Russian Journal of Genetics, Vol. 41, No. 1, 2005, pp. 79–84. Translated from Genetika, Vol. 41, No. 1, 2005, pp. 93–99.
Original Russian Text Copyright © 2005 by Malyarchuk.

HUMAN GENETICS

Distribution of Nucleotide Substitutions in Human Mitochondrial DNA Genes

B. A. Malyarchuk

Institute of the Biological Problems of the North, Russian Academy of Sciences, Magadan, 685000 Russia; fax: (41322) 3-44-63; e-mail: malyar@ibpn.kolyma.ru

Received February 4, 2004

Abstract—To analyze the distribution pattern of nucleotide substitutions in human mitochondrial DNA (mtDNA), mutational spectra of the mitochondrial genes were reconstructed. The reconstruction procedure is based on the mutation distribution data for 47 monophyletic mtDNA clusters, to which 794 examined mtDNA sequences encoding for tRNAs, rRNAs, and mitochondrial proteins are attributed. One of specific features of mitochondrial mutational spectra revealed was homoplasy of the mutations (the mean mutation number per variable nucleotide site in the coding region varied from 1.09 to 1.43). It was established that in the mtDNA genes maximum mutational constraint fell onto the guanine bases, albeit the content of these bases in the mtDNA L-chains was minimal. Maximal bias towards parallel G to A transitions was observed for rRNA genes, with the protein- and tRNA-encoding genes ranking next. Despite the fact that the differences in the average G-nucleotides content and variability between the genes of two mtDNA segments located between the OriH and OriL were statistically significant, the results did not provide the conclusion that the G-nucleotide instability observed in the mtDNA L-spectra was determined by the mechanism of asynchronous mtDNA replication, along with the deamination of cytosines in the H-chain regions, which remained single-stranded during replication.

INTRODUCTION

Development of human mitochondrial genomics started in the 1980s. At present, a large amount of population data on the variability of complete mitochondrial DNA (mtDNA) sequences has been accumulated [1–5]. The size of human mitochondrial genome constitutes 16569 base pairs (bp), and it contains the genes for two rRNAs, 22 tRNAs, and 13 respiratory chain protein subunits. The main functional elements essential for the mtDNA expression are located within the 1122-bp main noncoding region [6]. The increase in the number of studies on the variation of complete genomic sequences in human populations is followed by the progress in investigations of high mtDNA variation and its functional consequences [7–9].

To date, it has been demonstrated that tRNA and rRNA genes are the most conservative mtDNA regions, and hypervariable segments 1 and 2 (HVS1 and HVS2), located within the main noncoding region, are the most polymorphic [3, 9]. It has been shown that the frequency of parallel mutations (homoplasy) in the main noncoding region was 31 times higher than that in the mtDNA coding sequences [3]. The existence of homoplasy within the mtDNA coding sequences indicates nucleotide site variability in the mutation rate, the reasons for which are still unknown [4]. Hypervariable positions account for 6% of all polymorphic positions within mtDNA. They are rather randomly dispersed over the coding region and are found in different genes, including *ND1*, *ATPase6*, *COIII*, *ND3*, *ND4*, *ND5*, *CytB* and tRNA(*Ser*) [9]. Several studies are devoted to the analysis of the distribution of synonymic and nonsynonymic mutations and their role in the evolution of the mitochondrial proteins [7, 8]. Thereby, at present, initial information essential for the investigation of one of the main challenges of mitochondrial genetics, i.e., the mechanisms of the mutation generation and the reasons for the high evolution rate of mtDNA, is available.

The present study was focused on the reconstruction of the mutational spectra of the mtDNA coding sequences, as well as on the analysis of the distribution of nucleotide substitutions. These data are important for further analysis of the mutational processes within human mitochondrial genome.

MATERIALS AND METHODS

We examined 794 human mtDNA sequences located between the nucleotide positions 577 and 16023 (15445 bp in size) obtained from different sources, including 560 mtDNA sequences from the MitoKor database (www.mitkor.com/science/560mtdnasrevision.php) [4], 42 sequences from the GenBank database (AF382013.1-AF381981.1) [2], and 192 sequences reported in [3]. The rCRS sequence ([10]; GenBank NC 001807), also available in the MITOMAP database (www.mitomap.org) was used as a reference sequence. To analyze the mutations distribution within the mtDNA sequences and to determine the nucleotide content, software package MEGA version 2.1 was used [11].

Mutational spectra of the mtDNA genes were reconstructed relative to the L-chains of the ancestral

1022-7954/05/4101-0079 © 2005 Pleiades Publishing, Inc.

Table 1. Nucleotide differences between the Cambridge Reference Sequence (rCRS) and the ancestral sequence of human mtDNA

MtDNA variants	Nucleotide differences from rCRS
L(x L1b/c, L2, L3, M, N)	1048T 2758A 2885C 4312T 6185C 7146G 8468T 10589A 11914A 12007A 13105G
L(x L2, L3, M, N)	825A 8655T 10688A 10810C 13506T 15301G
L(x L3,M, N)	3594T 4104G 7256T 13650T
N	8701G 9540C 10398G 10873C 15301A
R	12705T
HV + preHV	11719A
HV	14766T
H	2706G 7028T
H2	1438G 4769G
rCRS polymorphism	750G 8860G 15326G

sequences identified earlier by phylogenetic analysis of mitochondrial genome variation data [2–4]. The order of emergence and the patterns of nucleotide substitutions were scored from the topology of human mtDNA phylogenetic tree (median network), represented, according to the data from [2–4], by the mtDNA clusters (haplogroups and subhaplogroups) prevailing in Africa (L1a/d, L1b, L1c, L1e, L2a, L2b, L3b, L3d, and L3e), Western Eurasia (H, HV, pre-V, J1, J2, K, U2, U3, U4, U5a, U5b, U6, U7, U8, U*, X, I1, I2, W1, W2, W*, T*, T1, T2, and N1b), and Eastern Eurasia (A, B1, B2, C, Z, D*, D5, E, G2, M*, M9a, M1, and R*). In accordance with the classification, mtDNA haplogroups are denoted by single Roman letters (excluding haplogroup HV), and subhaplogroups within the haplogroups, by digits added to the letter code of the group [12]. The mtDNA types that belong to a particular haplogroup, yet cannot be attributed to any of the known subhaplogroups, are designated by asterisks (e.g., T = T* + T1 + T2, or T* = T(x T1, T2), where x is the elimination sign.

Statistical significance of the differences between the mutation frequencies within mutational spectra of the mtDNA genes was evaluated using *t*-test (STATISTICA/w5.0 software package).

The bias in the nucleotide frequency distributions within mtDNA regions (*B*) was measured using the formula [13]:

$$B = |(G + T) - (A + C)|/(A + C + G + T).$$

In the absence of mutational asymmetry (A + C = G + T), the expected *B* values are close to zero [13].

Comparative analysis was performed using mutational spectra for the HVS1 and HVS2 of the main noncoding region of human mtDNA. These spectra were reconstructed earlier with the help of phylogenetic analysis of the population data on variability of coding and noncoding regions of mitochondrial genome [14].

RESULTS AND DISCUSSION

In the studies of human mtDNA, mutations are usually registered relative to the reference sequence, which is the Cambridge sequence of mtDNA (rCRS), representing the first mitochondrial genome, whose nucleotide sequence was published in 1981 [6] (as well as a revision of this human mtDNA sequence [10]). Since rCRS occupies the peripheral position on mtDNA phylogenetic tree (as it belongs to one of the subgroups (H2) of the large European haplogroup H) [3], spectra of mtDNA genes should be reconstructed relative to the ancestral sequences, occupying the most ancient nodes of the mtDNA phylogenetic tree. According to the phylogenetic analysis data, the most ancient is the node separating the root mtDNA type of African subhaplogroups L1a/d and the root type, which is the ancestral for most of human mtDNA haplogroups [4]. In Table 1, this node is defined as L (x L1b/c, L2, L3, M, N). The data in Table 1 show that rCRS reference sequence differs from the ancestral one in the mutations at 36 positions, which makes up only 0.2% of 15 445 nucleotides examined. Hence, the error would not be large if mutational spectra of the mtDNA genes are reconstructed relative to the rCRS rather than the ancestral sequence. Nevertheless, in the present study mutational spectra of the mtDNA genes were reconstructed relative to the ancestral mtDNA sequence (Table 1), identified through analysis of median networks presented in [2–4].

The data on the variable nucleotide frequency distribution in the mtDNA genes and intergenic regions are presented in Table 2. To reconstruct mutational spectra, a phylogenetic approach was used. It provides identification of parallel mutations independently arising in different mtDNA clusters during the evolution [14]. For this reason, the data on the mutation frequencies in Table 2 are presented taking into consideration their homoplasy. It can be seen that all mtDNA segments examined are characterized by the broad range of variable nucleotide frequency variation (from 4.9 to 16.9%). The genes for ribosomal and transport RNAs are the most conservative. The genes for cytochrome-c-oxidase (subunits I and II) and NADH-dehydrogenase (subunits 4 and 4L) are the most conservative among the protein-encoding genes. The most variable are the *ATPase6*, *ATPase8*, *ND6*, and *CytB* genes. Analysis of the short noncoding intergenic regions variability showed that the average variable nucleotide frequency, typical of these segments, was similar to the corresponding frequency in moderately-variable coding mtDNA sequences (11.3%). However, certain intergenic regions of mitochondrial genome exhibit higher variability. These are the regions 5577–5586 and 8270–8294, characterized by the presence of 30 and 40% of

Table 2. Frequency distribution of variable nucleotides and nucleotide substitutions within the genes and noncoding intergenic regions of human mtDNA

MtDNA segments	Location	Length	Number of variable positions (frequency, %)	Number of mutations/number of mutations per variable position
12S	648–1601	954	56 (5.87)	80/1.43
16S	1671–3229	1559	76 (4.87)	103/1.36
ND1	3307–4262	956	102 (10.67)	127/1.25
ND2	4470–5511	1042	110 (10.56)	145/1.32
COI	5904–7445	1542	134 (8.69)	179/1.36
COII	7586–8269	684	62 (9.06)	71/1.15
ATPase8	8366–8572	207	35 (16.91)	38/1.09
ATPase6	8527–9207	681	91 (13.36)	104/1.14
COIII	9207–9990	783	82 (10.47)	104/1.27
ND3	10059–10404	346	38 (10.98)	50/1.32
ND4L	10470–10766	297	21 (7.07)	30/1.43
ND4	10760–12137	1378	131 (9.51)	162/1.24
ND5	12337–14148	1812	200 (11.04)	244/1.22
ND6	14149–14673	525	70 (13.33)	96/1.37
CytB	14747–15887	1141	153 (13.41)	209/1.37
22 tRNA genes	Different	1509	93 (6.16)	107/1.15
Noncoding regions	Different	168	19 (11.31)	25/1.32
GVS1	16093–16365	273	202 (73.99)	2212/10.95
GVS2	72–297	226	92 (40.71)	500/5.43

variable nucleotide positions, respectively. These values are close to those observed in the hypervariable segments of the main noncoding region (over 40% in HVS2 and over 70% in HVS1) [14]. However, the HVS1 and HVS2 mutational spectra are characterized by extremely high (>5 mutations) values of mutational constraint (i.e., the number of independent mutations per each variable position). At the same time, in the mutational spectra of mutations observed in the genes and short intergenic regions of mtDNA the values of this index did not exceed 2 (Table 2). Interestingly, maximum mutational constraint values (1.43) were observed in the most conservative mtRNA genes (*12S* rRNA and *ND4L*).

The data on the distribution of nucleotide substitutions within mutational spectra of the mtDNA coding sequences are reported in Table 3. The ratio between transitions and transversions was shown to be 19.3 : 1 in the rRNA genes, 15 : 1 in the protein-encoding genes, and, on average, 14.3 : 1 in the tRNA genes. Similar ratio between transitions and transversions was observed in HVS1 of the main noncoding region (13.4 : 1) [14]. Mutational spectra of noncoding hypervariable segments of mtDNA are characterized by substantial prevalence of pyrimidine transitions over the purine ones, 3.3 : 1 in HVS1 and 1.7 : 1 in HVS2 [14]. The spectra of the mtDNA coding sequences, however, have a quite different pattern, i.e., nearly equal proportions of these substitutions (0.9 : 1, on average). It was shown that the degree of variation of certain nucleotide position types (A, G, T, or C) statistically significantly ($P = 0.04$) correlated with the nucleotide content only in the tRNA genes. The coefficient of correlation for the protein-encoding genes was $r = 0.8$ ($P = 0.2$), and for rRNA genes it was $r = 0.4$ ($P = 0.6$). In the protein-encoding genes the proportion of variable C-nucleotides was statistically significantly higher, while the proportion of variable G-nucleotides was statistically significantly lower than expected on the basis of the nucleotide content ($P < 0.0002$ in both cases). In the rRNA genes an increase of the variability in T-positions along with its decrease in A-positions was observed ($P = 0.0006$ and 0.029, respectively). Comparison of the nucleotide content and the mutational spectra of the protein-encoding genes showed that the frequency of mutations at all nucleotide types, except T, was not consistent with the nucleotide content ($P < 0.03$). In the rRNA genes nucleotide content was not consistent with the mutational spectra for all nucleotide types ($P < 0.02$), and especially for G-nucleotides (the average content of the latter in the rRNA genes constituted 18.2%, while their proportion among variable positions was 25%, and the proportion of mutations, falling on G bases in the mutational spectrum was as high as 35%). In the tRNA genes, however, only in C-positions the proportion of mutations

Table 3. The spectra of nucleotide substitutions in human mitochondrial genome genes

Mutations	*n*	*m*	*m* – *n*	*m*/*n*
Protein-encoding genes (G11.3%, T25.3%, C33%, A30.4%)				
C ⟶ T	295	315	20	1.07
T ⟶ C	298	401	103	1.35
G ⟶ A	203	343	140	1.69
A ⟶ G	345	412	67	1.19
C ⟶ A	26	27	1	1.04
A ⟶ C	7	7	0	1.0
A ⟶ T	16	16	0	1.0
T ⟶ A	8	9	1	1.13
T ⟶ G	10	10	0	1.0
G ⟶ T	3	3	0	1.0
C ⟶ G	19	19	0	1.0
G ⟶ C	7	7	0	1.0
rRNA genes (G18.2%, T21.8%, C26%, A34%)				
C ⟶ T	27	30	3	1.11
T ⟶ C	44	51	7	1.16
G ⟶ A	23	62	39	2.7
A ⟶ G	30	31	1	1.03
C ⟶ A	2	2	0	1.0
A ⟶ C	1	1	0	1.0
A ⟶ T	2	2	0	1.0
T ⟶ A	1	1	0	1.0
T ⟶ G	1	1	0	1.0
G ⟶ T	1	1	0	1.0
C ⟶ G	0	0	0	0
G ⟶ C	1	1	0	1.0
tRNA genes (G14.5%, T26.9%, C23.4%, A35.2%)				
C ⟶ T	12	12	0	1.0
T ⟶ C	29	32	3	1.1
G ⟶ A	13	17	4	1.31
A ⟶ G	32	39	7	1.22
C ⟶ A	2	2	0	1.0
A ⟶ C	0	0	0	0
A ⟶ T	3	3	0	1.0
T ⟶ A	0	0	0	0
T ⟶ G	1	1	0	1.0
G ⟶ T	0	0	0	0
C ⟶ G	1	1	0	1.0
G ⟶ C	0	0	0	0

Note: *n*, the number of polymorphic positions; *m*, the number of mutations; *m*/*n*, mutational constraint, showing the mean number of mutations per polymorphic position. In brackets is the nucleotide content.

was statistically significantly lower ($P = 0.03$), than it could be expected based on the nucleotide content.

Asymmetry of the DNA chain nucleotide content is a characteristic feature of mitochondrial genome [6]. As shown in the present study, in different genome regions the nucleotide frequency distribution bias (*B*) values varied within the wide range and constituted, on average, 26.8% in the protein-encoding genes, 20% in the rRNA genes, and 17.2% in the tRNA genes.

The lowest *B* value was observed for HVS2 of the main noncoding region (14%), while the highest value (39.4%) was observed in the HVS1. In the protein-encoding genes, however, both higher (over 40% in the *ATP8* and *ND6* genes) and lower *B* values (14 to 18% in the *COI*, *COIII*, and *ND3* genes) were observed. It is important to note that in the nuclear genes the nucleotide content bias is expressed to a much lower extent ($B = 4.3\%$ for 13870 gene [15]). In human mtDNA, the nucleotide content bias is mostly caused by a low frequency of guanine bases (Table 3). However, exactly guanine residues are subjected to highest mutational pressure (*m*/*n* constitutes 1.7 and 2.7 for the protein- and rRNA-encoding genes, respectively).

The appearance of the DNA chain asymmetry may be related to the processes of transcription and replication [15–17]. The mtDNA replication is asynchronous and transcription-dependent, since it is initiated by the synthesis of the short RNA segment within the D-loop (the main noncoding region), which serves as a primer for subsequent synthesis of the daughter H-chain of mtDNA [18]. The mtDNA replication asynchrony is manifested in the fact that during this process, the parental H-chain long remains single-stranded. For this reason, it may be more prone to mutation, than the L-chain [19, 20]. One of the possible mechanisms of generation of mutations (including those in mtDNA) is deamination of cytosines, leading to C : G ⟶ T : A transitions [21]. Hence, the probability of the C ⟶ T mutations is higher in the H-chain region, which remains single-stranded during replication [19]. This region includes hypervariable segments and all mtDNA genes, excluding the *12S* rRNA, *16S* rRNA, *ND1*, and *ND2* genes along with the genes for some tRNAs located between the OriH and OriL (between positions 577 and 5721). In the mutational spectra reconstructed relative to the L-chain, C ⟶ T mutations that occurred on the H-chain are registered as G ⟶ A. Thus, if the initial proposal is true, it is expected that the *12S* rRNA, *16S* rRNA, *ND1*, and *ND2* genes on the L-chain, first, contain more guanine bases, and second, exhibit lower G-nucleotide variability than other mtDNA genes.

The guanine content distribution in mtDNA genes and hypervariable segments along with the data on G ⟶ A mutation intensity are presented in Table 4. The mitochondrial genome segments compared were shown to differ in guanine prevalence. The average frequency of G-nucleotides in the first segment, including the *12S* rRNA, *16S* rRNA, *ND1*, and *ND2* genes, was

higher than in the second segment (14.4% versus 11.4%, $P = 0$). The average frequency of variable guanine bases was lower in the first segment (9.9% versus 19.0%, $P = 0$). Variable G-nucleotide frequencies, however, were different for different genes with the minimal value observed for the *16S* rRNA gene (3%). At the same time, in general, the frequency value range for the rRNA, *ND1*, and *ND2* genes (3 to 15.2%) overlapped with that for the other mtDNA genes (9.6 to 38.5%). Maximum values of variable G-nucleotide frequencies were observed in some mtDNA genes, the H-chain of which remains single-stranded during replication. These are the genes *ATPase6* and *ATPase8* (positions 8366–9207), and also *ND6* and *CytB* (positions 14 149–15 887). However, the intensity of parallel mutations at G-positions (the *m/n* index in Table 4) was the highest both in the four genes of the first segment (*12S* rRNA, *16S* rRNA, *ND1*, and *ND2*; positions 648–5511) and in the genes of the opposite mitochondrial genome segment (*ND3*, *ND4L*, and *ND4*; positions 10 059–12 137). Taken together, these results still do not conclusively show that the increased mutational pressure on the guanine bases observed in the L-spectra of the mtDNA genes exclusively results from the specific features of the mtDNA replication along with the mechanisms intensifying cytosine deamination at the stage of the appearance of single-stranded regions in the H-chain. This can be explained by the fact that the mtDNA segments compared statistically significantly differed only by the average frequencies of G-nucleotides and the mutations occurred therein. On the other hand, the proportions of the C ⟶ T mutations on the H-chain and the G ⟶ A mutations on the L-chain in different mtDNA genes are unknown. Consequently, the contribution of the H-chain C ⟶ T mutations to the differences in the L-spectra of the genes from the first and second groups is also unknown.

Transcription of the mtDNA chains also has its specific features. Among the protein-encoding genes, only the *ND6* is transcribed from the L-chain [18]. Because of this, based on the hypothesis on transcription-associated mutational asymmetry of the DNA chains [17], it can be expected that mutational L-spectrum of this gene will be different from those of the neighboring *ND5* and *CytB* genes in the increased frequency of the G ⟶ A transitions caused by the intense cytosine deamination in the H-chain remaining single-stranded during transcription.

Similarly, mutational spectra of the genes transcribed from the H-chain should differ from the *ND6* spectrum by higher frequency of the C ⟶ T transitions. Finally, taking into consideration different probability of the appearance of the G ⟶ A mutations in different genes due to specific features of the mtDNA replication and transcription, it can be expected that the G ⟶ A frequency will be the highest in the *ND6* gene, while it will be the lowest in the rRNA, *ND1*, and *ND2* genes and intermediate in the other mtDNA genes. However, the analysis showed the absence of statistically significant differences between mutational spectra of the *ND6* and the neighboring genes ($P > 0.2$ for the C ⟶ T and G ⟶ A transitions).

Table 4. Distribution of the G-nucleotides content, variable positions (*n*), and the G ⟶ A mutations (*m*) in different segments of mtDNA

MtDNA segment	Length, bp	Proportion of the G bases (%)	*n* (%)	*m*	*m*/*n*
HVS1	273	25 (9.1)	16 (64.0)	195	12.19
HVS2	226	37 (16.4)	18 (48.6)	73	4.06
(**OriH**)					
12S	954	182 (19.1)	15 (8.2)	38	2.53
16S	1559	270 (17.3)	8 (3.0)	24	3.0
ND1	956	112 (11.7)	17 (15.2)	36	2.12
ND2	1042	99 (9.5)	13 (13.1)	29	2.23
OriL					
COI	1542	250 (16.2)	24 (9.6)	39	1.63
COII	684	102 (14.9)	16 (15.7)	24	1.5
ATPase8	207	13 (6.3)	5 (38.5)	6	1.2
ATPase6	681	71 (10.4)	19 (26.8)	26	1.37
COIII	783	116 (14.8)	18 (15.5)	28	1.56
ND3	346	37 (10.7)	7 (18.9)	13	1.86
ND4L	297	36 (12.1)	4 (11.1)	9	2.5
ND4	1378	136 (9.9)	13 (9.6)	26	2.0
ND5	1812	192 (10.6)	26 (13.5)	42	1.62
ND6	525	37 (7.0)	10 (27.0)	14	1.4
CytB	1141	137 (12.0)	31 (22.6)	51	1.65

Note: OriH and OriL divide the mtDNA genes into two segments.

In conclusion, the obtained results suggest that such characteristic feature of mitochondrial genes as the L-spectra bias towards the G ⟶ A transitions is not exclusively determined by the specificity of the mtDNA replication and transcription. As seen in Table 4, some genes (*12S* rRNA, *16S* rRNA, *ND4*, and *ND4L*), characterized by relatively low G-nucleotide variability (3 to 11%) display high frequency of parallel mutations (2.0 to 3.0 per variable position). This can be explained by the dependence of mutational processes on the DNA context. Earlier, it was demonstrated that context-related mechanisms of mutations generation were among the most important mechanisms of mutagenesis of the main noncoding region of mtDNA [14, 22]. Hence, analysis of the context properties of the mtDNA genes seems important for further investigations of the mechanisms underlying the emergence of mutations in mitochondrial genome.

ACKNOWLEDGMENTS

The author thanks I.B. Rogozin (Institute of Cytology and Genetics, Siberian Branch of the Russian

Academy of Sciences), M.V. Derenko (Institute of the Biological Problems of the North, Far Eastern Branch of the Russian Academy of Sciences), and I.A. Kutuev (Institute of Biochemistry and Genetics, Ufa Research Center of the Russian Academy of Sciences) for their help in this study and for discussion of the results.

This work was supported by the Far Eastern Branch of the Russian Academy of Sciences (grant no. 03-3-A-06-096).

REFERENCES

1. Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U., Mitochondrial Genome Variation and the Origin of Modern Humans, *Nature*, 2000, vol. 408, pp. 708–713.
2. Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., *et al.*, Major Genomic Mitochondrial Lineages Delineate Early Human Expansions, *BMC Genet.*, 2001, vol. 2, p. 13.
3. Finnila, S., Lehtonen, M.S., and Majamaa, K., Phylogenetic Network for European mtDNA, *Am. J. Hum. Genet.*, 2001, vol. 68, pp. 1475–1484.
4. Herrnstadt, C., Elson, J.L., Fahy, E., *et al.*, Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian and European Haplogroups, *Am. J. Hum. Genet.*, 2002, vol. 70, pp. 1152–1171.
5. Hedges, S.B., A Start for Population Genomics, *Nature*, 2000, vol. 408, pp. 652–653.
6. Anderson, S., Bankier, A.T., Barrel, B.G., *et al.*, Sequence and Organization of the Human Mitochondrial Genome, *Nature*, 1981, vol. 290, pp. 457–465.
7. Mishmar, D., Ruiz-Pesini, E., Golik, P., *et al.*, Natural Selection Shaped Regional mtDNA Variation in Humans, *Proc. Natl. Acad. Sci. USA*, 2003, vol. 100, pp. 171–176.
8. Moilanen, J.S. and Majamaa, K., Phylogenetic Network and Physicochemical Properties of Nonsynonymous Mutations in the Protein-Coding Genes of Human Mitochondrial DNA, *Mol. Biol. Evol.*, 2003, vol. 20, pp. 1195–1210.
9. Meyer, S. and von Haeseler, A., Identifying Site-Specific Substitution Rates, *Mol. Biol. Evol.*, 2003, vol. 20, pp. 182–189.
10. Andrews, R.M., Kubacka, I., Chinnery, P.F., *et al.*, Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA, *Nat. Genet.*, 1999, vol. 23, p. 147.
11. Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M., *MEGA2: Molecular Evolutionary Genetics Analysis Software*, Arizona State Univ. Press, 2000.
12. Macaulay, V., Richards, M., Hickey, E., *et al.*, The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs, *Am. J. Hum. Genet.*, 1999, vol. 64, pp. 232–249.
13. Sueoka, N., Intrastrand Parity Rules of DNA Base Composition and Usage Biases of Synonymous Codons, *J. Mol. Evol.*, 1995, vol. 40, pp. 318–325.
14. Malyarchuk, B.A., Rogozin, I.B., Berikov, V.B., and Derenko, M.V., Analysis of Phylogenetically Reconstructed Mutational Spectra in Human Mitochondrial DNA Control Region, *Hum. Genet.*, 2002, vol. 111, pp. 46–53.
15. Majewski, J., Dependence of Mutational Asymmetry on Gene-Expression Levels in the Human Genome, *Am. J. Hum. Genet.*, 2003, vol. 73, pp. 688–692.
16. Tanaka, M. and Ozawa, T., Strand Asymmetry in Human Mitochondrial DNA Mutations, *Genomics*, 1994, vol. 22, pp. 327–335.
17. Green, P., Ewing, B., Miller, W., *et al.*, Transcription-Associated Mutational Asymmetry in Mammalian Evolution, *Nat. Genet.*, 2003, vol. 33, pp. 514–517.
18. Clayton, D.A., Replication and Transcription of Vertebrate Mitochondrial DNA, *Annu. Rev. Cell Biol.*, 1991, vol. 7, pp. 453–478.
19. Reyes, A., Gissi, C., Pesole, G., and Saccone, C., Asymmetrical Directional Mutation Pressure in the Mitochondrial Genome of Mammals, *Mol. Biol. Evol.*, 1998, vol. 15, pp. 957–966.
20. Tamura, K., On the Estimation of the Rate of Nucleotide Substitution for the Control Region of Human Mitochondrial DNA, *Gene*, 2000, vol. 259, pp. 189–197.
21. Beletskii, A. and Bhagwat, A.S., Transcription-Induced Mutations: Increase in C to T Mutations in the Nontranscribed Strand during Transcription in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA*, 1996, vol. 93, pp. 13 919–13 924.
22. Malyarchuk, B.A., The Effect of the Nucleotide Context on Induction of Mutations in Hypervariable Segment 1 of Human Mitochondrial DNA, *Mol. Biol.* (Moscow), 2002, vol. 36, no. 3, pp. 418–423.