Rate variation between mitochondrial domains and adaptive evolution in humans

Max Ingman* and Ulf Gyllensten

Department of Genetics and Pathology, Rudbeck Laboratory, University of Uppsala, Uppsala, Sweden

Received February 27, 2007; Revised and Accepted July 4, 2007

It has been proposed that as a result of human adaptation to different climates, mitochondrial genes have been affected by natural selection. To further study the selective pressure on human mitochondrial DNA, we have analysed polymorphism at the gene, domain and nucleotide site level in four geographic regions. The ratio of non-synonymous relative to synonymous substitutions is elevated for ATP6 in North Asia, ND3 in Africa and cytb in Europe relative to other mitochondrial genes. In addition, non-synonymous substitutions appear nearly five times more frequently outside core functional domains as compared with within functional domains. Therefore, a large part of the rate variation between mitochondrial genes is explained by differences in the length of the functional domain relative to the length of the gene. The accumulation of non-synonymous substitutions in the ATP6 gene among sequences from North Asia has been a gradual process over many thousands of years, consistent with relaxed purifying selection rather than recent strong selective sweeps. It is not necessary to invoke adaptive selection to explain the pattern of polymorphism in mitochondrial genes, when the most parsimonious explanation is relaxed purifying selection and the action of genetic drift.

INTRODUCTION

Mitochondria function in the oxidative phosphorylation (OXPHOS) pathway that uses dietary calories to produce ATP, the energy source for many cellular processes. The majority of gene products that are engaged in this function are encoded from the nuclear genome and transported to the mitochondrial vesicles in the cytoplasm, but 13 are encoded by mitochondrial DNA (mtDNA). mtDNA has a nucleotide substitution rate roughly 10 times higher than that of a typical nuclear gene (1) and has a very high degree of polymorphism. Interest in mtDNA polymorphism comes from several fields. A number of human diseases have been associated with mutations in the mitochondrial genome, indicating that dysfunction of the components of oxidative phosphorylation encoded by the mitochondrial genome can be deleterious (2). In order to test the hypothesis that common mtDNA variation influences human physiology and disease, Saxena et al. (3) studied all mitochondrial SNPs with a frequency >1% in Europeans for their association with type 2 diabetes. No association was found to either disease or metabolic variables and they failed to identify any physiological effect of alleles that were previously proposed to have been adaptive for energy metabolism in human evolution. It appears that some polymorphisms that have previously been identified as disease causing have a weak effect at best.

Interest in the functional role of mtDNA polymorphism also comes from the field of population genetics and evolutionary biology. mtDNA is widely used in studies of human evolution in which selective neutrality is often assumed. While all mitochondrial lineages of a species are under the same selective pressure, deviations from the assumption of selective neutrality would not be expected to affect genetic analyses of the relationship among lineages, or the degree of differentiation of lineages or populations. If selection for a specific amino acid occurs on some lineages but not others, unequal evolutionary rates on different branches of a phylogenetic tree may result, causing erroneous conclusions regarding the evolutionary relationships and time-scale for demographic events.

Finally, the extensive polymorphism in mtDNA has made it a suitable marker in forensic medicine for individual identification. In addition to the wealth of polymorphism, high copy number per cell and maternal inheritance contribute to the usefulness of mtDNA in such studies. It has been argued that forensic analyses should focus on polymorphism that is not disease-associated or otherwise functionally important.

*To whom correspondence should be addressed. Tel: +46 18 471 4969; Fax: +46 18 471 4931; Email: max.ingman@genpat.uu.se
such that no undesired information regarding the disease susceptibility of the investigated individuals is obtained (4,5). Also, the sites investigated should not be hotspots for mutation, such that the risk of somatic heterogeneity, heteroplasm and mutational differences between maternal relatives is low. Consequently, for forensic studies it is important to understand the selective constraints on mtDNA polymorphisms in order to select appropriate polymorphic sites and correctly interpret results.

Studies of the determinants of mtDNA polymorphism have been based on several methods of analysis. In a recombining system, the null hypothesis of selective neutrality has usually been tested by comparing the ratio of the number of non-synonymous changes per such site (ka) with synonymous changes per such site (ks). Under selective neutrality, ka is expected to be equal to ks, and if ka is higher than ks, positive selection is invoked, while a lower ka is indicative of purifying selection (6). Complications arise when applying this test to the analysis of mtDNA as the zero (7,8), or very low (9), recombination rate means that if a substitution is under positive selection then every other substitution on that lineage will be carried with it, complicating the identification of the site under selection. Also, since mtDNA lineages show strong geographic structuring, the polymorphism within lineages differs substantially from that observed between lineages. Not taking into account the evolutionary age of the polymorphism may lead to erroneous conclusions regarding the effect of selection.

A number of previous studies have addressed the pattern of selection on mtDNA coding regions using different data sets and methods, all with their inherent limitations. Analyses of the entire human mtDNA (10) have demonstrated a preponderance of ks over ka among humans, indicating that all genes encoded by the mitochondrial genome are subject to purifying (negative) selection maintaining their structure and function. Mishmar et al. (11) have proposed that varying selective pressures in different geographic regions have influenced the evolution of human mtDNA, in particular mtDNA lineages from ‘Arctic’ environments. Since ATP is essential for the maintenance of body temperature, a cold environment could select for advantageous non-synonymous variants (12). It was noted that the ATP6 gene is less conserved than other mitochondrial genes and has a higher ka/ks ratio among North Asian (denoted ‘Arctic’) lineages than lineages from other regions. Similar results were described for cytochrome b in European (‘temperate’) lineages and cytochrome oxidase I among African (‘tropical’) lineages. The conclusion by Mishmar et al. (11) has been criticized on several grounds. First, many of the pairwise sequence comparisons used to calculate ka and ks involved division by zero requiring a constant to be added to each estimate, therefore introducing a bias. Also, the use of Fisher’s exact test and Wilcoxon’s test have been criticized and when applied appropriately on the original data have not yielded any significant results (12). Lastly, the extent to which the samples used by Mishmar et al. (11) represent the different geographic regions has also been questioned.

Revisiting the issue of selection on mtDNA lineages, Elson et al. (14) could not detect a difference in the ka/ks ratio between lineages from different geographic regions. They noted that in a phylogenetic tree, polymorphisms found on terminal branches have a higher ka/ks ratio than those on internal nodes. This effect may confound comparison of data sets that differ with respect to the evolutionary age of the polymorphism. As an alternative to balancing or positive selection, Elson et al. (14) suggested that ATP6 has been subject to a relaxation in purifying selection in European and Asian populations. Their analysis was based mainly on European lineages and lacked a suitable dataset for addressing selection on Arctic populations. Ruiz-Pesini et al. (12), following-up the study by Mishmar et al. (11) also noted the higher ka/ks ratio for internal versus terminal substitutions. In contrast to Elson et al., Ruiz-Pesini et al. interpret their data as evidence of positive selection on certain lineages. The issue of selection on the mtDNA genome was recently discussed by Kivisild et al. (15). They find no support for cold climate adaptation in mtDNA and agree with the conclusion that the higher ka/ks for some genes is most likely due to a relaxation in purifying selection. Again, this study lacks appropriate information for Arctic populations.

Here, we re-examine the claims for positive selection on mtDNA polymorphism with a data set and methodology that remedies some of the limitations of previous studies. This was achieved through (i) a geographically balanced data set of mtDNA genome sequences of sufficient and equal number from each region, (ii) a method for calculation of ka and ks that is less affected by pairwise comparisons between sequences that are very similar and (iii) a hierarchical analysis of the pattern of substitution at the level of the gene, the domain and the nucleotide, in order to explore different explanations for variation in substitution rate.

RESULTS

Generation of a geographically balanced data set
To complement previously published data, we have sequenced the complete mitochondrial genomes of 61 individuals from indigenous populations of northern Asia and assembled data sets of 104 complete mitochondrial coding region sequences from human populations of each of four regions: North Asia, South Asia, Europe and Africa. The new sequences from North Asia are derived from 14 populations (Canadian Inuit (n = 7), Chukchi (n = 8), Chukotkan Inuit (n = 6), Kazakh (n = 6), Kchanti (n = 4), Koryak (n = 4), Mansi (n = 4), Mongol (n = 4), Nanaitci (n = 3), Nivkchi (n = 4), Shortci (n = 3), Tuvinian (n = 4), Yakut (n = 4)). The limited sample size for each population restricts the inferences that can be drawn regarding individual populations, but the entire sample provides an overview of the mtDNA diversity in the region. The frequency distribution of haplogroups in the different regions is shown in Table 1.

Gradual and episodic modes of selection
An obstacle in the analysis of selection on mitochondrial lineages is that the age of the polymorphism may confound the result, such that a higher ka/ks ratio is expected among younger polymorphisms where purifying selection has had less time to act (14). We reconstructed a neighbour-joining
(16) tree using the complete coding sequences from North Asia, dated the nodes and compared the accumulation of ka and ks in two of the genes previously discussed (ATP6 and COI) among lineages that have diverged at various time points (Fig. 1). It has previously been claimed that ATP6 is evolving faster at the amino acid level than other mitochondrial genes, particularly among sequences from North Asia (11), while COI is one of the most highly conserved genes. A gene subjected to strong positive selection may be expected to show a rapid change in evolutionary rate during the time when selection occurred. Our results show that the accumulation of non-synonymous substitutions for ATP6 in North Asia is a gradual process over many thousands of years, indicative of relaxed purifying selection rather than the accelerated evolution that one might expect if certain variants were being driven to fixation by positive selection. COI, however, appears to be under relatively strict purifying selection. This method has limitations in that the expected result of gradual evolution involving several sites cannot be distinguished from a model where a single site is affected by strong selection or where most changes are deleterious but a few are adaptive. Nevertheless, the data for ATP6 do not seem to indicate a rapid increase in ka at any time point.

**Ka/ks varies 5-fold between domains of mitochondrial genes**

Calculation of the mean ka/ks for closely related sequences can be problematic due to the fact that a number of the pairwise comparisons will have no synonymous changes and therefore require division by zero. Also, in comparing mean pairwise sequence differences between groups, high frequency group-specific substitutions will be under-represented since no polymorphism will be present in the majority of comparisons. In the extreme case, a fixed group-specific non-synonymous change will not register at all in an estimate of within-group ka. For these reasons, we have used an alternative method that avoids both of these problems. We concatenated all sequences from each geographic region and calculated the number of synonymous and non-synonymous changes relative to a majority-rule consensus of all 416 sequences. These values constitute the mean ka and ks for each data set and by this approach we avoid the problems inherent to individual comparisons.

Previous studies have noted the highest ka/ks for ATP6, cytB, ND3 and ND5 (11). The highest ka/ks ratios in our 416 sequences are found for ATP6 and ND3, while cytB and ND5 have levels similar to those of the other genes (Fig. 2), indicating that different methods for calculating ka/ks have only a partial overlap in the genes with extreme values.

Since functional domains within each coding region may be under different types of selection than other parts of the genes, each mitochondrial gene was divided into two parts based on the Pfam annotations reported in Ensembl v36: (i) the part encoding the core functional domain(s) and (ii) the remainder of the gene including transmembrane sequences. On average, the non-synonymous sites outside the core functional domain evolve nearly five times faster than those within the functional domain (\(ka = 0.00156\) and 0.00032; \(ka/ks = 0.423\) and 0.085, respectively). Due to this large difference, the ka/ks values of the core functional domains of ATP6, ND3 and cytB are significantly lower than for their complete genes (\(P = 1.2 \times 10^{-8}\), \(P = 5.1 \times 10^{-11}\), \(P = 7.5 \times 10^{-10}\), respectively; Fig. 2). The higher density of non-synonymous substitutions in a specific region of these genes has a profound effect on the overall ka/ks for these

---

**Table 1.** Table showing the number of sequences from each of the four regions belonging to major mitochondrial haplogroups

<table>
<thead>
<tr>
<th></th>
<th>North Asia</th>
<th>South Asia</th>
<th>Africa</th>
<th>Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>0</td>
<td>1</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>L2</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>L3</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>52</td>
<td>15</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Z</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>52</td>
<td>88</td>
<td>10</td>
<td>101</td>
</tr>
<tr>
<td>R</td>
<td>24</td>
<td>82</td>
<td>8</td>
<td>89</td>
</tr>
<tr>
<td>UK</td>
<td>3</td>
<td>19</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>HV</td>
<td>5</td>
<td>11</td>
<td>2</td>
<td>48</td>
</tr>
<tr>
<td>JT</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WI</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>21</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 1.** Accumulation of non-synonymous substitutions over time. ka (black bars) and ks (open bars) for ATP6 (A) and COI (B) for North Asian clades that diverged at various time points.
genes. There is a negative relationship between \( \alpha/\kappa \) and length of the functional domain relative to gene length \( (r^2 = 0.61, P = 0.0016; \text{Fig. 3, Supplementary Material, Table S1}) \). To a large extent, the substitutions implicated in the putative adaptive evolution reported for these genes are located outside the core domain. The three genes implicated above (ATP6, ND3, and cytb) also have the smallest core domains relative to total gene length. The high \( \alpha/\kappa \) ratios observed for these three genes are likely to be due to relaxed purifying selection on sites where the need for amino acid conservation is lower. In order to test if this difference in degree of conservation is visible even in intraspecific comparisons, we compared sequences from five species (Human, Macaque, Dog, Mouse, Rat). The results show that for ATP6, ND3, and cytb, regions outside the core functional domain have a \( \alpha/\kappa \) ratio that is more than double that of the core functional domains (Table 2). Consequently, relatively relaxed purifying selection on regions outside the core domain is not specific to human mitochondrial sequences, but appears to be a general characteristic of these genes.

### Variation in the \( \alpha/\kappa \) ratio between geographic regions

In order to address the claim that the mitochondrial genome has adapted to differences in climate between geographic regions \((11, 12)\), we examined the three genes (ATP6, ND3, and cytb) with significant differences in \( \alpha/\kappa \) between the complete gene and the core functional domain(s) among the 104 sequences from each of the four regions. Consistent with some previous reports \((11)\), we find the \( \alpha/\kappa \) for the complete ATP6 gene to be highest among the North Asian sequences \((\chi^2 = 30.3, P < 0.0001; \text{Fig. 4A})\). ND3 among African sequences \((\chi^2 = 30.8, P < 0.0001; \text{Fig. 4B})\) and cytb among European sequences \((\chi^2 = 121.7, P < 0.0001; \text{Fig. 4C})\).

However, the \( \alpha/\kappa \) of the core functional domains among sequences from these geographic regions in these three genes is significantly lower than for the total gene \((P = 6.6 \times 10^{-3}, 1.3 \times 10^{-3}, 8.3 \times 10^{-4})\), respectively. Although the difference in \( \alpha/\kappa \) between the functional domains of these three genes is smaller than when considering the entire gene, there is still a significant increase observed between the geographic regions (ATP6, \( P = 0.006 \); ND3, \( P = 0.001 \); cytb, \( P < 0.0001 \)). This indicates that some positions within the core functional domain also contribute to differences in gene \( \alpha/\kappa \) levels between regions.

Examining the individual region-specific changes indicates that a similar pattern exists for each gene/region. Of the eight non-synonymous substitutions in the functional domain of ATP6 among North Asian sequences, five are singletons and two occur in two sequences. The last (np 8794 C → T; His → Tyr) is a defining mutation for haplogroup A and is found in 21 sequences in our North Asian data set, predominantly in Chukchi and Inuit populations in which it has almost reached fixation. Among the African ND3 sequences there are two positions with non-synonymous substitutions in the functional domain, one is a singleton and the other is present in eight sequences (np 10321 T → C; Val → Ala). In cytb, six positions have non-synonymous substitutions in the functional domain among Europeans, of which five are present in three sequences or less and the last (np 14798 T → C; Phe → Leu) is found in 18 sequences and is a defining mutation for haplogroup K, present at considerable frequency among European mtDNA sequences \((17)\). Outside the core functional domain, ATP6 has one high frequency North Asian-specific substitution (np 8584 G → A; Ala → Thr) at codon 20 in the transmembrane domain, which is a defining substitution for haplogroup C. One of the substitutions that delineate the macro-haplogroups M and N is also present in this part of the sequence.
DISCUSSION

In order to address claims for adaptation of the mitochondrial ATP-producing machinery to cold climate, we have studied mitochondrial genome variation on lineages from different geographic regions at the gene, domain and nucleotide site level. In every gene and region, the $k_a/k_s$ ratio is $<1$, supporting that purifying selection is acting on all mitochondrial genes. Consistent with some previous studies, we observe a set of genes with a somewhat higher $k_a/k_s$ ratio relative to other mitochondrial genes. In an attempt to understand whether the variability at these genes reflects recent positive selection, we determined the temporal mode of non-synonymous substitution for these genes. The increase in $k_a$ for ATP6 over the past 60,000 years is continuous (gradual) rather than episodic, while COI shows as very slow $k_a$ increase. Assuming that human populations colonized the far north of Asia during the last 10,000–20,000 years and became adapted to cold climate as a result of selection for mitochondrial substitutions, we would expect to observe the most dramatic increase in $k_a$ for ATP6 in the most recent time period rather than the continuous pattern observed. The data represent cumulative numbers that have lower resolution than using discrete non-overlapping categories, but nevertheless the result suggests a relatively gradual accumulation of substitutions in ATP6. The pattern of accumulation of changes in both genes is consistent with that expected through the action of purifying selection, although somewhat more relaxed in ATP6. The results do not allow us to distinguish between the above model of selection (gradual accumulation of extended periods of time) and alternative models of selection (e.g. a model in which most changes are deleterious, but a few are adaptive). More extensive modelling is needed to determine the different types of selection scenarios that can be distinguished using this methodology. However, for ATP6, the gradual increase in accumulation of amino acid changes spans a time period of up to 60,000 years, which essentially covers the full time-span since modern humans left Africa and is clearly much longer than the time that humans have been exposed to cold climate in association with the colonization of North Asia and the Americas. This indicates that the relatively high $k_a/k_s$ for ATP6 is due to relaxed purifying selection specific to this gene, rather than a selection pressure that has occurred in a specific geographic region due to climate.

In studying the distribution of non-synonymous changes within genes, we noted a striking difference in $k_a/k_s$ between the functional domain and the remaining part of the gene. This result points to the underlying complexity in studying the pattern of selection and provides further support that purifying selection is the main determinant of the $k_a/k_s$ ratio. The pattern of replacement of substitutions in ATP6. The pattern of accumulation of changes in both genes is consistent with that expected through the action of purifying selection, although somewhat more relaxed in ATP6. The results do not allow us to distinguish between the above model of selection (gradual accumulation of extended periods of time) and alternative models of selection (e.g. a model in which most changes are deleterious, but a few are adaptive). More extensive modelling is needed to determine the different types of selection scenarios that can be distinguished using this methodology. However, for ATP6, the gradual increase in accumulation of amino acid changes spans a time period of up to 60,000 years, which essentially covers the full time-span since modern humans left Africa and is clearly much longer than the time that humans have been exposed to cold climate in association with the colonization of North Asia and the Americas. This indicates that the relatively high $k_a/k_s$ for ATP6 is due to relaxed purifying selection specific to this gene, rather than a selection pressure that has occurred in a specific geographic region due to climate.

In studying the distribution of non-synonymous changes within genes, we noted a striking difference in $k_a/k_s$ between the functional domain and the remaining part of the gene. This result points to the underlying complexity in studying the pattern of selection and provides further support that purifying selection is the main determinant of the $k_a/k_s$ ratio. The pattern of accumulation of changes in both genes is consistent with that expected through the action of purifying selection, although somewhat more relaxed in ATP6. The results do not allow us to distinguish between the above model of selection (gradual accumulation of extended periods of time) and alternative models of selection (e.g. a model in which most changes are deleterious, but a few are adaptive). More extensive modelling is needed to determine the different types of selection scenarios that can be distinguished using this methodology. However, for ATP6, the gradual increase in accumulation of amino acid changes spans a time period of up to 60,000 years, which essentially covers the full time-span since modern humans left Africa and is clearly much longer than the time that humans have been exposed to cold climate in association with the colonization of North Asia and the Americas. This indicates that the relatively high $k_a/k_s$ for ATP6 is due to relaxed purifying selection specific to this gene, rather than a selection pressure that has occurred in a specific geographic region due to climate.

In studying the distribution of non-synonymous changes within genes, we noted a striking difference in $k_a/k_s$ between the functional domain and the remaining part of the gene. This result points to the underlying complexity in studying the pattern of selection and provides further support that purifying selection is the main determinant of the $k_a/k_s$ ratio. The pattern of accumulation of changes in both genes is consistent with that expected through the action of purifying selection, although somewhat more relaxed in ATP6. The results do not allow us to distinguish between the above model of selection (gradual accumulation of extended periods of time) and alternative models of selection (e.g. a model in which most changes are deleterious, but a few are adaptive). More extensive modelling is needed to determine the different types of selection scenarios that can be distinguished using this methodology. However, for ATP6, the gradual increase in accumulation of amino acid changes spans a time period of up to 60,000 years, which essentially covers the full time-span since modern humans left Africa and is clearly much longer than the time that humans have been exposed to cold climate in association with the colonization of North Asia and the Americas. This indicates that the relatively high $k_a/k_s$ for ATP6 is due to relaxed purifying selection specific to this gene, rather than a selection pressure that has occurred in a specific geographic region due to climate.

In studying the distribution of non-synonymous changes within genes, we noted a striking difference in $k_a/k_s$ between the functional domain and the remaining part of the gene. This result points to the underlying complexity in studying the pattern of selection and provides further support that purifying selection is the main determinant of the $k_a/k_s$ ratio. The pattern of accumulation of changes in both genes is consistent with that expected through the action of purifying selection, although somewhat more relaxed in ATP6. The results do not allow us to distinguish between the above model of selection (gradual accumulation of extended periods of time) and alternative models of selection (e.g. a model in which most changes are deleterious, but a few are adaptive). More extensive modelling is needed to determine the different types of selection scenarios that can be distinguished using this methodology. However, for ATP6, the gradual increase in accumulation of amino acid changes spans a time period of up to 60,000 years, which essentially covers the full time-span since modern humans left Africa and is clearly much longer than the time that humans have been exposed to cold climate in association with the colonization of North Asia and the Americas. This indicates that the relatively high $k_a/k_s$ for ATP6 is due to relaxed purifying selection specific to this gene, rather than a selection pressure that has occurred in a specific geographic region due to climate.

In studying the distribution of non-synonymous changes within genes, we noted a striking difference in $k_a/k_s$ between the functional domain and the remaining part of the gene. This result points to the underlying complexity in studying the pattern of selection and provides further support that purifying selection is the main determinant of the $k_a/k_s$ ratio. The pattern of accumulation of changes in both genes is consistent with that expected through the action of purifying selection, although somewhat more relaxed in ATP6. The results do not allow us to distinguish between the above model of selection (gradual accumulation of extended periods of time) and alternative models of selection (e.g. a model in which most changes are deleterious, but a few are adaptive). More extensive modelling is needed to determine the different types of selection scenarios that can be distinguished using this methodology. However, for ATP6, the gradual increase in accumulation of amino acid changes spans a time period of up to 60,000 years, which essentially covers the full time-span since modern humans left Africa and is clearly much longer than the time that humans have been exposed to cold climate in association with the colonization of North Asia and the Americas. This indicates that the relatively high $k_a/k_s$ for ATP6 is due to relaxed purifying selection specific to this gene, rather than a selection pressure that has occurred in a specific geographic region due to climate.
and vice versa, mitochondrial-encoded residues that are in close proximity to nuclear-encoded residues are evolving more rapidly, suggesting a functional dependence between subunits encoded by the two genomes (18). The need for functional complementarity has also been demonstrated by experimental inter-specific combinations of mitochondrial and nuclear genomes (19). We found the evolutionary rate outside the functional domain to be almost 5-fold higher in comparisons within humans, when compared with about 2-fold higher in inter-species comparisons. The fact that the functional domain is more conserved than the remainder of the gene in both intra- and inter-specific comparisons supports the notion that this difference is due to varying levels of purifying selection. The underlying difference in the length of the sequence encoding the functional domain relative to gene length explains a major part of the gene-specific differences in ka/ks.

A number of sites contribute to the relatively high region-specific ka for ATP6 in North Asia, ND3 in Africa and cytB in Europe. These sites represent candidates for positive selection, but we believe the variability pattern is most parsimoniously explained by relaxed purifying selection and genetic drift. One such substitution is np 8794 in ATP6. This site contributes to the high ka/ks in North Asians and is found at high frequency in the Chukchi people of northeastern Siberia and in Canadian Inuits, while virtually absent among the other North Asian sequences in our data set. It could be argued that this is the expected pattern for a selective sweep. However, both the Chukchi and Canadian Inuit groups have had very small population sizes and under such conditions selection would have to be very strong to override the effect of genetic drift (20).

In summary, our examination does not lend support to the hypothesis of mtDNA adaptation to cold climate. It is not necessary to invoke adaptive selection to explain differences between genes or geographic regions when the data can be explained by relaxed purifying selection at some sites coupled with genetic drift. The mode by which amino acid changes have accumulated in ATP6 over time is also consistent with relaxed purifying selection operating over long-time periods, rather than recent strong selective sweeps. None of the substitutions implicated have been associated with any human physiological difference (2). Given the lack of a discernible effect of the mtDNA polymorphism in the European population on vital traits for survival (3), it seems premature to conclude that mtDNA variants have been selected for in the past in order to provide more efficient cold climate adaptation. Definitive evidence for adaptive selection on individual sites in human mtDNA requires functional tests in vitro using cell-based assays such as those recently employed in a study of cell respiration of four mouse mtDNA haplotypes.

SNP typing and DNA sequencing
MtDNA sequencing used primers described by Reider et al. (22) and the BigDye Primer Cycle Sequencing kit from Applied Biosystems (CA, USA). DNA sequencing was performed on an ABI3700 automated fragment analysis machine and analysed with DNA Sequencing Analysis software from Applied Biosystems (CA, USA). Sequence alignments were achieved using Sequencher (Gene Codes Corporation, Ann Arbor, MI, USA) software.

Genetic analyses

\[ka \text{ and } ks \text{ between human sequences were estimated using the modified Nei Gojobori method (23) by MEGA 2.1 software (24). These values were calculated over concatenated sequences relative to a majority-rule consensus of all 416 sequences. For interspecies comparisons, } ka \text{ and } ks \text{ were calculated pairwise using MEGA with the modified Nei Gojobori method and the Jukes–Cantor correction (25). The Neighbour-Joining tree (16) was reconstructed using PAUP 4.0b10 (26) using the Kimura 2 parameter model for nucleotide substitution (27). The ages of different clades were calculated using the mean branch length within the clade and a substitution rate of } 1.70 \times 10^{-8} \text{ substitutions per site per year (8).}\]

Statistical analyses

Statistical comparisons of two proportions [i.e. counts of non-synonymous substitutions (n) and synonymous substitutions (s)] were made with a two-tailed Fisher’s Exact test. Comparisons across more than two proportions were made with \(\chi^2\) tests.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

The samples used in this study were kindly provided by Dr Mikhail Voevoda, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia. Funding for this study was provided by the Swedish Natural Sciences Research Council and the Beijer Foundation.

Conflict of Interest statement. None declared.

REFERENCES


