Russian Journal of Genetics, Vol. 41, No. 3, 2005, pp. 301–305. Translated from Genetika, Vol. 41, No. 3, 2005, pp. 385–390.

HUMAN GENETICS

The Role of Nucleotide Context in the Induction of Mutations 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—Based on the mutations distribution patterns in the mitochondrial DNA (mtDNA) genes, context analysis of the regions, including mutable positions characterized by the appearance of more than two parallel mutations, was performed. It was demonstrated that the mechanism of dislocation mutagenesis, leading to the appearance of mismatches within the frameshift regions of either primer or template mtDNA chains during replication, accounts for the induction of 21% of unstable positions in the mtDNA genes. Context analysis showed that pyrimidine bases in the positions +1 and +2 (gYRNS, gYY, and gR consensuses, where g is mutable position) had the highest influence on the induction of mutations in G positions of the mtDNA genes. The highest effect on the mutagenesis in T positions was excreted by the bases in the positions –1 and +1 (RtY and tA consensuses, where t is mutable position). In general, these data point to the prevalence of the context-dependant mechanisms of the mutations induction in human mitochondrial genome.

INTRODUCTION

Investigations of mutational spectra of the main noncoding region of human mitochondrial DNA (mtDNA) showed that the appearance of mutations within hypervariable segments (HVS1 and HVS2) of this region mostly depended on the DNA context (i.e., specific features of the DNA primary structure) in the vicinity of variable positions [1]. The process of the mutations appearance is irregular, since, as demonstrated in a number of studies that the rates of induced and spontaneous mutations are increased only in certain DNA positions, called the mutational "hotspots" [2–5]. Analyses of the mutational spectra in human samples with the inherited diseases have demonstrated that neighboring nucleotide bases, located ±2 nucleotides apart from the variable regions, had a pronounced influence on the generation of mutations at these positions [2, 3]. Although the effect of more distant bases on the mutagenesis cannot be excluded, the neighboring bases are the most probable candidate agents affecting the mutation rates [6]. The effect of the DNA context on the mutagen-induced mutagenesis was observed during the analyses of mutational spectra in *Escherichia coli* [6, 7] and mammals [8].

The DNA context effects on mutagenesis (both spontaneous and induced) can be manifested at each stage of the DNA replication and repair processes [4, 9]. It is suggested that the appearance of mismatches during replication can be determined by specific features of the template–primer duplexes primary structure within the active DNA region, polymerase complex [10]. Neighboring bases can stabilize the mismatches, thereby accelerating the mutation rate [3]. The importance of dislocation mutagenesis, causing the appearance of mismatches (and, consequently, mutations) as a result of a frameshift of primer or template DNA chains in the monotonous nucleotide tracts during replication [11], should also be considered. Analysis of the mutational spectra in the main noncoding region of human mtDNA showed that dislocation mutagenesis model could explain the appearance of more than 25% of hotspots in HVS1 and HVS2. Within the mtDNA HVS1 the hotspot consensus sequences, CC and KTNCNK (hotspots are underlined) have been described [1]. Thus, the data on the mutational spectra of the mtDNA main noncoding region indicated that within this region context-dependant mechanisms of the mutation generation prevailed.

In the present study, using the context analysis of the nearest nucleotide neighbors of the most variable positions of mtDNA genes, mutational spectra of the mtDNA genes encoding for two rRNAs, twenty-two tRNAs, and thirteen respiratory chain protein subunits were examined.

MATERIALS AND METHODS

We analyzed mutational spectra of human mtDNA genes that were obtained based on the data [12–14] on variability of 794 15445-bp sequences (located between nucleotide positions 577 and 16023, according to Cambridge Reference Sequence [15]). Mutational spectra were reconstructed by use of median-network method, based on maximum parsimony algorithm [16]. Mutational spectrum is defined as a frequency distribution of the appearance of independent mutations of each type (transitions, transversions, deletions, and insertions) in

1022-7954/05/4103-0301

each nucleotide position of the DNA fragment examined [4].

To describe the distribution patterns of nucleotide substitutions in mtDNA in accordance with their variability, statistical methods (nonparametric statistics and distribution fitting) as implemented in the STATISTICA 5.0 (StatSoft, Inc., Tulsa, United States) software package were used. Mutational spectra were analyzed using methodical approaches described in [1, 6]. To investigate the role of the neighboring bases in the appearance of nucleotide substitutions, DNA regions located ±5 bases from the variable positions (mutable motifs) were examined. These regions were manually aligned relative to variable positions (mutable positions) of a certain type (A, G, C, and T) analyzed. Mutable motif consensus sequences were generated for the variable bases G and T, i.e., substitutions for the C and A bases were converted into the complementary forms. Consensus nucleotides in mutable motifs were designated using generally accepted single-letter nucleotide coding nomenclature: N = A, T, G; C; W = A or T; S = G or C; R = A or G; Y = T or C; M = A or C; K = G or T; B = T or G, or C; V = A or G, or C; H = A or T, or C; D = A or T, or G [17].

RESULTS AND DISCUSSION

Analysis of the rRNA, tRNA and protein-encoding genes sequence variations (with the summarized size of 15 366 bp) showed that 1454 positions within mtDNA coding sequence (9.4% of the length of mtDNA segment examined) were variable. Application of the method of phylogenetic analysis enabled distinguishing of the parallel mutations, which independently arose in one and the same nucleotide positions, but at different stages of the mtDNA evolution and on different brunches of its phylogenetic tree. For these reasons, taking into consideration the parallel mutations, mutational spectra of the mtDNA genes are characterized by 1849 independent mutational events, i.e., on average, 1.3 mutations per each position. Note, however, that the value of this index vary within a broad interval, constituting up to 14 parallel mutations per one position. It was demonstrated that the distribution pattern of the mtDNA positions depending upon their variability was geometrical ($P < 0.01$). This means that more than 90% of the mtDNA coding regions are monomorphic, in 7% of the positions single mutations are observed, and only the remaining less than 3% of the positions are characterized by homoplasy (from 2 to 14 parallel mutations per position).

Nucleotide positions characterized by the increased mutation rate are the most interesting in respect of identification of the role of the DNA context factors in mutagenesis [1, 4]. For these reasons, further analysis was performed using mtDNA fragments which included mutable positions characterized by the emergence of more than two parallel mutations per position. Table 1 demonstrates the results of the context analysis performed at a distance of ±5 bases from variable G nucleotides on the L-chain of mtDNA. The regions, where C nucleotides are variable, are recorded in the complementary form in accordance with the mtDNA H-chain sequence. The data show that all mutable motifs, without exclusions, can be subdivided into several types of short consensus sequences: gYRNS, gYY, and gR (g, mutable position). It should be noted that most of the sequences, described by consensus gR, are of the gRY type (9 out of 12 sequences). Substitutions in regions gY arise almost three times more frequently, compared to gR ($P = 0.003$), while in gC regions they are statistically significantly more frequent than in gT, or in gR ($P = 0.03$). Substitutions in the regions gNY are also statistically significantly more frequent than in gNR ($P = 0.003$), while in gNC (but not in gNT) the substitutions are statistically significantly more frequent than in gNR ($P = 0.02$). Since no context regularities were revealed in the regions 5' to the mutable positions, the main role of the pyrymidine bases in the positions +1 and +2 on the generation of mutations in G positions of the mtDNA genes can be suggested. The model of DNA chains dislocation during replication can explain the origin of unstable positions only in five out of forty-two segments analyzed (12%).

The results of the analysis of T-nucleotide mutable motifs are presented in Table 2. Interestingly, in this case dislocation mutagenesis model is able to explain the generation of mutations in 32.4% of mutable positions. Context analysis of the position +1 showed that mutable motifs split into two main groups, tY and tA (t, mutable position). Analysis of position +2 showed that consensus tNY was statistically significantly more widely distributed than tNR ($P = 0.0004$). In position –1 adenine and guanine residues were statistically significantly more frequent (the At frequency was substantially higher compared to other sequence variants; $P = 0.03$). Thus, mutable motifs of T positions can be described with the help of two consensuses, RtY and tA (Table 2). In this case, it is also evident that nearest neighboring bases in positions –1 and +1 have the highest effect on the mutagenesis in T positions of the mtDNA genes.

Distribution of the mutations within the segments of protein-encoding genes largely depends on such evolutionary factor as selection, determined by different functional importance of nucleotide substitutions in different codon positions. Earlier studies showed that 32% of the mutations within mtDNA protein-encoding genes were non-synonymous. Furthermore, only a quarter of all parallel mutations within the coding sequences led to amino acid substitutions [18]. Analysis of mutable motifs carried out in the present study, also showed that parallel mutations in 30.3% of the positions resulted in amino acid substitutions. Among these, mutations in positions 3316, 13 708, and 5460 (6, 7, and 8 independent substitutions of alanine for threonine, respectively), and also in positions 14 766 (six substitutions of isoleucine for threonine) and 14 470 (six

Table 1. Context DNA analysis of the mtDNA segments characterized by variable G bases

Position	Number of parallel mutations	MtDNA sequence											
14364	3	C	C	A	C	A	g	C	A	C	C	A	
15110	4	T	G	C	T	T	g	C	A	A	C	T	
12346H	5	A	G	T	G	T	g	C	A	T	G	G	
13708	7	G	C	C	T	G	g	C	A	G	C	C	
3705	4	G	C	A	C	T	g	C	G	A	G	C	
15043	4	A	T	C	G	G	g	C	G	A	G	G	
9449H	3	A	T	C	C	C	g	T	A	T	C	G	
10373	3	T	A	T	G	A	g	T	G	A	C	T	
3666	3	T	C	A	G	G	g	T	G	A	G	C	
							g	Y	R		S		
11176	3	A	A	C	C	A	g	C	C	A	G	A	
8251	3	A	T	A	G	G	g	C	C	C	G	T	
1888	4	G	G	A	G	A	g	C	C	A	A	A	
15884	4	A	A	T	G	G	g	C	C	T	G	T	
9554	4	C	A	C	T	G	g	C	C	C	C	C	
1719	6	C	C	T	T	A	g	C	C	A	A	A	
3316	6	C	C	A	T	G	g	C	C	A	A	C	
5460	8	T	C	A	T	C	g	C	C	C	T	T	
5046	3	T	A	G	C	A	g	T	T	C	T	A	
11914	7	A	C	C	A	C	g	T	T	C	T	C	
709	14	T	C	C	C	C	g	T	T	C	C	A	
12630	3	A	C	A	T	G	g	T	C	C	A	T	
3918	4	G	G	G	G	A	g	T	C	C	G	A	
3438	3	T	A	C	G	G	g	C	T	A	C	T	
6023	4	G	C	C	G	A	g	C	T	G	G	G	
14569	4	A	C	A	C	C	g	C	T	A	A	C	
							g	Y	Y				
9266	3	A	C	A	G	G	g	G	C	C	C	T	
2706	3	G	G	C	A	T	g	A	C	A	C	A	
12372	3	A	C	C	C	T	g	A	C	T	T	C	
15257	3	C	A	G	T	A	g	A	C	A	G	T	
5147	5	A	C	C	A	C	g	A	C	C	C	T	
8697	3	C	A	A	A	T	g	A	T	A	A	C	
15930	3	C	C	G	G	A	g	A	T	G	A	A	
3010	6	A	T	C	C	C	g	A	T	G	G	T	
10685	4	G	C	A	G	C	g	G	T	G	G	G	
3915	3	G	A	A	G	G	g	G	A	G	T	C	
12007	5	C	A	A	T	G	g	G	G	C	T	C	
1438	4	A	A	A	C	T	g	A	G	A	G	T	
							g	R					
7337	3	G	C	T	T	C	g	A	A	G	C	G	d
1598	4	T	G	G	A	C	g	A	A	C	C	A	d
6455H	4	C	A	G	A	C	g	A	A	G	A	G	d
930	3	T	A	G	A	A	g	C	C	G	G	C	d
6260	5	G	T	G	G	A	g	G	C	C	G	G	d

Note: Presented are the mtDNA L-chain sequences in the 5'–3' direction. The cases, when complementary mtDNA H-chains are analyzed, are marked by the addition of letter "H" to the position numbers. Consensuses are indicated. MtDNA segments, instability of which is explained by dislocation mutagenesis model are marked by letter d.

Table 2. Context DNA analysis of the mtDNA segments characterized by variable T bases

Position	Number of parallel mutations	MtDNA sequence											
14766	6	C	A	A	A	A	t	T	A	A	C	C	
13105H	3	G	A	A	G	A	t	T	C	C	T	G	
14233H	3	T	A	T	G	A	t	T	A	T	G	G	
15236H	3	T	C	A	G	A	t	T	C	A	T	T	
15758H	3	T	C	C	G	A	t	T	C	A	G	G	
14769H	3	G	G	G	G	G	t	T	A	G	T	T	
15924H	7	T	C	C	G	G	t	T	T	A	C	A	
13020	3	T	T	A	G	G	t	C	T	C	C	A	
10915	3	A	G	C	T	G	t	T	C	C	C	C	
9677H	3	G	G	T	T	G	t	T	T	T	C	T	
14798	3	A	C	T	C	A	t	T	C	A	T	C	
5442	4	C	C	C	C	A	t	T	C	C	T	C	
10084	4	A	A	T	A	A	t	C	A	A	C	A	
						R	t	Y					
4216	4	T	A	T	G	A	t	A	T	G	T	C	
6671	3	T	C	C	C	A	t	A	T	T	G	T	
3394	3	T	A	G	G	C	t	A	T	A	T	A	
10598H	4	G	T	A	G	C	t	A	T	A	A	T	
6827	3	T	C	C	G	C	t	A	C	C	A	T	
4561	3	C	T	G	A	G	t	A	G	G	C	C	
8705	3	A	A	C	C	A	t	A	C	A	C	A	
14180	3	C	A	A	T	A	t	A	T	A	C	A	
14182	3	A	T	A	T	A	t	A	C	A	C	C	
15670	4	C	T	C	C	A	t	A	T	A	T	C	
							t	A					
12414	4	A	A	C	C	C	t	A	A	C	A	A	d
6152	3	C	T	A	G	T	t	C	C	C	C	T	d
6221	5	T	T	A	C	C	t	C	C	C	T	C	d
8618	3	A	T	T	G	A	t	C	C	C	C	A	d
14308	3	C	A	G	C	T	t	C	C	T	A	C	d
14470	6	G	T	A	T	A	t	C	C	A	A	A	d
7055H	3	A	C	A	G	C	t	C	C	T	A	T	d
9545H	5	T	G	C	C	C	t	C	C	T	A	A	d
11719H	5	G	T	A	A	G	t	C	C	G	T	G	d
6680	3	G	T	A	A	C	t	T	A	C	T	A	d
15784	8	T	A	C	C	C	t	T	T	T	A	C	d

substitutions of serine for praline) were most frequent. The role of multiple amino acid substitutions in the mitochondrial genome evolution is still unclear, although some of these substitutions are known to be associated with hereditary diseases. For instance, among "fast" positions mentioned above, only the mutation in position 13 708 belongs to the secondary pathological mutations associated with Leber's disease [19].

The data obtained indicate that variability of the mitochondrial genome genes, as well as of its main noncoding region is caused by the effects of nucleotide context on the mutation generation. Consensus regions of variable mtDNA positions are rather heterogeneous, pointing to the diverse mechanisms of spontaneous mutagenesis in human mitochondria. One of these, the mechanism of dislocation mutagenesis, leads to the appearance of mismatches in the frameshift regions of primer or template mtDNA chains during replication. It can explain the emergence of 21% of unstable positions within mtDNA genes. The value obtained is comparable with the results of the analogous examination of the hotspots within hypervariable segments of the main noncoding region [1]. Note, that the data obtained point to the most important role of the nearest neighboring nucleotides in the generation of mutations within mtDNA genes. The possible influence of nucleotides located more than five positions apart from mutable positions, however, cannot be excluded, though an analysis of this kind was not carried out in the present study. It is also noteworthy that relatively high proportion of the positions susceptible to the influence of dislocation mutagenesis also results from their closeness, or affiliation with the short monotonous repeats, indicating the important role of the context-dependant mutagenesis in the generation of mutations in human mitochondrial genome.

ACKNOWLEDGMENTS

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

REFERENCES

1. 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.

2. Cooper, D.N. and Youssoufian, H., The CpG Dinucleotide and Human Genetic Disease, *Hum. Genet.*, 1988, vol. 78, pp. 151–155.

3. Krawczak, M., Ball, E.V., and Cooper, D.N., Neighboring-Nucleotide Effects on the Rates of Germ-Line Single-Base-Pair Substitution in Human Genes, *Am. J. Hum. Genet.*, 1998, vol. 63, pp. 474–488.

4. Rogozin, I.B. and Pavlov, Y.I., Theoretical Analysis of Mutation Hotspots and Their DNA Sequence Context Specificity, *Mutat. Res.,* 2003, vol. 544, pp. 65–85.

5. Morton, B.R., Oberholzer, V.M., and Clegg, M.T., The Influence of Specific Neighboring Bases on Substitution Bias in Noncoding Regions of the Plant Chloroplast Genome, *Mol. Evol.*, 1997, vol. 45, pp. 227–231.

6. Rogozin, I.B., Berikov, V.B., Vasunina, E.A., and Sinitsina, O.I., The Effect of the Primary Structure of DNA on Induction of Mutations by Alkylating Agents, *Rus. J. Genet.*, 2001, vol. 37, no. 6, pp. 704–710.

7. Broschard, T.H., Koffel-Schwartz, N., and Fuchs, R.P.P., Sequence-Dependent Modulation of Frameshift Mutagenesis at NARI-Derived Mutation Hot Spots, *J. Mol. Biol.*, 1999, vol. 288, pp. 191–199.

8. Shibutani, S., Suzuki, N., Tan, X., *et al.*, Influence of the Flanking Sequence Context on the Mutagenicity of Acetylaminofluorene-Derived DNA Adducts in Mammalian Cells, *Biochemistry*, 2001, vol. 40, pp. 3717–3722.

9. Zavolan, M. and Kepler, T.B., Statistical Inference of Sequence-Dependent Mutation Rates, *Curr. Opin. Genet. Dev.*, 2001, vol. 11, pp. 612–615.

10. Timsit, Y., DNA Structure and Polymerase Fidelity, *J. Mol. Biol.*, 1999, vol. 293, pp. 835–853.

11. Kunkel, T.A. and Soni, A., Mutagenesis by Transient Misalignment, *J. Biol. Chem.*, 1988, vol. 263, pp. 14 784–14 789.

12. 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.

13. Finnila, S., Lehtonen, M.S., and Majamaa, K., Phylogenetic Network for European mtDNA, *Am. J. Hum. Genet.*, 2001, vol. 68, pp. 1475–1484.

14. 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.

15. 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.

16. Bandelt, H.-J., Forster, P., Sykes, B.C., and Richards, M.B., Mitochondrial Portraits of Human Populations Using Median Networks, *Genetics*, 1995, vol. 141, pp. 743–753.

17. Cornish-Bowden, A., Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences: Recommendation, *Nucleic Acids Res.*, 1985, vol. 13, pp. 3021–3030.

18. 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.

19. Wallace, D.C., Mitochondrial Diseases in Man and Mouse, *Science,* 1999, vol. 283, pp. 1482–1488.