ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 6, pp. 702-706. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © B. A. Malyarchuk, 2011, published in Biokhimiya, 2011, Vol. 76, No. 6, pp. 858-863.

Adaptive Evolution Signals in Mitochondrial Genes of Europeans

B. A. Malyarchuk

Institute of Biological Problems of the North, Far-Eastern Branch of the Russian Academy of Sciences, ul. Portovaya 18, 685000 Magadan, Russia; fax: (4132) 634-463; E-mail: malyarchuk@ibpn.ru

Received December 27, 2010
Revision received February 1, 2011

Abstract—Since modern Europeans appear to be descendants of the Late Pleistocene European peoples who survived the last glacial period, it is quite reasonable to expect the presence of adaptive genetic variants that originated in the Ice Age in the modern gene pool of Europeans. To find such adaptive variants, mitochondrial genomes have been analyzed of the modern population from Eastern and Central Europe belonging to haplogroups U4, U5, and V, that diversified during the Late Pleistocene and Holocene periods. Analysis of distribution of nonsynonymous and synonymous substitutions, as well as results of search for radical amino acid changes that arose under the influence of adaptation (positive destabilizing selection) allowed us to detect signals of molecular adaptation in different mitochondrial genes and haplogroups of mtDNA. However, there were very few strong adaptive signals ($z > 3.09$, $P < 0.001$) that could be due to the loss of adaptive mtDNA haplotypes during the Holocene warming.

DOI: 10.1134/S0006297911060113

Key words: human mitochondrial genome, adaptive evolution, selection

The human mitochondrial genome consists of circular DNA molecules 16,569 nucleotide pairs (bp) in size and containing genes encoding two rRNAs, 22 tRNAs, and 13 subunits of the respiratory chain proteins [1]. The main functional elements required for transcription and replication of mtDNA are concentrated in the major noncoding region, and ~70% of the mitochondrial genome consists of genes encoding proteins that are subunits of protein complexes of the mitochondrial respiratory chain. Mitochondrial proteins are involved in functioning of four out of five oxidative phosphorylation complexes of mitochondria – these are seven subunits of NADH dehydrogenase complex I (ND1, 2, 3, 4, 4L, 5, and 6), one subunit (cytochrome *b*) of the cytochrome bc_1 complex III, three subunits of cytochrome *c* oxidase complex IV (CO1, 2, and 3), and two subunits of the ATPase complex V (AT6 and 8) [2].

Data on the variability of mtDNA in human populations are widely used in phylogeographic studies directed to reconstruction of the genetic history of ethnoracial groups, but little attention is paid to the problem of adaptive evolution of the mitochondrial genome [3-7]. This can be associated with the evolutionary conservativeness of the genes encoding components of the mitochondrial electron transport chain [8-11]. Therefore, any deviation of mitochondrial genes from neutral evolution always attracts attention, first of all because of a possible contribution of such molecular changes to adaptation caused by various changes (climatic, nutritional, environmental, etc.).

Assessment of the number of nonsynonymous substitutions in different genes of mtDNA is one of the most adequate approaches for elucidating the role of selection in the evolution of the mitochondrial genome [12]. Usual approaches are based on analyzing the ratio of the number of nonsynonymous substitutions by nonsynonymous site to the number of synonymous substitutions by synonymous site or on analysis of the ratio between the number of nonsynonymous substitutions by synonymous ones in phylogenetic clusters of DNA. Under conditions of selective neutrality, it is expected that these parameters for nonsynonymous substitutions must correspond to those for synonymous ones; in the case of prevalence of the nonsynonymous number over the synonymous ones the effect of positive selection is supposed, whereas in the opposite case the effect of negative (purifying) selection is supposed.

Studies on the variability of human mtDNA genes revealed the prevalence of the number of synonymous substitutions over the number of nonsynonymous ones, and this suggested a great importance of negative selection for functioning of the mitochondrial genome [3, 6,

7]. However, some researchers supposed that the selective pressure could vary in different mtDNA genes and, moreover, such differences could be a consequence of the adaptation of humankind to changes in the environment. Thus, Mishmar et al. [5] have shown than in the population of the Arctic zone ATPase subunit 6 (AT6) is the most variable, whereas in the Temperate zone cytochrome *b* (CYTB) is the most variable, and in the Tropical zone the highest variability is displayed by subunit 1 of cytochrome *c* oxidase (CO1). It was supposed in this and the subsequent works that appearance of a great number of variants of mtDNA polymorphism should be caused by adaptation of human populations to environmental conditions, first of all, to the cold of high latitudes during the last glacial period [4, 5]. It was also supposed that on the warming these earlier adaptive variants of polymorphism could become somewhat pathologic [4, 5]. However, other studies on mtDNA failed to detect variability or interregional differences in the rate of nonsynonymous mutations [6, 7]. Thus, it should be noted that results of studies on adaptive evolution of the human mitochondrial genome are very contradictory.

This work presents results of analysis of distribution of nonsynonymous and synonymous substitutions in groups of phylogenetically related haplotypes of mtDNA (haplogroups and their subgroups) in the European population. Data were analyzed on the variability of total mitochondrial genomes of haplogroups U4, U5, and V, which are specific for the population of Eastern and Central Europe. Carriers of these haplogroups survived the last glaciations in refuges in the South of Europe, but during the Holocene period they occupied territories to the North as they became free of ice [3]. Therefore, it is quite reasonable to expect signals of adaptive evolution in the mtDNA genes of Europeans. The purpose of this work was to search for such molecular changes.

MATERIALS AND METHODS

Characteristics of molecular data. Analysis of nonsynonymous and synonymous substitutions can be performed only on representative samples; therefore, the study was carried out only within the limits of the best studied haplogroups U4, U5, and V, with sample sizes of 93, 213, and 66 genomes, respectively. For the analysis, previously published data were used on the full-genomic variability of mtDNA in populations of Eastern and Central Europe [13-15], as well as data of other authors for populations of Southern and Northern Europe [16, 17]. Using a phylogeographic approach, it was established that carriers of the U4 and U5 haplogroups had survived the last Wurm glaciation period (with maximum at 18-25 thousand years ago) in refuges in the South of Europe (in the Pyrenees, the Balkan Peninsula, and the South of Eastern Europe), and thus are descendants of the Upper Pleistocene population of Europe [3, 13-15]. Haplogroup V arose in Europe (possibly in the Pyrenees) already in the Holocene (about 11 thousand years ago), but its sources indicate to haplogroup HV0, the carriers of which also survived in Southern Europe refuges [3, 14].

Phylogenetic and statistical analysis of data. Nucleotide sequences of mitochondrial genomes and statistical and phylogenetic analysis of molecular data were leveled using the MEGA 4.0.2 program [18]. The phylogenetic trees were designed using a Neighbor-Joining (NJ) method based on *p*-distances between the DNA sequences, which were calculated from the number of nucleotide substitutions per position in the comparison in pairs. The mitochondrial genome of chimpanzee (No. X93335 in the GenBank database) was used as the external group.

To assess the effect of natural selection on mtDNA variability, the distribution of nonsynonymous (NS) and synonymous (S) mutations was studied in the groups of substitutions associated with haplogroups (H) and unique (or private (P)) substitutions in the terminal branches of the phylogenetic tree according to the approach proposed by Elson et al. [6] and Ruiz-Pezini et al. [4]. In both cases the analysis was based on comparison of the NS/S ratios in the H and P groups using a precise Fisher test. The neutrality index NI was determined as the ratio $(NS/S)_P/(NS/S)_H$. Statistical analysis was performed using the mtPhyl program (http://eltsov.org). The evolutionary age of mtDNA clusters was calculated also with the mtPhyl program for the mutation rate determining one substitution in the encoding region of mtDNA over the period of 5140 years [5].

Searching for molecular adaptation on the level of individual amino acids. To reveal adaptive changes in the mtDNA genes, changes in physicochemical features of amino acids were analyzed during the evolution (i.e. following the topology of the phylogenetic NJ-tree) using the TreeSAAP 3.2 program [19]. The algorithm of this program allowed us to compare the observed changes in physicochemical features of amino acids (altogether 31 features were analyzed) in the phylogenetic tree of mtDNA with the expected distribution based on the hypothesis of the random character of amino acid substitutions under conditions of selective neutrality. The *z*-test can be used to assess the significance of amino acid substitutions in eight magnitude categories (mc) and also to define the selection type. According to an MM01 model used in the TreeSAAP 3.2 program, on detecting the positive selection in the most radical significance categories (mc = 6, 7, and 8; $z > 3.09$; $P < 0.001$), the features of amino acids are supposed to be changed under the influence of positive destabilizing selection [10, 11]. Such selection results in the most radical amino acid substitutions; therefore, the detection of this selection is evidence of molecular adaptation; radical substitutions supported by such selection change the structure and functions of

proteins in a definite direction and thus promote the better adaptation of the organism. According to the MM01 model, conservative amino acid substitutions (mc = 1, 2, and 3; $z > 3.09$; $P < 0.001$) suggest the influence of positive stabilizing selection [10].

RESULTS AND DISCUSSION

Statistical analysis of distribution of synonymous and nonsynonymous substitutions in haplogroups U4, U5, and V revealed that for haplogroups U4 and V values of the NI parameter were <1 (0.91 and 0.5, respectively) and >1 for U5 (NI = 1.19). However, in all cases, deviations to either side were insignificant ($P > 0.2$), and the results indicated a neutral character of evolution in the mtDNA haplogroups analyzed in Europeans (Table 1). The hypothesis of selection influence on the mitochondrial gene pool of Europeans was also tested on the level of individual genes encoding proteins. However, this study also indicated the absence of a significant influence of any selection type on mtDNA genes within the limits of the haplogroups studied.

Table 1. Assessment of the influence of natural selection on mtDNA variability

Haplogroups of mtDNA and their subgroups	Size (in genomes)	Value of neutrality index NI	Statistical significance of differences (P)
V	66	0.5	0.2
V1	22	1.12	0.8
U4	93	0.75	0.4
U4a	56	0.56	0.3
U4a1	21	2.25	0.4
U4a2	33	0.17	0.09
U4b	16	0.5	0.6
U4c	8	0.42	0.9
U4d	13	4.5	0.28
U5	213	1.19	0.3
U5a	120	1.18	0.4
U5a1	67	1.05	0.6
U5a1a	22	0.71	0.8
U5a1b	23	0.82	0.8
U5a1c	7	0	0.4
U5a1d	7	8.33	0.1
U5a2	53	1.56	0.3
U5a2a	22	1.2	0.7
U5a2b	22	0.92	0.7
U5b	93	1.22	0.4
U5b1	56	11.79	0.004
U5b2	33	0.41	0.1

Note: It is suggested that in the absence of selection the neutrality index NI has values close to 1.0; at NI > 1.0 the effect of negative (purifying) selection is expected, at NI < 1.0 the effect of positive selection is expected.

Because mtDNA haplogroups consist of subgroups with regionally specific distribution, i.e. are associated with definite geographical population groups, the further analysis was performed on the level of mtDNA subgroups. Table 1 shows that negative selection has a highly significant influence on subgroup U5b1 (NI = 11.79, $P = 0.004$) in only one case. The neutrality indices were high also for subgroups U5a1d and U4d (NI = 8.3 and 4.5, respectively), but these values were unreliable. Within limits of the used statistical approaches, the negative selection was a consequence of the prevalence of nonsynonymous substitutions in the terminal (evolutionarily younger) branches of the phylogenetic tree as compared to the stalks, which determined the mtDNA clusters. Note that the three mentioned subgroups of mtDNA have a high evolutionary age: 24.8 ± 7.4 thousand years for U4d, 19.1 ± 6.1 thousand years for U5a1d, and 18.8 ± 5.2 thousand years for U5b1. This means that nonsynonymous substitutions were accumulated during the time period corresponding to the transition from the glaciation maximum (18-25 thousand years ago) to the current interglacial period. However, it is unclear what caused the increase in the accumulation of nonsynonymous substitutions in the terminal branches of these clusters of mtDNA – either the weakening of negative selection cutting the substitutions capable of changing structural–functional organization of protein molecules, or the necessity for adaptation to the warming, that, by contrast, could require changes in amino acid sequences.

The only example suggesting a possibility of the influence of positive selection (and, consequently, of adaptation) on the mtDNA cluster is presented by subgroup U4a2, for which a very pronounced (but unreliable) deviation of neutrality was shown (NI = 0.17, $P = 0.09$) (Table 1). Within our approaches, it was suggested that the selection should be favorable for accumulation of nonsynonymous substitutions at all stages of haplogroup formation, in both the stalks and the branch terminals. However, the age of subgroup U4a2 is only 6.5 ± 1.4 thousand years; therefore, the weakening of the negative selection in this case occurred during a relatively warm period.

It seems very likely that during the glaciation maximum mitochondrial genomes of humans were already functioning in the optimal regimen for those climatic conditions. This was promoted by both the selection against cold-unadapted genotypes and arising of adaptive variants. It was supposed that the mitochondrial gene pool of Europeans should undergo corresponding changes on warming, because the previously useful genetic variants could become deleterious and new adaptive

Table 2. Analysis of changes in physicochemical features of amino acid substitutions associated with mtDNA haplogroups of different evolutionary age

Haplogroup of mtDNA	Age, thousand years	Amino acid substitution	Gene	Positive selection ($P < 0.001$), significance categories, and type of changes in amino acid features	
				stabilizing	positive destabilizing (adaptation)
U4	21.2	M316T	*Cytb*	–	6, pH_i
U4c	24.0	F50L	*ND4*	2, C_α	–
U5	31.0	V91I	*CO3*	1, P_β	–
U5a	20.0	H16R	*Cytb*	–	–
U5a1	19.2	T158A	*Cytb*	–	6, pH_i
U5a1a1b	19.0	H4Y	*ND5*	–	7, α_m
U5b2	28.5	Q434R	*ND5*	–	–
U5b2a	20.8	N88S	*ND2*	–	–
U5b2b	24.0	I100V	*ND5*	–	–
U5b2b	24.0	T432A	*ND5*	–	–
U4a1	12.3	M201V	*ND5*	–	–
U4c1	11.1	S182A	*AT6*	–	–
U5a1b1c	10.0	A177T	*AT6*	–	–
U5a1b	10.3	N154S	*CO3*	–	–
U5a2a	8.5	S531T	*ND5*	–	–
V1	9.8	I57T	*ND2*	–	–
V1	9.8	M115V	*AT6*	–	–

Note: The following changes in physicochemical features of amino acids are indicated: isoelectric point (pH_i), contact region of the helix (C_α), β-structure (P_β), α-helix (α_m). A line means the absence of significant ($P < 0.001$) changes in the amino acid features.

variants of mtDNA would be necessary [4, 5]. No doubt, such a scenario could be tested the best by comparing mitochondrial genomes of contemporary humans and our ancestors from different epochs. Thus, the recent study of the mtDNA variability in the collared lemming, an Arctic animal from contemporary and fossil populations, revealed that during the last 25 thousand years the genetic variety sharply decreased; moreover, some genetic variants disappeared (e.g. of the cytochrome *b* gene), which could be important for adaptation during the glaciations [20].

However, there is also another explanation for the accumulation of nonsynonymous substitutions in the terminal branches of phylogenetic trees. This process was earlier shown to depend on the age of mtDNA clusters: the number of nonsynonymous substitutions increased with a decrease in the evolutionary age [7, 21]. Consequently, the negative selection directed to stabilize the structure of mitochondrial genes was operating at all stages of mtDNA evolution, but the nearer it was to the present time, the higher was the rate of nonsynonymous substitutions that simply had not yet passed all stages of the purifying selection. In this case further selection would be directed to elimination of haplotypes carrying nonsynonymous substitutions. However, this hypothesis is based on the idea that all arising nonsynonymous substitutions are unfavorable, but in fact this is not so. Thus, it is still unclear what is the ratio between the number of nonsynonymous substitutions in the terminal clusters of mtDNA that are unfavorable for the present (i.e. which appeared at the stage of adaptation to cold) and favorable now (i.e. which appeared during the warming).

Results of molecular dating of phylogenetic clusters of mtDNA within the limits of haplogroups U4, U5, and V have indicated that the evolutionary age of these groups is usually not more than 30 thousand years (Table 2). Many mtDNA clusters are determined by nonsynonymous substitutions, which are the most interesting for studies on molecular adaptation (Table 2). To assess the adaptive importance of molecular changes that occurred at different stages of formation of mtDNA haplogroups, we used TreeSAAP analysis, which allowed us to detect radical changes in amino acids resulting in changes in physicochemical features of proteins. The most radical changes seemed to indicate the influence of adaptation, i.e. positive destabilizing selection [11]. The analysis revealed only three radical substitutions (Table 2). One of these substitutions appeared concurrently with formation of haplogroup U4 specific for populations of Eastern Europe and Western and Southern Siberia, whereas two

substitutions appeared concurrently with formation of subgroups U5a1 and U5a1a1b. These subgroups are characteristic mainly for the population of the Eastern and Central Europe. Subgroups U5a1 and U5a1a1b arose at the time of maximal decrease in temperature; therefore, the detected radical substitutions seemed to indicate adaptive changes in the genes of cytochrome *b* and *ND5* in Europeans during the glaciation period. Note that both amino acid substitutions in cytochrome *b* are located in the loop regions of the protein, which are much more conservative than the transmembrane regions [8]. The substitution T158A occurred in the *cd*2-helix of the *cd*-loop of the Q_0-redox site of cytochrome *b* directly interacting with cytochrome c_1 encoded by the nuclear genome [11]. This region is extremely important for functioning of the respiratory chain; therefore, amino acid substitutions arising in it can be significant for adaptation [9-11]. The substitution M316T in the *fg*-loop of the functionally important cytochrome *b* template domain is also very important [11].

Thus, analysis of nucleotide sequences of entire mitochondrial genomes in European populations revealed adaptive variants in different mitochondrial genes and haplogroups of mtDNA, which, according to molecular dating, appeared during the last glacial period. However, such adaptation signals are very rare, and it is very likely that adaptive variants of mtDNA were lost during the Holocene as the temperature was warming, but this hypothesis needs to be tested by comparison of genetic materials of contemporary and ancient inhabitants of Europe.

This work was supported by the Far-Eastern Branch of the Russian Academy of Sciences (project No. 09-3-A-06-221).

REFERENCES

1. Anderson, S., Bankier, A. T., Barrel, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. R., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. M., Staden, R., and Young, I. G. (1981) *Nature*, **290**, 457-467.
2. Saraste, M. (1999) *Science*, **283**, 1488-1493.
3. Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H. J., Torroni, A., and Richards, M. B. (2010) *Curr. Biol.*, **20**, 174-183.
4. Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., and Wallace, D. C. (2004) *Science*, **303**, 223-226.
5. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A., and Wallace, D. C. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 171-176.
6. Elson, J. L., Turnbull, D. M., and Howell, N. (2004) *Am. J. Hum. Genet.*, **74**, 229-238.
7. Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P. A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L. L., and Oefner, P. J. (2006) *Genetics*, **172**, 373-387.
8. Irwin, D. M., Kocher, T. D., and Wilson, A. C. (1991) *J. Mol. Evol.*, **32**, 128-144.
9. Da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., and Antunes, A. (2008) *BMC Genomics*, **9**, e119.
10. McClellan, D. A., and McCracken, K. G. (2001) *Mol. Biol. Evol.*, **18**, 917-925.
11. McClellan, D. A., Palfreyman, E. J., Smith, M. J., Moss, J. L., Christensen, R. G., and Sailsbery, J. K. (2005) *Mol. Biol. Evol.*, **22**, 437-455.
12. Kimura, M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5773-5777.
13. Malyarchuk, B. A., Grzybowski, T., Derenko, M., Perkova, M., Vanecek, T., Lazur, J., Gomolcak, P., and Tsybovsky, I. (2008) *Mol. Biol. Evol.*, **25**, 1651-1658.
14. Malyarchuk, B., Derenko, M., Denisova, G., and Kravtsova, O. (2010) *Mol. Biol. Evol.*, **27**, 2220-2226.
15. Malyarchuk, B., Derenko, M., Grzybowski, T., Perkova, M., Rogalla, U., Vanecek, T., and Tsybovsky, I. (2010) *PLoS ONE*, **5**, e10285.
16. Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A. S., Bandelt, H. J., Semino, O., and Torroni, A. (2005) *Am. J. Hum. Genet.*, **76**, 883-886.
17. Finnila, S., Lehtonen, M. S., and Majamaa, K. (2001) *Am. J. Hum. Genet.*, **68**, 1475-1484.
18. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) *Mol. Biol. Evol.*, **24**, 1596-1599.
19. Woolley, S., Johnson, J., Smith, M. J., Crandall, K. A., and McClellan, D. A. (2003) *Bioinformatics*, **19**, 671-672.
20. Prost, S., Smirnov, N., Fedorov, V. B., Sommer, R. S., Stiller, M., Nagel, D., Knapp, M., and Hofreiter, M. (2010) *PLoS ONE*, **5**, e10447.
21. Derenko, M. V., and Malyarchuk, B. A. (2010) *Molecular Phylogeography of Population of the North Eurasia by Data on Variability of Mitochondrial DNA* [in Russian], North-Eastern Research Center, Far-East Branch, Russian Academy of Sciences, Magadan.