ISSN 0026-8933, Molecular Biology, 2008, Vol. 42, No. 3, pp. 399–402. © Pleiades Publishing, Inc., 2008.
Original Russian Text © B.A. Malyarchuk, 2008, published in Molekulyarnaya Biologiya, 2008, Vol. 42, No. 3, pp. 454–458.

GENOMICS. TRANSCRIPTOMICS. PROTEOMICS

UDC 575.174:599.9

Cold Spots of Human Mitochondrial DNA Hypervariable Segment 1

B. A. Malyarchuk

Institute of Biological Problems of the North, Far East Division, Russian Academy of Sciences, Magadan, 685000 Russia; e-mail: malyarchuk@ibpn.ru

Received December 6, 2007
Accepted for publication January 24, 2008

Abstract—The distribution of mutations in hypervariable segment 1 (HVS1) of mitochondrial DNA (mtDNA) was analyzed for more than 37000 individuals from various regions of the world. The results were used to estimate the intensity of mutation processes and the features of the cold spot distribution in mtDNA. Analysis of the structural–functional organization and variation of HVS1 made it possible to associate a lower variation with functionally important HVS1 regions. The distribution of CAT cold spots in secondary DNA structures revealed a lack of correlation between the cold spot location and the structural type of the mtDNA region.

DOI: 10.1134/S0026893308030072

Key words: human mitochondrial DNA, major noncoding region, hypervariable segment 1, cold spots, DNA secondary structure

INTRODUCTION

Phylogenetic analysis has revealed the so-called hot spots (nucleotide positions characterized by a higher mutation rate) in hypervariable segments 1 and 2 (HVS1 and HVS2) of the major noncoding region of mtDNA [1, 2]. The hypervariability of nucleotide positions results in identical (parallel) mutations appearing in phylogenetically unrelated mtDNA types.

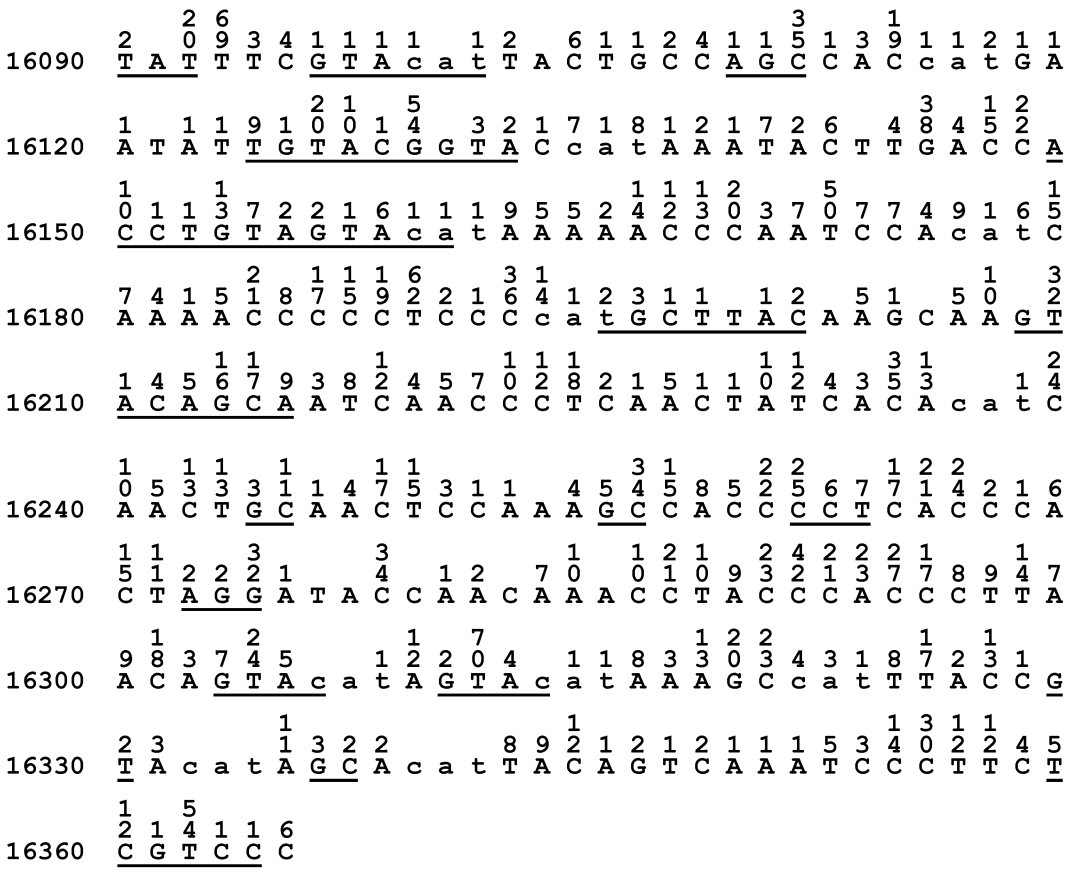
Earlier analysis of the mutation spectra of mtDNA HVS1 has revealed four classes of nucleotide regions with different levels of variability [1]. Two of these classes contain hot spots, and one class includes obvious cold sites, where the number of independent (parallel) mutations varies from 0 to 2. Context analysis of HVS1 has identified two consensus sequences for hot spots in T and C positions. The sequences are KTNCNK and CC, where K is G or T and N is any nucleotide (the hot spot is underlined). It is interesting that the GTACAT sequence, which is hypervariable in some cases and conserved in some others [1, 2], is a particular case of the KTNCNK consensus. It is known that some HVS1 regions are practically monomorphic; i.e., there are no mutations in these regions, whereas the cold spot class is wider and includes 0–2 mutations [1]. However, the results of a monomorphism study greatly depend on the size of databases and the accepted allowance for the ethno-racial diversity. The data on the mtDNA variability in the native populations of South Asia, Australia, and America have been published only in recent years. These data make it possible to obtain more reliable results of cold spot distribution analysis.

The objective of this work was to study the distribution of mutations in mtDNA HVS1 by analyzing a database containing data about more than 37000 persons from different regions of the world.

EXPERIMENTAL

Reconstruction of the HVS1 mutation spectrum. To make such a reconstruction, data on the distribution of variable positions along a DNA region are insufficient. The most precise estimation of the hot spot distribution in HVS1 is provided by analysis of the variability of HVS1 sequences combined into monophyletic clusters (mtDNA groups) according to polymorphism of coding mtDNA regions, whose variation rate is significantly lower than that of HVS1 [1–3]. When mtDNA groups are isolated, HVS1 polymorphism is taken into account as far as HVS1 variants are group specific. Thus, according to the earlier accepted methodology [1, 2], variable positions were searched for within HVS1 and the number of parallel mutations was determined for each mtRNA group. Parallel mutations were identical polymorphic variants that independently appeared in different mtDNA groups.

Population data. Several sets of population data were analyzed. The HVS1 variation was analyzed using a set of 7482 HVS1 sequences (positions 16090–16365), which belonged to 88 mtDNA groups



Mutation spectrum of human mtDNA HVS1. The number of independent mutations occurring in 88 phylogenetic mtDNA groups is indicated at the top of a nucleotide. DNA regions included into the stems of hairpin structures according to the analysis of the HVS1 secondary structure are underlined. The CAT regions of mtDNA are shown with lowercase letters.

[2]. This database represented West Eurasian (n = 3834), East Eurasian (n = 801), and African (n = 2847) populations. The regional distribution of mtDNA groups was as follows: 28 groups for West Eurasia (H, HV*, pre-V, pre-HV, R*, T1, T*, J*, J1a, J1b, J2, K, U*, U1, U2, U3, U4, U5, U7, U8a, U8b, N1a, N1b, N1c, N*, I, W, and X), 34 groups for East Eurasia (C, Z, M8a, D4, D5, G2, G3, G4, E, M*, M7*, M7b, M7c, M9, M10, A, N9a, N2, N*, Y, R9a, R*, F*, F1a, F1b, F1c, F2, B*, B4*, B4a, B4b, B5*, B5a, and B5b), and 26 groups for Africa (L0a1, L0a2, L1b, L1c*, L1c1, L1c2, L1c3, L1d, L1e, L2a*, L2a1a, L2a1b, L2b, L2c, L2d1, L2d2, L3b1, L3b2, L3d, L3e1, L3e2, L3e3, L3e4, L3f*, L3f1, and L3g).

HVS1 variation analysis. The HVS1 variability in the native populations of South Asia (India, Pakistan, Iran, China, Thailand, and Bangladesh), North and South America, North Asia (South Siberia, Mongolia, Korea, Tadjikistan, and Iran), and Australia, including Melanesia and Papua New Guinea, was analyzed using the databases [4] (n = 4653), [5] (n = 601), [6] (n = 1432), and [7] (n = 1865), respectively. In addition, the HV1_polymorphic database [8] (n = 21 141), representing mainly the Caucasian population of the United States, was analyzed. In total, 37 174 HVS1 sequences were examined. The mtDNA lineages from the databases were included into the above 88 mtDNA groups, since they cover all existing variants of human mtDNA. Some rare mtDNA groups typical for the native populations of Australia, New Guinea, and Southeast Asia were included into the R*, N*, and M* paragroups at this step.

Consideration of mutations. Mutations were identified relative to the L chain of the Cambridge reference sequence of human mtDNA [9]. In all cases, I did not take into account the point insertions/deletions and transversions in unstable HVS1 region 16 184–16 193. The secondary structure of mtDNA regions was reconstructed using the RNAdraw 1.1b2 software (United States).

RESULTS AND DISCUSSION

The mutation spectrum reconstructed for mtDNA HVS1 is shown in the figure. Analysis of the large body of population data revealed some positions with a low variation, including positions that displayed 0–2 independent (parallel) mutations per site in all

Table 1. Cold spots of human mtDNA HVS1

Number of parallel mutations per site	Nucleotide position
0–2	16090, 16091, 16096–16103, 16105–16107, 16109, 16110, 16112, 16115–16123, 16125, 16127, 16128, 16130, 16132, 16133, 16135, 16137–16139, 16141, 16143, 16149, 16151, 16152, 16155–16157, 16159–16161, 16165, 16177, 16182, 16190, 16191, 16194, 16195, 16197–16202, 16204, 16205, 16208, 16210, 16225, 16226, 16228, 16229, 16236–16238, 16246, 16251–16253, 16267, 16268, 16272, 16273, 16275–16277, 16279–16282, 16285, 16306–16308, 16313–16315, 16323, 16326, 16328–16330, 16332–16334, 16337–16341, 16345–16351, 16361, 16363, 16364

Table 2. Conserved regions of human mtDNA HVS1 and their location within DNA secondary structures

Position	Nucleotide sequence	DNA region
16096–16103	GTACATTA	Hairpin stem
16105–16107	TGC	Hairpin stem
16115–16123	CATGAATAT	Interhairpin region
16137–16139	AAA	Hairpin loop
16155–16157	AGT	Hairpin stem
16159–16161	CAT	Hairpin stem
16197–16202	CTTACA	Hairpin stem
16236–16238	CAT	Interhairpin region
16251–16253	CAA	Hairpin loop
16275–16277	ATA	Interhairpin region
16279–16282	CAAC	Interhairpin region
16306–16308	CAT	Hairpin loop
16313–16315	CAT	Interhairpin region
16328–16330	CGT	Hairpin stem
16332–16334	CAT	Hairpin loop
16337–16341	CACAT	Hairpin stem and interhairpin region
16345–16351	AGTCAAA	Interhairpin region

88 mtDNA groups and belonged to the cold spot class [1] (Table 1). This class comprised 115 positions, 42% of the total HVS1 size. In addition to separate cold spots, HVS1 contains short nucleotide sequences with a low variation (16096–16103, 16105–16107, 16115–16123, 16137–16139, 16155–16157, 16159–16161, 16197–16202, 16236–16238, 16251–16253, 16275–16277, 16279–16282, 16306–16308, 16313–16315, 16328–16330, 16332–16334, 16337–16341, and 16345–16351). A lack of polymorphism was revealed for 29 nucleotide positions, which account for about 10.5% of the HVS1 size.

The conservation of some mtDNA regions can be stipulated by their functional importance, since HVS1 contains elements necessary for the initiation and regulation of replication and transcription of the mitochondrial genome. For example, a lower variation was observed for the SP region (16104–16106) and the 7S DNA termination-associated sequence (TAS, 16157–16172). Only the 5'-terminal TAS region, located between positions 16155–16157 and 16159–16161, was quite conserved. A lower variation is typical for the control element (CE, 16194–16208). This DNA region contains a series of cold spots (16194, 16195, 16197–16202, 16204, 16205, and 16208).

About half of the conserved HVS1 regions indicated in Table 2 belong to the CAT type (16099–16101, 16159–16161, 16306–16308, 16313–16315, and 16332–16334). They form part of the GTACAT sequence, typical for mtDNA regions associated with transcription termination: ETAS1 (16081–16138), TAS (16157–16172), and ETAS2 (16294–16353). Note that this sequence can be both hypervariable and conserved. For example, the GTACAT sequence is practically monomorphic in regions 16096–16101, 16156–16161, and 16329–16334, whereas region 16303–16315 contains two GTACAT sequences that each have a hot spot at the first of the two thymine res-

idues. Possibly, GTACAT sequences participate in secondary structure formation, since they are organized as direct repeats and are partially inverted with respect to each other, being able to form alternative palindromic hairpin structures. For example, it is known that rodent ETASs, including GTACAT and its derivatives, form highly stable secondary structures with a stem of 11–17 nt and a loop of 5–9 nt [10]. Moreover, in some animals, repeated GTACAT monomers occurring in the major noncoding mtDNA region allow the formation of various secondary structures, differing in the number of hairpins and the number of monomers in the stems of palindromic hairpin structures [10]. It is quite possible that such an organization is typical for the major noncoding region of human mtDNA. In this case, the appearance of both hot and cold spots in GTACAT sequences can be determined by their location within secondary structures and be explained by specific mutagenesis mechanisms such as correction of imperfect heteroduplexes within palindromic hairpin structures. To check this hypothesis, I studied the cold spot arrangement within HVS1 secondary structures. The RNAdraw 1.1b2 software was used to reconstruct the secondary structure of HVS1 region 16070–16368. This mtDNA region contains eight hairpin structures, and GTACAT sequences participate in the organization of five of these structures. However, CAT trinucleotides are practically not involved in the formation of hairpin stems (figure, Table 2). The conserved HVS1 regions indicated in Table 2 are located both within hairpin structures (in stems (seven cases) and loops (four cases)) and between them (seven cases).

Thus, analysis of the large body of data on the variation of nucleotide sequences belonging to different groups of mtDNA made it possible to estimate the intensity of mutation processes and the features of the cold spot distribution in mtDNA HVS1. Analysis of the structural–functional organization and variation of HVS1 suggests a lower variation for HVS1 regions that potentially are of functional significance. However, the distribution of CAT cold spots in secondary structures shows the lack of correlation between the cold spot location and the structural type of the mtDNA region.

ACKNOWLEDGMENTS

I am grateful to M.V. Derenko (Institute of Biological Problems of the North) for his help in the conduction of this study.

This work was supported by the Far East Division of the Russian Academy of Sciences (project no. 03-3-A-06-096).

REFERENCES

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

2. Malyarchuk B.A., Rogozin I.B. 2004. Mutagenesis by transient misalignment in human mitochondrial DNA control region. *Ann. Hum. Genet*. **68**, 324–339.

3. Bandelt H.-J., Quintana-Murci L., Salas A., Macaulay V. 2002. The fingerprint of phantom mutations in mtDNA data. *Am. J. Hum. Genet.* **71**, 1150–1160.

4. Metspalu M., Kivisild T., Metspalu E., Parik J., Hudjashov G., Kaldma K., Serk P., Karmin M., Behar D.M., Gilbert M.T., Endicott P., Mastana S., Papiha S.S., Skorecki K., Torroni A., Villems R. 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**, e26.

5. Tamm E., Kivisild T., Reidla M., Metspalu M., Smith D.G., Mulligan C.J., Bravi C.M., Rickards O., Martinez-Labarga C., Khusnutdinova E.K., Fedorova S.A., Golubenko M.V., Stepanov V.A., Gubina M.A., Zhadanov S.I., Ossipova L.P., Damba L., Voevoda M.I., Dipierri J.E., Villems R., Malhi R.S. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE*. **2**, e829.

6. Derenko M., Malyarchuk B., Grzybowski T., Denisova G.A., Dambueva I., Perkova M., Dorzhu C., Luzina F., Vanecek T., Villems R., Zakharov I. 2007. Phylogeographic analysis of mitochondrial DNA in North Asian populations. *Am. J. Hum. Genet*. **81**, 1025–1041.

7. Ingman M., Gyllensten U. 2006. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res*. **34**, D749–D751.

8. Behar D.M., Rosset S., Blue-Smith J., Balanovsky O., Tzur S., Comas D., Mitchell R.J., Quintana-Murci L., Tyler-Smith C., Wells R.S., Genographic Consortium. 2007. The genographic project public participation mitochondrial DNA database. *PLoS Genet*. **3**, e104.

9. Andrews R.M., Kubacka I., Chinnery P.F., Lightowlers R.N., Turnbull D.M., Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genet*. **23**, 147.

10. Larizza A., Pesole G., Reyes A., Sbisa E., Saccone C. 2002. Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. *J. Mol. Evol*. **54**, 145–155.