transversion (Kivisild et al. 1999a). In addition, haplogroup U2i is represented by three clusters: U2a, U2b, and U2c (Kivisild et al. 1999b; Palanichamy et al. 2004; Quintana-Murci et al. 2004). Haplogroup U2a is characterized by the rare and stable HVS1 transversion 16206C. Haplogroup U2b is defined by the mutations 146, 2706, 5186T, 12106, 13149, and 15049. Haplogroup U2c is recognized by the mutations 152, 5790A, 14935, 15061, and 16234. The distribution of these sister clades within haplogroup U2 is essentially restricted to the Indo-Pakistani regions. They have not been observed in Europe and the Near East, and they are rare in Iranian plateau and Central Asian populations (Metspalu et al. 2004; Quintana-Murci et al. 2004). The estimated coalescence times for these haplogroups are 45,700 $\pm$ 14,400 years for haplogroup U2a, 35,900 $\pm$ 9,000 years for haplogroup U2b, and 45,200 $\pm$ 10,400 years for haplogroup U2c (Quintana-Murci et al. 2004). Based on the diversity of the U2a, U2b, and U2c lineages in India, Quintana-Murci et al. (2004) estimated a potential founder age of this part of the U phylogeny of 49,900 $\pm$ 7,900 years. Thus an

[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, Plzen, Czech Republic.

Human Biology, October 2008, v. 80, no. 5, pp. 565–571.

KEY WORDS: MITOCHONDRIAL DNA, HAPLOGROUP U2D, HUMAN POPULATIONS, MOLECULAR PHYLOGEOGRAPHY.

entry of haplogroup U2 in India more recent than 40,000 years ago is not plausible (Palanichamy et al. 2004).

Phylogeographic studies have shown that haplogroup U2e is present mainly in western Eurasian populations at a frequency of 1% on average (Richard et al. 2007). Meanwhile, one more haplogroup U2 member in western Eurasia is haplogroup U2d, which is rare and is found only occasionally in some populations of western Asia, the Caucasus, and Ethiopia (Kivisild et al. 2004). Haplogroup U2d is defined only by HVS1 motif 16051-16189-16234-16294. The phylogenetic position of this haplogroup is also obscure, despite the fact that one complete mitochondrial genome has been sequenced (Maca-Meyer et al. 2001). On the basis of coding and control-region mutations, Palanichamy et al. (2004) suggested that haplogroups U2d and U2e form a sister cluster. However, this assumption is tentative because it is based on only the diagnostic mutation at position 16189, which is a hypervariable site (Bandelt et al. 2002; Malyarchuk et al. 2002).

The aim of this study is to reconstruct the phylogeographic pattern of haplogroup U2d in different Eurasian populations based on the mtDNA control-region and coding region variability data. As a result of the analysis of the published data, we have found that in west Eurasian populations, haplogroup U2d consists mainly of two related groups of lineages—with HVS1 motifs 16051-16189-16234-16294 and 16051-16184-16234-16294-16342, respectively (Table 1). Both groups are present at low frequencies in populations of the Middle East, the Caucasus, and eastern and central Europe. A monophyletic origin of these clusters is confirmed by the common HVS2 motif 73-152-199-471 with a back-mutation at position 263. A single exception to this rule is the HVS2 sequence of a Jordanian individual (GenBank number AF382000) sequenced by Maca-Meyer et al. (2001); however, it is likely that there was an error in Maca-Meyer et al.'s sequence, so we propose that the HVS2 motif 73-152-199-471 is specific for haplogroup U2d.

To reconstruct the phylogeny of haplogroup U2d, we have completely sequenced the mitochondrial genome of a Czech individual from western Bohemia. These data were compared with the U2d genome of the Jordanian sample previously studied by Maca-Meyer et al. (2001) (Figure 1). The comparison allows us to reveal a U2d trunk defined by four coding-region mutations at positions 1700, 4025, 11893, and 14926. Whether or not the transition at np 16189 is trunk specific is unclear because U2d2 sequences may or may not have a mutation at position 16189 (see Table 1). This suggests that the ancestral U2d lineage was characterized by the 16189T variant and that the transition from T to C was generated several times in both haplogroups U2d1 and U2d2, taking into account the instability of the T base flanked by poly-C tracts (Bendall and Sykes 1995). This further suggests that, in fact, haplogroup U2d might share two mutations—at positions 152 and 16234—with haplogroup U2c (see Figure 1). If so, the Indian U2c and the western Eurasian U2d lineages separated from each other long ago, close to the root of their putative common ancestor.

It is worth noting that haplogroup U2d has an interesting geographic pattern of distribution (see Table 1). Both of its subgroups, U2d1 and U2d2, are present

Table 1. Distribution of Haplogroup U2d Control-Region Sequences

HVS1 Sequence	*HVS2 Sequence*	*Population Origin and Frequency (%)*	*Reference*
U2d1			
16051-16189-16234-16294		Ethiopian (0.4)	Kivisild et al. (2004)
		Turkish Kurds (0.9)	Nasidze et al. (2005)
		Armenians (1.0)	Richards et al. (2000)
		Romanians (1.0)	Richards et al. (2000)
		Mordva (2.0)	Bermisheva et al. (2002)
16051-16189-16234-16294	73-152-199[a]	Kirghiz (0.9)	Comas et al. (1998, 2004)
16051-16189-16234-16294	73-152-263-471[b]	Jordanian	Complete mtDNA genome from Maca-Meyer et al. (2001)
16051-16189-16294		Mordva (1.0)	Bermisheva et al. (2002)
16051-16093-16111-16189-16234-16294		Ethiopian (0.4)	Kivisild et al. (2004)
16051-16189-16234-16266-16294	73-152-199-471[b]	Hungarian Roma (1.0)	Egyed et al. (2007)
16051-16189-16234-16266-16294		Nogays (0.5)	Bermisheva et al. (2004)
16051-16189-16266-16294	73-152-199[c]	Italians (0.5)	Babalini et al. (2005)
		Croatian Italians (2.4)	Babalini et al. (2005)
16051-16189-16234-16266-16294-16352	73-152-199-471[b]	Romanians (0.6)	Egyed et al. (2007)

(*continued on next page*)

Table 1. (*continued*)

HVS1 Sequence	*HVS2 Sequence*	*Population Origin and Frequency (%)*	*Reference*
U2d2			
16051-16184-16234-16294-16342	73-152-199-471[b]	Dubai Arabs (0.4)	Alshamali et al. (2008)
16051-16184-16234-16291-16294-16342		Palestinians (0.9)	Richards et al. (2000)
16051-16093-16184-16189-16234-16294-16342		Persians (0.2)	Quintana-Murci et al. (2004), Metspalu et al. (2004), Derenko et al. (2007)
16051-16145-16172-16184-16189-16192-16234-16294-16342	73-152-199-471[b]	Czechs (0.6)	Malyarchuk et al. (2006)
16051-16093-16148-16184-*16210-16233*-16234-16266-16294[d]		Darginians (2.7)	Nasidze and Stoneking (2001)
16051-16148-16184-16234-16294-16342-16357		Darginians (2.7)	Nasidze and Stoneking (2001)
16051-16093-*16137*-16148-16184-16189-16234-16294-16342-16357[d]		Avars (3.1)	Nasidze et al. (2004)

Length variation and transversions in the poly-C stretches at positions 16180–16193 and 309–315 and the transition at 16519 have not been considered because of instability of these nucleotide positions.

a. Sequenced between positions 63 and 322.

b. Sequenced between positions 1 and 576.

c. Sequenced between positions 30 and 408.

d. Probable sequencing errors shown in italics.

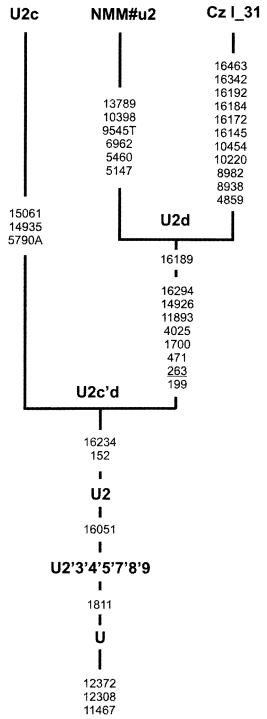

Figure 1. Complete genome-based phylogenetic tree of haplogroups U2d and U2c. The tree is rooted in haplogroup U. Numbers along links refer to substitutions scored relative to the revised Cambridge reference sequence (Andrews et al. 1999). Transversions are further specified; recurrent mutations are underlined. For subhaplogroup U2c only diagnostic mutations are shown according to classification (Palanichamy et al. 2004). The complete mitochondrial genome of Czech individual Cz I_31 was sequenced by means of the procedures described by Torroni et al. (2001). This sequence has been submitted to GenBank (accession number EU440736). An additional U2d complete sequence was taken from the literature (Maca-Meyer et al. 2001) and was designated NMM followed by # and the original sample code. For phylogeny construction, the A/C stretch length polymorphism in regions 16180–16193 and 303–315, 522–523del, and mutation 16519, all known to be hypervariable, were disregarded.

in populations of the Middle East, the Caucasus area, and eastern Europe. Meanwhile, U2d2 lineages have been detected mostly in southern regions of western Eurasia—in the United Arab Emirates, in Israeli Palestinians, and in Iran as well as in the Caucasus among Dagestani populations (such as Darginians and Avars). Haplotype U2d2 has also been found in Czechs from western Bohemia. This population is characterized by a relatively high frequency (2.8%) of east Eurasian mtDNA lineages, which were probably inherited from Asian nomadic tribes that were assimilated by Europeans in the early Middle Ages (Malyarchuk et al. 2006).

An east to west direction of gene flow can also be suggested for U2d1 lineages, because these mtDNA lineages are observed in the steppe regions of southern Russia (among Mordva and Nogay people) and further in Transylvania (among Romanians and Hungarian Roma) and even on the Adriatic coast (among the Croatians and Croatian-Italians), besides being found in the Middle East, Anatolia, and the Caucasus (see Table 1). Overall, the data obtained have allowed us to reconstruct the phylogeny of haplogroup U2d with greater accuracy than phylogenies reached earlier, and they reveal an interesting phylogeographic pattern of distribution of haplogroup U2d in Europe.

Acknowledgments This work was supported by the Program of Basic Research of the Russian Academy of Sciences through grant 06-I-P1-032 ("Biodiversity and Gene Pools Dynamics").

Received 4 August 2008; revision received 8 September 2008.

Literature Cited

Alshamali, F., A. Brandstätter, B. Zimmermann et al. 2008. Mitochondrial DNA control region variation in Dubai, United Arab Emirates. *Forensic Sci. Intl. Genet.* 2:e9–e10.

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.

Babalini, C., C. Martínez-Labarga, H. V. Tolk et al. 2005. The population history of the Croatian linguistic minority of Molise (southern Italy): A maternal view. *Eur. J. Hum. Genet.* 13:902–912.

Bandelt, H. J., L. Quintana-Murci, A. Salas et al. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* 71:1150–1160.

Bendall, K. E., and B. C. Sykes. 1995. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am. J. Hum. Genet.* 57:248–256.

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.

Comas, D., F. Calafell, E. Mateu et al. 1998. Trading genes along the silk road: MtDNA sequences and the origin of central Asian populations. *Am. J. Hum. Genet.* 63:1824–1838.

Comas, D., S. Plaza, R. S. Wells et al. 2004. Admixture, migrations, and dispersals in Central Asia: Evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* 12:495–504.

Derenko, M., B. Malyarchuk, T. Grzybowski et al. 2007. Phylogeographic analysis of mitochondrial DNA in North Asian populations. *Am. J. Hum. Genet.* 81:1025–1041.

Egyed, B., A. Brandstätter, J. A. Irwin et al. 2007. Mitochondrial control region sequence variations in the Hungarian population: Analysis of population samples from Hungary and from Transylvania (Romania). *Forensic Sci. Intl. Genet.* 1:158–162.

Kivisild, T., M. J. Bamshad, K. Kaldma et al. 1999a. Deep common ancestry of Indian and western Eurasian mitochondrial DNA lineages. *Curr. Biol.* 9:1331–1334.

Kivisild, T., K. Kaldma, M. Metspalu et al. 1999b. The place of the Indian mtDNA variants in the global network of maternal lineages and the peopling of the Old World. In *Genomic Diversity: Applications in Human Population Genetics*, R. Deka and S. S. Papiha, eds. New York: Kluwer Academic Plenum, 135–152.

Kivisild, T., M. Reidla, E. Metspalu et al. 2004. Ethiopian mitochondrial DNA heritage: Tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet.* 75:752–770.

Maca-Meyer, N., A. M. González, J. M. Larruga et al. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* 2:13.

Malyarchuk, B. A., I. B. Rogozin, V. B. Berikov et al. 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. *Hum. Genet.* 111:46–53.

Malyarchuk, B. A., T. Vanecek, M. A. Perkova et al. 2006. Mitochondrial DNA variability in the Czech population with application to the ethnic history of Slavs. *Hum. Biol.* 78:681–696.

Metspalu, M., T. Kivisild, E. Metspalu et al. 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* 5:26.

Nasidze, I., E. S. Ling, D. Quinque et al. 2004. Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann. Hum. Genet.* 68:205–221.

Nasidze, I., D. Quinque, M. Ozturk et al. 2005. MtDNA and Y-chromosome variation in Kurdish groups. *Ann. Hum. Genet.* 69:401–412.

Nasidze, I., and M. Stoneking. 2001. Mitochondrial DNA variation and language replacements in the Caucasus. *Proc. Biol. Sci.* 268:1197–1206.

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.

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.

Richard, C., E. Pennarun, T. Kivisild et al. 2007. An mtDNA perspective of French genetic variation. *Ann. Hum. Biol.* 34:68–79.

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.

Torroni, A., C. Rengo, V. Guida et al. 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.* 69:1348–1356.